MATH5805 – A Modern Introduction to Extreme Value Theory

Week 5 - Univariate Extremes: the Point Process representation (PP)

Dr Boris Béranger
School of Mathematics and Statistics
UNSW Sydney

CRICOS Provider Code 00098G

Lecture outline

1. Introduction
2. Limit laws
3. Using the limiting result
4. Link with existing models
5. Statistical application
6. Analysis of precipitation data

1. Introduction



Goals



This lecture is about introducing the **Point Process** representation of extremes.

In particular, we

- Define what **Point Process** are
- Introduce a framework where Point Processes can be used to model univariate extremes
- Establish the links with the Block Maxima and Peaks-over-Threshold approaches.

Note that throughout this lecture we will make the assumption that $X_1, \ldots, X_n$ is a sequence of i.i.d. random variables with **common distribution function** F.

Why Point Processes?

We have seen so far that we can characterise extremes using **block maxima (BM)**, **threshold exceedances (POT)** or ***r*-largest order statistics** (another alternative not considered in this course).

In fact, the BM and POT approaches are **special cases of a point process** representation, under which we use a Poisson process to approximate the occurrence of those values that exceed a (high) threshold.

Characterising extremes using point processes will lead to the same modelling results.

Advantages of the Point Process representation of extremes

- An interpretation of extreme value behaviour unifying previously introduced models
- Enables a more natural formulation of non-stationarity in threshold excesses.

Point Processes

Think of a **point process** simply as a **random distribution of points X_i in space**.

E.g.: times of avalanches, positions of stars in the sky, time/position of earthquakes times/epicentres, etc.

Definition [Point Process]

Let E denote the state space where the points live and $\{X_i, i > 1\}$ be a countable collection of points in E. A point process is characterised by a set of non-negative integer-valued random variables $N(A)$ such that

$$N(A) = \text{card}\{i : X_i \in A\}, \qquad \text{for } A \subset E,$$

i.e. $N(A)$ **counts the number of** $X_i \in A$.

E.g.: A can be a period of time.

The features of a point process can be summarised by

- its **intensity measure**

$$\Lambda(A) = \mathrm{E}\left(N(A)\right),$$

  i.e. the expected number of points in any subset $A \subset E$.

- its **intensity (density) function**

$$\lambda(\boldsymbol{x}) = \frac{\partial \Lambda(A)}{\partial x_1 \cdots \partial x_d},$$

  where $A = [a_1, x_1] \times \cdots \times [a_d, x_d] \subset \mathbb{R}^d$.

Definition [Poisson Point Process]

A point process N is called Poisson Point Process if the following conditions hold:

- For $A \subset E$ and $k \geq 0$,

$$\mathbb{P}(N(A) = k) = \begin{cases} e^{-\Lambda(A)} \frac{\Lambda(A)^k}{k!} & \text{if } \Lambda(A) < \infty, \\ 0 & \text{if } \Lambda(A) = \infty, \end{cases}$$

- For any $m \geq 1$, if $A_1, \ldots, A_m$ are mutually disjoint subsets of E, then $N(A_1), \ldots, N(A_m)$ are independent random variables.

Poisson Point Processes

A Poisson Point Process is said to be **homogenous** if its intensity Λ is proportional to the length, area, volume, etc. of a set A. The intensity function is $\lambda(\boldsymbol{x}) = \lambda$ where λ represents the average number of point per unit of length, area, volume, etc.

E.g: For a univariate Poisson Process and $A = [t_1, t_2] \subset E$, $\Lambda(A) = \lambda(t_2 - t_1)$.

A Poisson Point Process is said to be **non-homogenous** if its intensity function is non-stationary, it depends on the location of the space A:

$$\Lambda(A) = \int_A \lambda(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}.$$

Inference for Poisson Point Processes

For simplicity we will **assume that A is a one-dimensional space**.

Suppose that $x_1, \ldots, x_n$ **have been observed in a region** $A \subset \mathbb{R}$ and that these are the realisation of a Poisson process on A, with intensity function $\lambda(\cdot; \theta)$ for some value of θ.

Information: the points have occurred at a number of known locations but nowhere else.

We let $I_i = [x_i, x_i + \delta_i], i = 1, \ldots, n$ be a set of small intervals around the observations, and $\mathcal{I} = A \setminus \bigcup_{i=1}^{n} I_i$. The **likelihood** is then defined by

$$\begin{aligned} L(\theta; x_1, \ldots, x_n) &= \mathbb{P}\left(N(I_1) = 1, \ldots, N(I_n) = 1, N(\mathcal{I}) = 0\right) \\ &= \mathbb{P}\left(\mathbf{N}(\mathcal{I}) = \mathbf{0}\right) \times \prod_{\mathbf{i=1}}^{\mathbf{n}} \mathbb{P}\left(\mathbf{N}(\mathbf{I_i}) = \mathbf{1}\right) \end{aligned}$$

Inference for Poisson Point Processes

Using $\Lambda(I_i;\theta) = \int_{x_i}^{x_i+\delta_i} \lambda(x)\mathrm{d}x \approx \lambda(x_i)\delta_i$, we have

$$\begin{aligned}\mathbb{P}\left(\mathbf{N(I_i) = 1}\right) &= \exp\left\{-\Lambda(I_i;\theta)\right\}\Lambda(I_i;\theta)\\ &\approx \exp\left\{-\lambda(x_i)\delta_i\right\}\lambda(x_i)\delta_i\\ &\approx \lambda(\mathbf{x_i})\delta_\mathbf{i},\end{aligned}$$

where we have used $\exp\{-\lambda(x_i)\delta_i\} \approx 1$ for small δ_i. Similarly, since the δ_i are small,

$$\mathbb{P}(N(\mathcal{I}) = 0) = \exp\{-\Lambda(\mathcal{I})\} \approx \exp\{-\Lambda(A)\}$$

Therefore the **likelihood** is

$$L(\theta; x_1, \ldots, x_n) \approx \exp\{-\Lambda(A)\} \times \prod_{i=1}^{n} \lambda(x_i)\delta_i \approx \exp\{-\Lambda(A)\} \prod_{i=1}^{n} \lambda(x_i)$$

Example

Consider a homogenous Poisson Point Process with points observed over $A = [t_1, t_2]$.

As previously shown $\Lambda(A) = \lambda(t_2 - t_1)$, therefore the likelihood is

$$L(\lambda; x_1, \ldots, x_n) = \exp\{-\lambda(t_2 - t_1)\}\, \lambda^n.$$

The log-likelihood is

$$\ell(\lambda; x_1, \ldots, x_n) = -\lambda(t_2 - t_1) + n\lambda,$$

leading to the maximum likelihood estimator $\hat{\lambda} = n/(t_2 - t_1)$ which is the *empirical rate of occurrence*.

Definition [Convergence in distribution of point processes]

A sequence of point processes $N_1, N_2, \ldots$ on a set E converges in distribution to a *limiting point process* N, written $N_n \xrightarrow{d} N$, if

$$(N_n(A_1), \ldots, N_n(A_k)) \xrightarrow{d} (N(A_1), \ldots, N(A_k)),$$

for all choices of k and all bounded sets $A_1, \ldots, A_k \subset E$ such that $\mathbb{P}(N(\partial A_j) = 0) = 1$ for $j = 1, \ldots, k$, where $\partial(A_j)$ is the boundary of A_j.

Recall the definition of convergence in distribution from the Week 2 lecture, slide 11.



2. Limit Laws

Basic idea

Assume that F **is unknown** and that **the Extremal Type Theorem applies**, i.e. $\exists (a_n) > 0, (b_n) \in \mathbb{R}$ s.t.

$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq y\right) \longrightarrow G_\xi(y; \mu, \sigma) = \exp\left\{-\left[1 + \xi\left(\frac{y-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right\},$$

for some $\mu, \sigma > 0$ and ξ. We construct a sequence of point processes on $\mathbb{R}^2$ by

$$N_n = \left\{\left(\frac{i}{n+1}, \frac{X_i - b_n}{a_n}\right) : i = 1, \ldots, n\right\} \tag{1}$$

Theorem

Let y_* and y^* be the lower and upper endpoints of G. On regions of the form $[0,1] \times [u, \infty)$,

$$N_n \xrightarrow{d} N, \quad \text{as } n \to \infty,$$

where N is a non-homogeneous Poisson point process.

Corollary

On $A = [t_1, t_2] \times [y, y^*)$, the intensity measure of the limiting Poisson point process N is

$$\Lambda(A) = (t_2 - t_1)\left[1 + \xi\left(\frac{y-\mu}{\sigma}\right)\right]^{-1/\xi}. \tag{2}$$

Some justification:

The probability of each point in N_n falling into the region $A_u = [0,1] \times [u, \infty]$ is

$$p_n = \mathbb{P}\left(\frac{X_i - b_n}{a_n} > u\right) \approx \frac{1}{n}\left[1 + \xi\left(\frac{u-\mu}{\sigma}\right)\right]^{-1/\xi}.$$

Let $N_n(A)$ be the number of points of N_n in region A. Because X_i's are i.i.d., $N_n(A_u)$ follows a binomial distribution $\text{Bin}(n, p_n)$.

Basic idea

As $n \to \infty$, the binomial distribution converges to a Poisson distribution with rate

$$\Lambda(A_u) = \left[1 + \xi \left(\frac{u - \mu}{\sigma}\right)\right]^{-1/\xi}.$$

Note that the rate is homogeneous in the time direction. Therefore, this is exactly the characterising property of a non-homogeneous Poisson process with intensity function

$$\lambda(t, v) = \frac{1}{\sigma}\left[1 + \xi \left(\frac{v - \mu}{\sigma}\right)\right]^{-1/\xi - 1}.$$

Since the process in homogeneous in time we obtain the intensity measure of the limiting process on any region A of the form $[t_1, t_2] \times (u, \infty)$, with $0 \leq t_1 < t_2 \leq 1$.

For all such non-overlapping set A the $N(A)$ are independent and Poisson distributed, therefore N is a Poisson point process.



3. Using the limiting result

Limiting Poisson point process - Examples

Example - Exponential(1)

The distribution function is $F(x) = 1 - e^{-x}$ for $x > 0$. We have previously shown that for $a_n = 1$ and $b_n = \log n$,

$$\frac{M_n - b_n}{a_n} \xrightarrow{d} G_0 \,(\text{Gumbel}).$$

The endpoints of G are $y_* = -\infty$ and $y^* = +\infty$, and the sequences of point process

$$N_n = \left\{ \left(\frac{i}{n+1}, X_i - \log n \right) : i = 1, \ldots, n \right\},$$

converge to a Poisson process with intensity measure on $A = [t_1, t_2] \times [y, \infty)$, for $y > -\infty$ given by

$$\Lambda(A) = (t_2 - t_1) \exp\{-y\}.$$

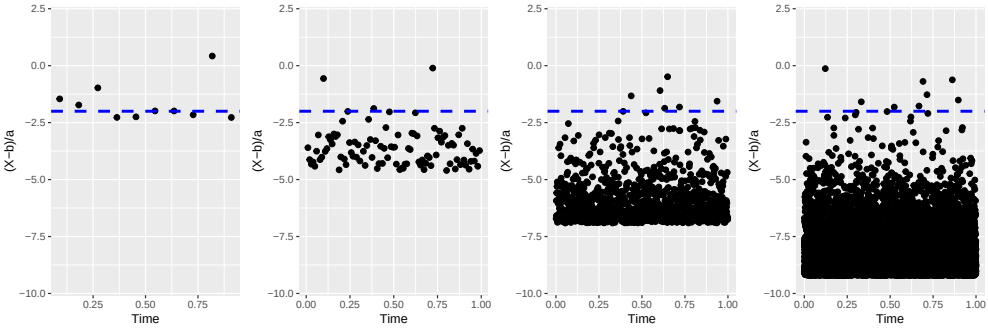
This is the limit of (2) in the case $\mu = 0, \sigma = 1$ as $\xi \to 0$.

Limiting Poisson point process - Examples

Example - Exponential(1)

Poisson limit when rescaling samples of size 10, 100, 1000 and 10000 (left to right).

Vertical dashed line indicate a high threshold $u = -2$.



Limiting Poisson point process - Examples

Example - Exponential(1) Note that when $u = -2$ the intensity measure on $A = [0, 1] \times [-2, +\infty)$ becomes

$$\Lambda(A) = \exp\{2\} \approx 7.39.$$

```
## For n= 10 , #obs above  -2  is  6
## For n= 100 , #obs above  -2  is  4
## For n= 1000 , #obs above  -2  is  6
## For n= 10000 , #obs above  -2  is  8
```

Limiting Poisson point process - Examples

```
n <- c(10, 100, 1000, 10000); t <- -2;
plot_list <- list()
set.seed(123)

for(i in 1:length(n)){
  dat <- data.frame(data = rexp(n[i]) - log(n[i]),
                    time = seq(1,n[i]) / (n[i]+1))

  plot_list[[i]] <- ggplot(dat, aes(time,data)) +
                      geom_point(size=2) +
                      scale_y_continuous(name="(X-b)/a", limits=c(-9.5, 2)) +
                      xlab("Time") +
                      geom_hline(yintercept=t, colour="blue", linetype="dashed",
                                 linewidth=1.2)
  cat("For n=", n[i],", #obs above ", t, " is ", sum(dat$data>t),"\n")
}
grid.arrange(grobs=plot_list, ncol=4) #from gridExtra package
```

Example - Fréchet(1)

The distribution function is $F(x) = \exp(-1/x)$ for $x > 0$. We have previously shown that for $a_n = n$ and $b_n = 0$,

$$\frac{M_n - b_n}{a_n} \xrightarrow{d} G_1 \text{ (Standard Fréchet).}$$

Here the endpoints of G are $y_* = 0$ and $y^* = +\infty$, and the sequences of point process

$$N_n = \left\{ \left(\frac{i}{n+1}, \frac{X_i}{n} \right) : i = 1, \ldots, n \right\},$$

converge to a Poisson process with intensity measure on $A = [t_1, t_2] \times [y, \infty)$, for $y > -\infty$ given by

$$\Lambda(A) = (t_2 - t_1) y^{-1}.$$

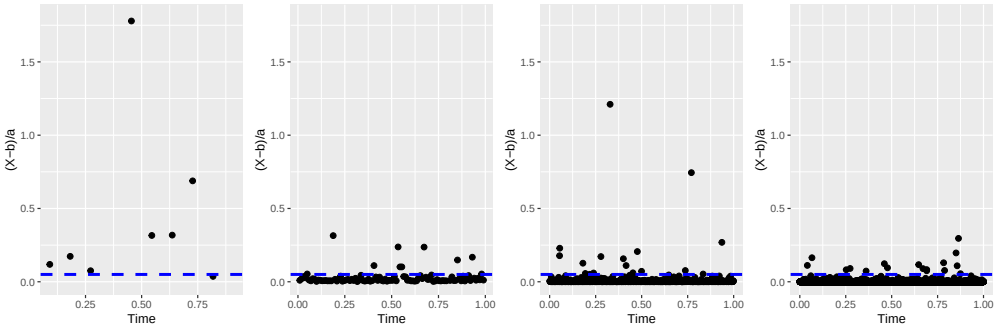
This is the limit of (2) in the case $\mu = 1, \sigma = 1$ and $\xi = 1$.

Limiting Poisson point process - Examples

Example - Fréchet(1)

Poisson limit when rescaling samples of size 10, 100, 1000 and 10000 (left to right).

Vertical dashed line indicate a high threshold $u = 0.05$.



Limiting Poisson point process - Examples

Example - Fréchet(1)

Note that when $u = 0.05$ the intensity measure on $A = [0,1] \times [0.05, +\infty)$ becomes

$$\Lambda(A) = 1/0.05 = 20.$$

```
## For n= 10 , #obs above  0.05  is  9
## For n= 100 , #obs above  0.05  is  11
## For n= 1000 , #obs above  0.05  is  17
## For n= 10000 , #obs above  0.05  is  17
```

Limiting Poisson point process - Examples

```
n <- c(10, 100, 1000, 10000); t <- 0.05;

plot_list <- list()
set.seed(123)
for(i in 1:length(n)){
  dat <- data.frame(data = rfrechet(n[i]) / n[i], # requires evd package
                    time = seq(1,n[i]) / (n[i]+1))

  plot_list[[i]] <- ggplot(dat, aes(time,data)) +
                      geom_point(size=2) +
                      scale_y_continuous(name="(X-b)/a", limits=c(0, 3.5)) +
                      xlab("Time") +
                      geom_hline(yintercept=t, colour="blue", linetype="dashed",
                                 linewidth=1.2)
  cat("For n=", n[i],", #obs above ", t, " is ", sum(dat$data>t),"\n")
}
grid.arrange(grobs=plot_list, ncol=4)
```

Example - Uniform(0,1)

The distribution function is $F(x) = x$ for $0 \leq x \leq 1$. We have previously shown that for $a_n = 1/n$ and $b_n = 1$,

$$\frac{M_n - b_n}{a_n} \xrightarrow{d} G_{-1} \text{ (Weibull)}.$$

Here the endpoints of G are $y_* = -\infty$ and $y^* = 0$, and the sequences of point process

$$N_n = \left\{ \left(\frac{i}{n+1}, n(X_i - 1) \right) : i = 1, \ldots, n \right\},$$

converge to a Poisson process with intensity measure on $A = [t_1, t_2] \times [y, 0]$, for $y < 0$ given by

$$\Lambda(A) = -(t_2 - t_1)y.$$

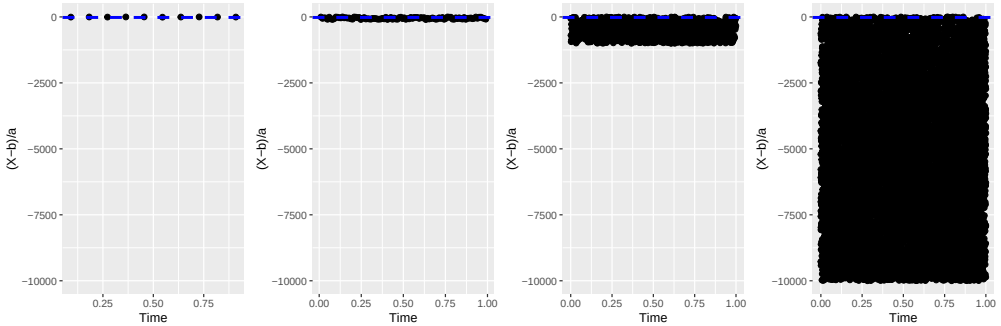
This is the limit of (2) in the case $\mu = -1, \sigma = 1$ and $\xi = -1$.

Limiting Poisson point process - Examples

Example - Uniform(0,1)

Poisson limit when rescaling samples of size 10, 100, 1000 and 10000 (left to right).

Vertical dashed line indicate a high threshold $u = -20$.



Limiting Poisson point process - Examples

Example - Uniform(0,1)

Note that when $u = -20$ the intensity measure on $A = [0,1] \times [-20, 0]$ becomes

$$\Lambda(A) = 20.$$

```
## For n= 10 , #obs above  -20  is  10
## For n= 100 , #obs above  -20  is  18
## For n= 1000 , #obs above  -20  is  19
## For n= 10000 , #obs above  -20  is  19
```

Limiting Poisson point process - Examples

```
n <- c(10, 100, 1000, 10000); t <- -20;

plot_list <- list()
set.seed(123)
for(i in 1:length(n)){
  dat <- data.frame(data = n[i] * (runif(n[i]) - 1),
                    time = seq(1,n[i]) / (n[i]+1))

  plot_list[[i]] <- ggplot(dat, aes(time,data)) +
                      geom_point(size=2) +
                      scale_y_continuous(name="(X-b)/a", limits=c(-10000, 0)) +
                      xlab("Time") +
                      geom_hline(yintercept=t, colour="blue", linetype="dashed",
                                 linewidth=1.2)
  cat("For n=", n[i],", #obs above ", t, " is ", sum(dat$data>t),"\n")
}
grid.arrange(grobs=plot_list, ncol=4)
```

Limiting Poisson point process - Examples

If we were interested in the set $A = [0, 1/2] \times [-20, 0]$, the intensity measure would be

$$\Lambda(A) = -\frac{1}{2} \times (-20) = 10.$$

```
n <- c(10, 100, 1000, 10000); t <- -20;

set.seed(123)
for(i in 1:length(n)){
  dat <- data.frame(data = n[i] * (runif(n[i]) - 1),
                    time = seq(1,n[i]) / (n[i]+1))

  cat("For n=", n[i],", #obs above ", t, "on [0,1/2], is ",
      sum((dat$data>t) * (dat$time<=1/2)),"\n")
}
```

Limiting Poisson point process - Examples

The previous code outputs:

```
## For n= 10 , #obs above  -20 on [0,1/2], is  5
## For n= 100 , #obs above  -20 on [0,1/2], is  9
## For n= 1000 , #obs above  -20 on [0,1/2], is  11
## For n= 10000 , #obs above  -20 on [0,1/2], is  9
```



4. Link with existing models

Relation with the GEV family of distributions

Let $M_n = \max_{i=1,\dots,n} X_i$ and consider the sequence of point processes

$$N_n = \left\{ \left(\frac{i}{n+1}, \frac{X_i - b_n}{a_n} \right) : i = 1, \dots, n \right\}$$

Taking $A_z = (0,1) \times (z, \infty)$, the event $\left\{ \frac{M_n - b_n}{a_n} \leq z \right\}$ is equivalent to the event $\{N_n(A_z) = 0\}$ and therefore

$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq z \right) = \mathbb{P}\left(N_n(A_z) = 0\right) \to \mathbb{P}\left(N(A_z) = 0\right).$$

Since N follows a Poisson distribution

$$\mathbb{P}\left(N(A_z) = 0\right) = \exp\left\{-\Lambda(A_z)\right\} = \exp\left\{ -\left[1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi} \right\}.$$

$\Longrightarrow$ The limiting distribution of normalised block maxima is the GEV

Relation with the GPD family of distributions

Let $N_n^i = \left(\frac{i}{n+1}, \frac{X_i - b_n}{a_n}\right)$ be the i-th point in N_n. For a sufficiently large u, we have

$$\begin{aligned}
\mathbb{P}\left(\frac{X_i - b_n}{a_n} > u + x \middle| \frac{X_i - b_n}{a_n} > u\right) &= \frac{\mathbb{P}\left(\frac{X_i - b_n}{a_n} > u + x\right)}{\mathbb{P}\left(\frac{X_i - b_n}{a_n} > u\right)} \\
&= \frac{\mathbb{P}\left(N_n^i \in (0,1) \times (u + x, \infty)\right)}{\mathbb{P}\left(N_n^i \in (0,1) \times (u, \infty)\right)} \\
&\approx \frac{\frac{1}{n}\left[1 + \xi\left(\frac{u+x-\mu}{\sigma}\right)\right]^{-1/\xi}}{\frac{1}{n}\left[1 + \xi\left(\frac{u-\mu}{\sigma}\right)\right]^{-1/\xi}} \\
&= \left[1 + \frac{\xi x}{\tilde{\sigma}}\right]^{-1/\xi}.
\end{aligned}$$

(For more details, refer to **Week 3 - Lecture, Slide 19**)

5. Statistical application



Statistical application

Following on **Smith (1989)**, for statistical application we:

- assume that the limiting process is a reasonable approximation for finite n above a high threshold u.
- absorb the unknown scaling coefficients a_n and b_n into the intensity function through μ and σ in order to work directly with the original series X_i's.
- rescale the intensity so that the **annual maximum** is $\mathrm{GEV}(\mu, \sigma, \xi)$ distributed.

Therefore, we model the series $\{(j, X_j) : i = 1, \ldots, n\}$ as a non-homogeneous Poisson point process over A_u with intensity function

$$\Lambda(A) = n_b(t_2 - t_1)\left[1 + \xi\left(\frac{y - \mu}{\sigma}\right)\right]^{-1/\xi},$$

where n_b is the number of blocks (e.g., year) and the time limits rescaled to $0 \leq t_1 < t_2 \leq 1$.

Likelihood maximisation is the most natural.

For a region of the form $A_v = [0,1] \times (v, \infty)$ where $v > u$, the likelihood is

$$L(\theta) = \exp\{-\Lambda(A_v)\} \prod_{i=1}^{N_{A_v}} \lambda(t_i, x_i)$$

$$\propto \exp\left\{-n_b \left[1 + \xi\left(\frac{v-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} \times \prod_{i=1}^{N_{A_v}} \frac{1}{\sigma}\left[1 + \xi\left(\frac{x_i - \mu}{\sigma}\right)\right]^{-1/\xi - 1},$$

where $x_1, \ldots, x_{N_{A_v}}$ is an enumeration of the N_{A_v} points that exceed the threshold v.

This method has the advantage that the model parameters are GEV without dependence on u, and that all parameters are estimated in a single step. Invariance of all parameters to u allows for a varying threshold, necessary for non-stationary series.

Return level estimation

Assuming that we are in the stationary case, we have

$$\begin{aligned} 1 - \frac{1}{\text{return period}} &= \mathbb{P}\left(\max_{i=1,\dots,n} X_i \leq \text{return level} \right) \\ &= \mathbb{P}\left(X \leq \text{return level} \right)^n \\ &\approx \left(1 - \frac{1}{n} \left[1 + \xi \left(\frac{\text{return level} - \mu}{\sigma} \right) \right]^{-1/\xi} \right)^n, \end{aligned}$$

from which we deduce

$$\text{return level} = \mu + \frac{\sigma}{\xi} \left\{ 1 - n^{-\xi} \left(1 - \left[1 - \frac{1}{\text{return period}} \right]^{1/n} \right)^{-\xi} \right\}.$$

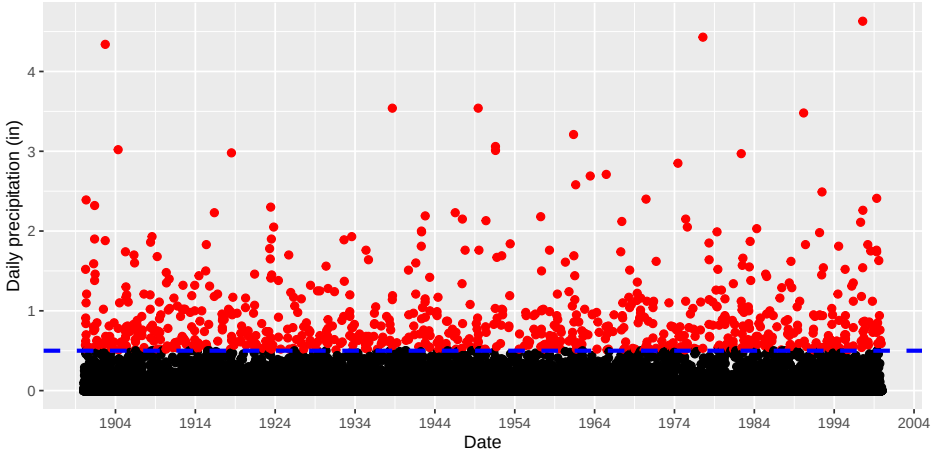
*** we need to fulfill the condition** $[1 + \xi(\textbf{return level} - \mu)/\sigma] > 0$.

6 -Analysis of precipitation data



Analysis of precipitation data

Recall the example of daily precipitation at Fort Collins, Colorado. In the last lecture we selected $u = 0.5$ resulting in 759 observations above the threshold.



Analysis of precipitation data

For the previously fitted models, the estimated parameters are

Model	Parameter	Estimate	Std. Error	Conf. Int.
Block Maxima	Location (μ)	1.347	0.062	$(1.226, 1.468)$
	Scale (σ)	0.533	0.049	$(0.437, 0.628)$
	Shape(ξ)	0.174	0.092	$(-0.07, 0.354)$
POT	Scale ($\tilde{\sigma}$)	0.361	0.021	$(0.321, 0.628)$
	Shape(ξ)	0.189	0.045	$(0.101, 0.276)$

Let's now fit the Poisson Point Process...

Analysis of precipitation data

```r
fit.pp <- fevd(Prec, Fort, threshold=0.5, type="PP", units="inches")
# fit.pp
ci(fit.pp, type="parameter")
```

```
## fevd(x = Prec, data = Fort, threshold = 0.5, type = "PP", units = "inch
##
## [1] "Normal Approx."
##
##           95% lower CI  Estimate 95% upper CI
## location     1.3059797 1.3912641    1.4765485
## scale        0.4587417 0.5291348    0.5995279
## shape        0.1013206 0.1886400    0.2759593
```

Analysis of precipitation data

The estimated parameters are

Model	Parameter	Estimate	Std. Error	Conf. Int.
Poisson Process	Location (μ)	1.391	0.046	$(1.306, 1.477)$
	Scale (σ)	0.529	0.036	$(0.459, 0.600)$
	Shape(ξ)	0.189	0.045	$(0.101, 0.276)$

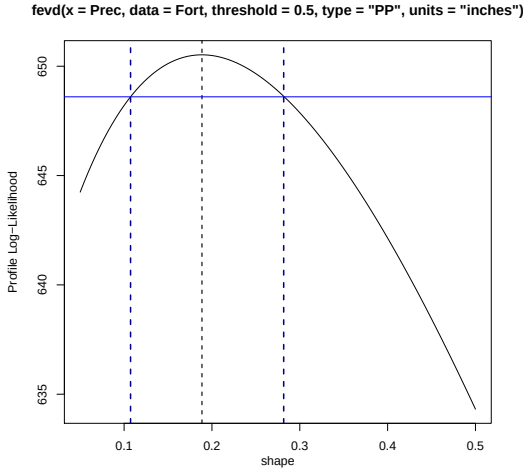
Seems to be heavy tailed!

Comparing with the estimates of the GPD model, recall that $\sigma_{\mathbf{GPD}} = \sigma_{\mathbf{GEV}} + \xi(u - \mu)$.
Here we have $0.529 + 0.189(0.5 - 1.391) \approx 0.361 = \sigma_{\mathbf{GPD}}$.

Let's have a closer look at the shape parameter ξ

```
ci(fit.pp, type="parameter", which.par=3, xrange=c(0.05,0.5), nint=100,
   method="proflik", verbose=TRUE)
```

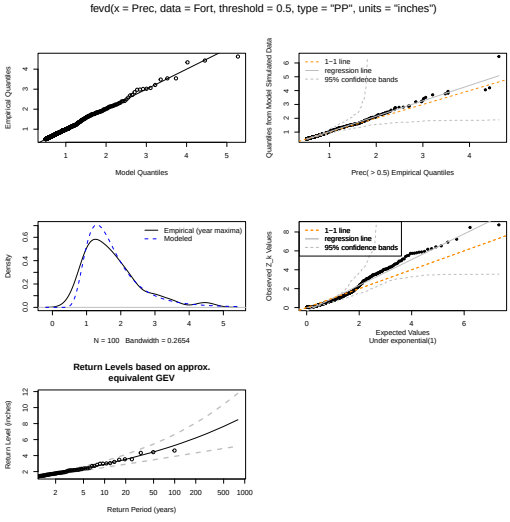
Analysis of precipitation data



Using the profile likelihood confirms that there is a heavy tail phenomena.

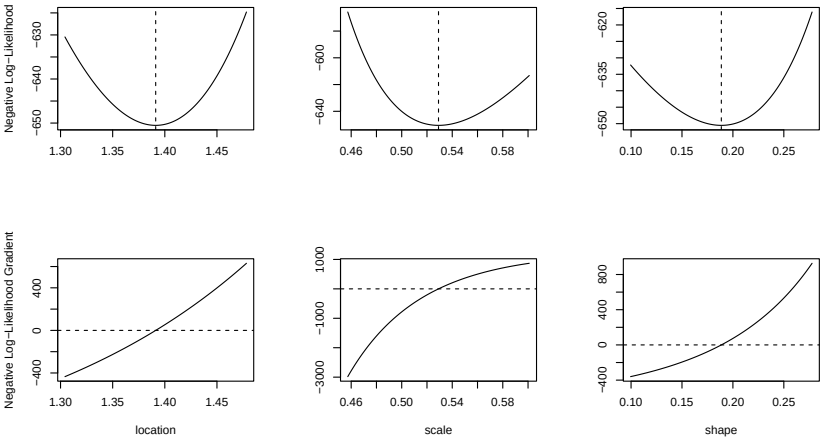
Analysis of precipitation data

We assess the quality of the fit by running some diagnostic plots



Analysis of precipitation data

fevd(x = Prec, data = Fort, threshold = 0.5, type = "PP", units = "inches")



Analysis of precipitation data

The return levels associated with return periods of 20, 50 and 100 years are

Return period	Estimate	Conf. Int.
20-year	3.498	(2.906, 4.091)
50-year	4.442	(3.468, 5.417)
100-year	5.267	(3.899, 6.634)

```
return.level(fit.pp, return.period=c(20,50,100), do.ci=TRUE)
```

```
## fevd(x = Prec, data = Fort, threshold = 0.5, type = "PP", units = "inches")
##
## [1] "Normal Approx."
##
##                        95% lower CI Estimate 95% upper CI
## 20-year return level       2.906144 3.498361     4.090577
## 50-year return level       3.467754 4.442253     5.416751
## 100-year return level      3.899323 5.266646     6.633970
```

Analysis of precipitation data

Question

What is the probability of a daily precipitation above the threshold of 0.5 in to be at least as large as that during the 28 July 1997 flood (4.63 in)?

We are interested in

$$\mathbb{P}(X \geq 4.63).$$

```
fit.pp.sum <-  summary(fit.pp, silent=TRUE)
pevd(4.63, loc=fit.pp.sum$par[1], scale=fit.pp.sum$par[2],
     shape=fit.pp.sum$par[3], type="PP",threshold=0.5, lower.tail=FALSE)
```

```
## [1] 0.01694536
```

Analysis of precipitation data

Question

What about the yearly frequency of daily precipitations above the 0.5 threshold?

Consider $A = (0,1) \times [u, \infty)$, we are interested in estimating the intensity measure:

$$\Lambda(A) = \left[1 + \xi \left(\frac{u - \mu}{\sigma}\right)\right]^{-1/\xi}.$$

Empirically estimate:

$$\hat{\Lambda}(A) = 365.25 \times \frac{\text{Number of } X_i > 0.5}{\text{Number of } X_i} = 365.25 \times \frac{759}{36524} \approx 7.590$$

Model estimate:

$$\hat{\Lambda}(A) = \left[1 + \hat{\xi} \left(\frac{u - \hat{\mu}}{\hat{\sigma}}\right)\right]^{-1/\hat{\xi}} \approx 7.590.$$

Analysis of precipitation data

```
# Empirical estimate
365.25 * sum(Fort$Prec > 0.5) / nrow(Fort)
```

```
## [1] 7.590208
```

```
# Model estimate
val <- 1 + fit.pp.sum$par[3] * ((0.5-fit.pp.sum$par[1])/fit.pp.sum$par[2])
val^(-1/fit.pp.sum$par[3])
```

```
##    shape
## 7.590203
```

References



Sources and recommending reading

1. **[BGST04]** Beirlant, J., Goegebeur, Y., Segers, J. & Teugels J. (2004) *Statistics of Extremes: Theory and Applications*. Wiley Series in Probability and Statistics.
2. **[C01]** Coles, S. (2001) An Introduction to Statistical Modeling of Extreme Values. Springer. Chapter 1-3.
3. **Smith, R. L. (1989)**. Extreme value analysis of environmental time series: an example based on ozone data (with discussion), *Statistical Science*, 4, pp.367–393.