Estimation of severe crash frequency using two surrogates

Zhankun Chen

Master's thesis
2022:E47

Lund University

Faculty of Science
Centre for Mathematical Sciences
Mathematical Statistics

Centrum Scientiarum Mathematicarum

Abstract

This thesis is concerned with the estimation of crash frequency based on the bivariate modeling of surrogate measures of safety (SMoS), which serve as indicators for traffic risk. Using the SMoS, any traffic conflict between two road users can be described by their proximity together with their hypothetical consequence. We quantify traffic conflicts of different severity as random vector of proximity SMoS and consequential SMoS, and define the traffic risk as the probability measure over the random vector of SMoS pair. We use EVT both in its bivariate context and in approximating the marginal distribution of proximity SMoS, which is combined with copula, to compute the probability of severe collision. The 10-year frequencies of severe collision are also computed based on the fitted models. From a methodological point of view, the copula approach with EV margin is more favorable than bivariate EV models, as collisions of lower severity can also be computed. From an implementation point of view, the bivariate EV model is more favorable, as the assumptions on the marginal distribution are defined by the model. The new approach that combines EV distributions and copula is found to have the most accurate estimated crash frequency given that the police report was used a reference, for our data set.

Acknowledgement

First of all I would like to thank my supervisor Nader Tajvidi. He offers two courses in Extreme Value Theory at Lund university, which I highly recommend to interested readers. Besides these two enlightening courses, Nader helped me to explore theory beyond the curriculum and guided me through the degree project, so that I can turn some of my wild fantasies into reality. He deserves my deepest gratitude.

Since this degree project is deeply related to road traffic safety, it would not have been possible without the Transport and Roads group from LTH. I am sincerely grateful for the hospitality from the whole Transport and Roads group, especially towards Aliaksei Laureshyn, Carmelo D'agostino, Oksana Yastremska-Kravchenko, for whom have offered me enormous help in the conceptual ideas of traffic safety as well as a cozy workplace. Their enthusiasm and optimism have been valuable inspiration for me.

Johannes Lindvall offer helped me in revising the report by pointing out many typos which I would otherwise not noticed. He also provided me with many valuable suggestions regarding the structure and the aesthetics. His contribution to the final draft must be noted.

Special thanks to my family and my friends, who always have faith in me, and to my feline friends Camila, Syla, Stella, Simba, Sixten and Tusse –they have brought so much joy to my life.

Contents

Abbreviations and notations

BM:Block Maxima
DoA: Domain of Attraction
EVT: Extreme Value Theory
FML: Full Maximum Likelihood
GEV: Generalized Extreme Value distribution
GoF: Goodness of Fit
GPD: Generalized Pareto distribution
i.i.d: Independently identically distributed
LminD: The left-turning vehicles under min distance measurements
LPETD: The left-turning vehicles under first PET distance measurements
max i.d: maximum infinite divisible
MEVD: Multivariate Extreme Value distribution
MLE: Maximum Likelihood Estimator
POT: Peaks over threshold in univariate case
POT1: Peaks over thresholds of type 1 exceedences in multivariate case
POT2: Peaks over thresholds of type 2 exceedences in multivariate case
PET: Post Encroachment Time
r.v: random variable
r.vt: random vector
r.vs: random variables
SMoS: Surrogate Measure of Safety
w.r.t: with respect to
w.l.o.g: without loss of generality
$\Phi(x)$: standard normal distribution
$x_F := \inf\{x \in \mathbb{R} : F(x) > 0\}$: the left end point of distribution function F
$x^F := \sup\{x \in \mathbb{R} : F(x) < 1\}$: the right end point of distribution function F
A^c: the complement of the set A.

1 Introduction

Road traffic accidents is one of the most frequent cause of death. As being summarized in the 2018 WHO report *"Global status report on road safety"*, there were 1.35 million people killed and more than 50 million people injured in traffic accidents. The traffic safety can be improved by infrastructures. To evaluate the effect of the improvement, the **expected** crash frequency is of great importance, as accidents are still extremely rare events relative to the number of all traffic interactions. The observation for collision in each cite is limited and thus considered as extreme events. Frequency of such extreme events can be interpreted as the risk of certain cite. We intend to formulate the problem more formally using Surrogate Measure of Safety (SMoS).

In the studying of traffic risk, the term "traffic conflict" is more general than accident. A traffic accident is defined as the collision of multiple road users while the traffic conflict refers to the situation where two or more road users are on the colliding course. In our project we focus on only two road users at the same time hence the definitions are simplified accordingly. A traffic conflict is terminated either when one road user has left the colliding course or the evasions of two road users are successful. Otherwise the conflict will end up in a collision. The severity of a collision is evaluated based on the unleashed energy and the protection offered by the vehicles.

Essentially we want to answer two questions: how close are two road users to a collision (proximity)? How bad is the collision going to be (severity)? Proximity and severity of collision are two aspects for investigating traffic risks in road safety context. To quantify the risk of traffic conflicts, we use SMoS. From the proximity aspect we employ proximity SMoS, which measures the closeness to collision. The characterization of traffic conflicts based on proximity SMoS can be visualized as a pyramid, this is a classic concept that appeared first in Hydén [1987]

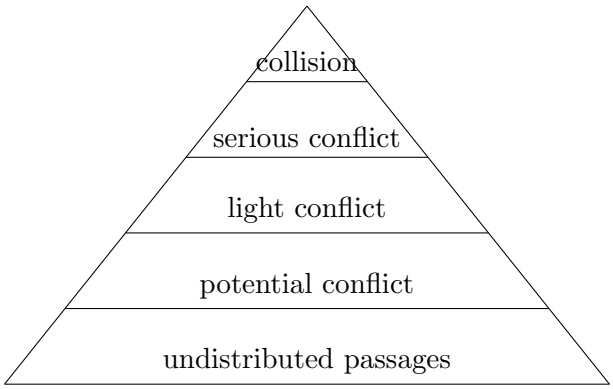


Figure 1: Hierarchy pyramid of traffic events

The hierarchy structure of traffic events motivates the use of order statistics to analyze collisions. We characterize the seriousness of a conflict by the low values of

proximity SMoS. Suppose $X_1, \cdots, X_N$ are samples of proximity SMoS of N encounters and $X_{(1)}, \cdots, X_{(N)}$ be its order statistics. Because of the well-order property of proximity indicators, if $X_{(1)}$ is greater than 0 i.e it is not a collision, then there is no collisions in $X_1, \cdots, X_N$. In another word:

$$P(\text{No collision}) = 1 - P(X_{(1)} > 0)$$

Let $X \mapsto T(X)$ be a monotonic decreasing transformation, then the probability above can be equivalently stated as:

$$P(\text{No collision}) = P\big(\max_i(T(X_1), \cdots, T(X_N)) < T(0)\big)$$

As the number of traffic encounters increases, the extreme order statistics will converge to a class of non-degenerate distributions, called **Extreme Value Distributions**. Hence we can use extreme value theory (EVT) to compute the probability of having a collision, which has the following advantages:

First and the most importantly is that the proximity is a continuous random variable that has support on the positive real axis, meaning that the collision $X \leq 0$ is an event that has zero Lebesgue measure. It is possible in EVT modeling that the limiting distribution of extreme order statistics has supports which strictly includes 0. Consequently we can compute a non-zero probability for having a collision. Another major improvement brought by the EVT approach is the robustness, since no demands are imposed on the underlying distribution of temporal SMoS. Next we will discuss briefly the applications of single variate EVT to traffic risk, more precisely, the proximity SMoS.

1.1 Extreme value theory in analysis of i.i.d near-crash events

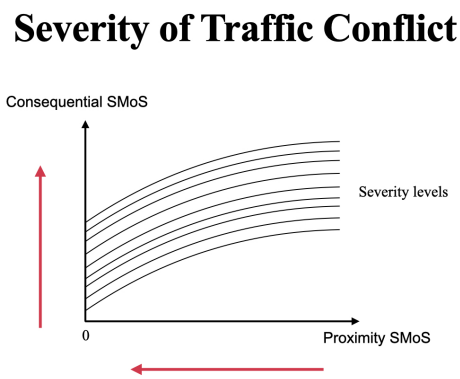
The use of EVT in analysing road safety arose in the recent twenty years. The earliest statistical modeling dated back to Songchitruksa and Tarko [2006], where block maxima models were used in analysing the post encroachment time (PET) at a signalized intersection. The framework of EVT analysis based on the Hyden event pyramid was developed later in Tarko [2012], where the formulation of peaks over threshold model in road safety context was discussed in details. The two univariate extreme value models were then applied to model the proximity under different situations, usually in the form of comparative studies. For example Zheng et al. [2014a] modeled the PET of freeway entries; Borsos et al. [2020] modeled *Time To Collision* of signalized intersections with left-turning and straight moving road users; Farah and Azevedo [2017] modeled minTTC between two road users moving towards the each other in a passing maneuver.

The first attempt of modeling crash proximity with bivariate extreme value theory appeared in Jonasson and Rootzén [2014]. In this paper bivariate modeling

with two proximity SMoS as margins were used to reduce the selection bias due to different types of proximity SMoS. [1] The bivariate extreme value models have been validated with different data sets and with different combinations of proximity SMoSes, such as in Zheng et al. [2018],Wang et al. [2019],Zheng et al. [2019] and Cavadas et al. [2020]. It was verified that the crash frequencies estimated from the bivariate extreme value distributions are more accurate, in the sense that the estimated crash frequency based on bivariate models are much closer to real life events than univariate models. Since the validity of different SMoS are questionable under different circumstances and different SMoSes interpret conflicts slightly differently, hence for a traffic conflict if we take what is agreed by different proximity SMoS, we get a more precise description of the conflict.

1.2 Severity of traffic conflicts

The collisions according to the Hyden pyramid does not specify the consequence of a collision, but the (potential) consequence of a traffic conflict is as well of our interest. As this thesis is motivated by the EU Vision Zero project, which is aiming at reducing casualty of traffic accidents to 0. To properly define severity, we employ the concept of Traffic Conflict Technique (TCT), where the severity of conflicts are level curves that represent the combinations of measurements in temporal and consequential SMoSes.



The severity of a traffic conflict is a combination of the proximity and potential consequences. The TCT extends the Hyden pyramid by including the (potential) consequences of a collision. **By this characterization a collision that has low**

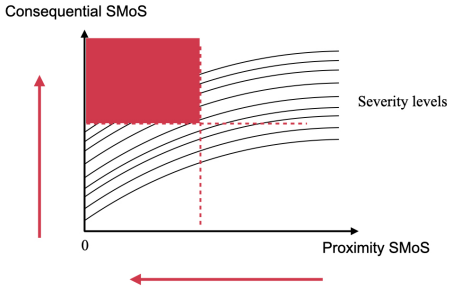
[1] for example, some proximity SMoSes indicate there is a collision while others do not. The selection bias can be reduced by improving data collection procedure, for example in Wang et al. [2019] the video footage were recorded from a drone above target site, which eliminates the selection bias due to the position of the cameras. More discussions of video footages can be found in Laureshyn [2010]

consequence may be ranked as less severe than a near-collision that has potentially high consequence.

Yet none of existing bivariate EVT models of two proximity SMoSes are incapable of predicting the severity of a conflict. As pointed out in Laureshyn et al. [2010], the proximity SMoS alone ignores the behaviours of road users, for example as two vehicles merge drivers will brake which result in decrease in consequential SMoS thus the severity of the conflict also decreases. One possible approach is to form a universal SMoS that accounts also for the behavioural aspect of road users. For example in Laureshyn et al. [2017] a SMoS extended Δv is introduced where braking is taken into consideration when measuring the Δv.[2]

Alternatively, Jonasson and Rootzén [2014] suggested that we can use bivariate extreme value models of a proximity SMoS and a consequential SMoS in their first paper regarding the bivariate modeling of SMoSes. There may exist certain dependence between proximity and consequence which is then captured by the joint distribution. This was experimented in Borsos [2021].

Attempt in Borsos (2021)



Borsos [2021] used a bivariate extreme value model called Peaks Over Threshold type 1 (POT1) to investigate the probability of severe conflicts, which are events located in the red coloured region in Figure 1.2. The conflicts of high severity, but not necessarily collisions are represented in the red region. We will define the risk to be the probability measure of the random vector $Z = (X, Y)$.

[2]similar approach can found in Bagdadi [2013] with the difference in the calculation of deceleration, where in Laureshyn et al. [2017] constant deceleration that represents two levels of braking were used while in Bagdadi [2013] the deceleration is the continuation of the derivative of the interpolated speed in terms of polynomials.

1.3 Partition of outcome space

Consider the probability space $(\Omega, \mathcal{F}, P)$ and a measurable function $X : \Omega \mapsto S \subset \mathbb{R}^+$. Moreover suppose that S is compact, then we can represent S as a finite union of open sets. In our case we may assume that $S = \cup_{i=1}^{n} S_i$ where S_i s are consecutive and disjoint intervals such that $S_i = [a_{i-1}, a_i)$. By construction we can partition the outcome space into finitely disjoint subsets: $\Omega = \cup_{i=1}^{n} \Omega_i$, where $\Omega_i = X^{-1}(S_i)$

We are going to partition the outcome space twice, with two distinct measurable functions X, Y. In our context we let Ω be the set of all traffic encounters, the two measurable functions X, Y represent the proximity and the (expected) consequential measurements of traffic conflicts, then we can partition the traffic encounters according to the two pyramids, such that

- Proximity: $\Omega = \cup_{i=1}^{n} \Omega_i$, where $\Omega_i = X^{-1}(D_i),\ X : \Omega \mapsto \boldsymbol{D} = \cup_{i=1}^{n} D_i$
- (Expected) consequence: $\Omega = \cup_{j=1}^{m} \tilde{\Omega}_j$, where $\tilde{\Omega}_j = Y^{-1}(S_j),\ Y : \Omega \mapsto \boldsymbol{S} = \cup_{j=1}^{m} S_j$

The partition $\Omega = \cup_{i=1}^{n} \Omega_i$ gives the Hyden pyramid (Figure 1)and the partition $\Omega = \cup_{j=1}^{m} \tilde{\Omega}_j$ can also be visualized as a pyramid:

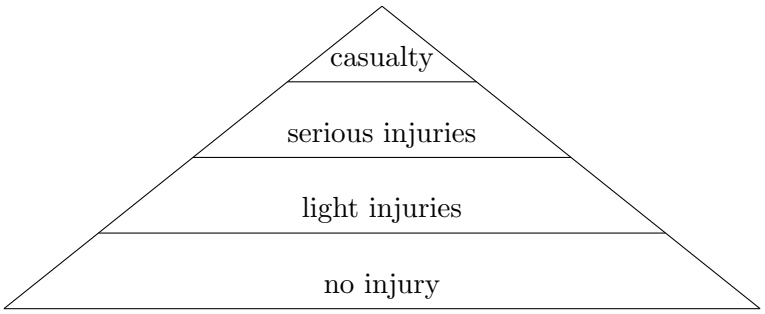


Figure 2: pyramid of consequence of (hypothetical) collisions

It is important to note that both $\boldsymbol{D}$ and $\boldsymbol{S}$ are unions of consecutive disjoint intervals on the positive real line. The intervals corresponding to the two pyramid summits are located towards the opposite ends. For example the set of collision is $\Omega_1 = X^{-1}(D_1)$ while the set of the most severe (hypothetical) collisions is $\tilde{\Omega}_m = Y^{-1}(S_m)$. By such construction it follows that the k th level of the proximity pyramid counting from the top is Ω_k; the k th level of the consequence pyramid counting from the bottom is $\tilde{\Omega}_k$. By such construction any rectangle in Figure 1.2 can be represented as a combination of levels of the two pyramids, as illustrated in the following Figure 3.

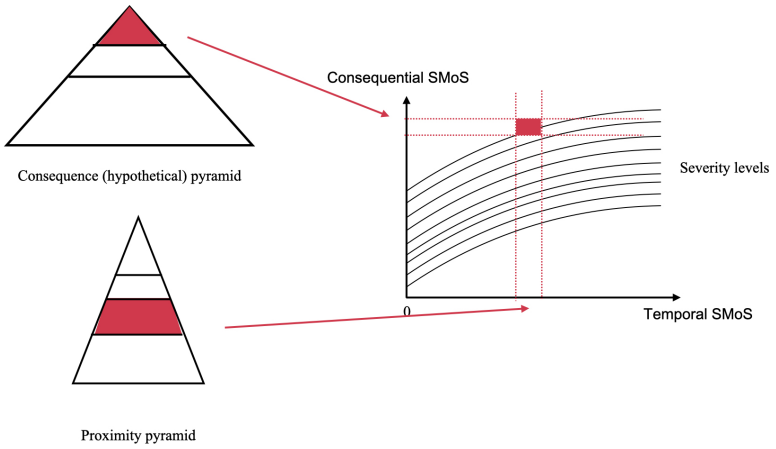


Figure 3: Connection between the pyramids and severity

1.4 Definition of risk

We start from defining the risk of an event formally:

Def 1.4.1 (Risk in TCT)**.** *Let Ω be collection of all traffic encounters, $(\Omega \times \Omega, \mathcal{F}, P)$ be the probability space, $X : \Omega \mapsto \boldsymbol{D} \overset{compact}{\subset} \mathbb{R}^+$, $Y : \Omega \mapsto \boldsymbol{S} \overset{compact}{\subset} \mathbb{R}^+$ be two continuous $\mathcal{F}$-measurable functions, $Z = (X, Y) : \Omega \times \Omega \mapsto \boldsymbol{S} \times \boldsymbol{D}$ to be a random vector. We define the risk of $\boldsymbol{Z} \in \boldsymbol{S} \times \boldsymbol{D}$ to be the probability measure of Z over $\boldsymbol{Z}$ i.e $P(Z \in \boldsymbol{Z})$.*

Consider the partitions $\cup_{i=1}^{n} \Omega_i$, $\cup_{j=1}^{m} \tilde{\Omega}_j$ of the outcome space Ω as discussed in previous Section 1.3. By such construction the intersection between any two arbitrary union $(\cup_i \Omega_i) \cap (\cup_j \tilde{\Omega}) \in \mathcal{F}$, thus we can apply the probability measure. This means for any combination of the levels of the two pyramids, there is a corresponding region in the TA/Speed plot (Figure 1.2) and we can compute the probability of the colored region $P(Z^{-1}(S \times D))$. By this definition an encounter is said to be of high risk if it has a combination of high consequential SMoS and low proximity SMoS. The collisions are concentrated on the consequential SMoS axis, thus the identification of collision severity becomes easier. For some severity level curve ξ, "collisions that are severer than level ξ" can be otherwise stated as "collisions that have consequential SMoS measurements greater than y, in which y is the intercept of ξ with the consequential SMoS axis. An illustration of the events is presented below (Figure 4):

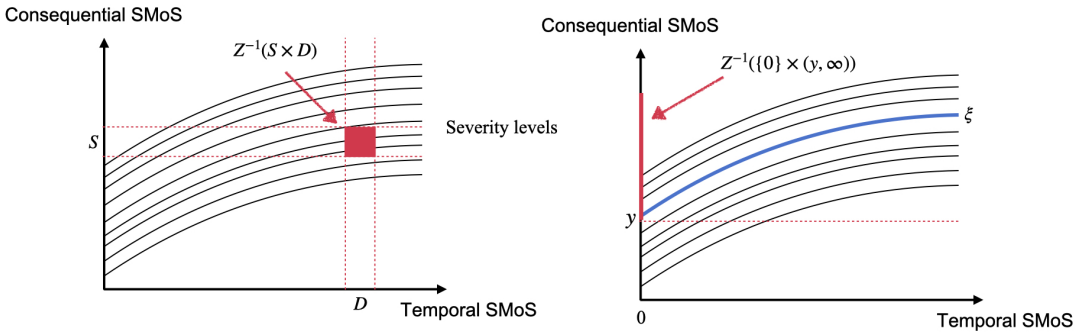


Figure 4: Left: Some traffic conflicts; Right: Collisions that are severer than ξ

We are interested in collisions that where one is injured seriously and we refer to the maximum abbreviated injury score (MAIS) for determining the injury levels. It is not possible to directly identify the corresponding MAIS level given the consequential SMoS. The relation between MAIS and the severity of a collision is probabilistic. The probability of reaching certain MAIS levels is affected by consequential SMoS as well as other factors, such as the hitting position, brands of vehicles etc, for which one can investigate by means of multinomial or ordinal logistic regressions.

We regard for simplicity that if given the consequential SMoS measurement of a collision is greater than 12 meters/second, then we consider it to be a collision where serious injury may occur. We choose $\Delta \geq 10$ according to the result from Gabauer and Gabler [2008] such that the probability of resulting in MAIS level 3 injury is lower than 0.1 for collision at $\Delta v \leq 10$. Therefore the collisions that of our interest is defined by $\{\omega \in \Omega : X^{-1}(0) \cap Y^{-1}((y, \infty)), y \geq 10\}$ and its risk:

$$P\big(\{\omega \in \Omega : X^{-1}(0) \cap Y^{-1}((y, \infty))\}\big) = P(X = 0, Y > y) \tag{1}$$

Looking at (1) we identify a problem immediately. By assumption both proximity and consequential SMoSes are continuous random variables, hence (1) will have value 0. This is apparently not the case in practice. To obtain a non-zero probability measure of collision events, we require the distribution function of the proximity SMoS to be supported on an interval that strictly includes 0 such that $\inf_{x \in \mathbb{R}}\{x : F(x) > 0\} =: x_F < 0$. Instead of (1), we make an extrapolation on the support of temporal SMoS and compute the following probability:

$$P\big(\{\omega \in \Omega : X^{-1}([x_F, 0]) \cap Y^{-1}((y, \infty))\}\big) = P(X \leq 0, Y > y) \tag{2}$$

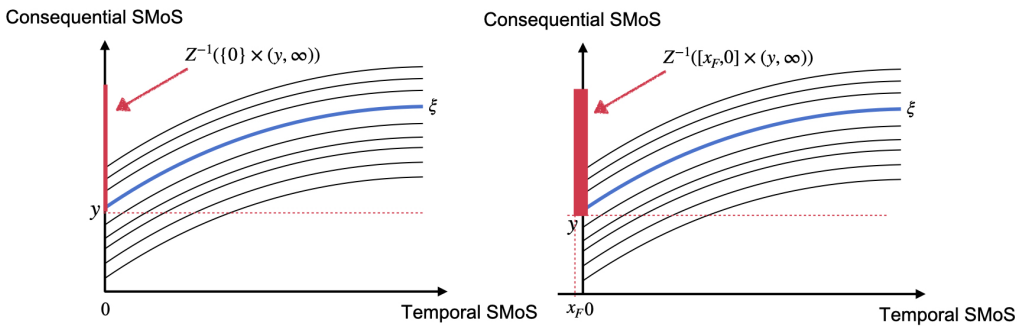


Figure 5: Before and after the extrapolation

The rest of the thesis is dedicated to the computation of (2), in particular we test two approaches: bivariate extreme value theory and copula. Intuitively severe collisions lie precisely in the region of high risk conflicts, as developed in Borsos [2021], so we may choose POT1 model. Under this circumstance we can use truncated BEV distribution to approximate the conditional tail distribution $P(X \leq 0, Y > y | X < u_1, Y > u_2)$. In order to apply this method it is required that the tails of two margins are towards the maximum of the their supports. Unfortunately this is not the case for severe collisions, thus it is necessary to make a transformation before fitting bivariate POT1 models. We will discuss the choice of transformation in Section 3.

Suppose now that $T(\cdot)$ is a decreasing transformation of X, then the probability $P(X \leq 0, Y > y)$ is equivalently stated as $P(T(X) \geq T(0), Y > y)$. The severe traffic conflicts separated by the thresholds [3]. The computation of (2) based on bivariate EVT approach is:

$$P(Y > y, T(X) > T(0)) = \bar{G}(T(0), y)\mathbf{1}\{y \geq u_2\} \tag{3}$$

where u_2 is the thresholds for the consequence SMoS. The advantage of POT1 models is that the marginal distributions of proximity and consequential SMoSes do not bother us as long as the thresholds u_1, u_2 are large enough. The compromise to be made is that we lose the flexibility of computing the probability for events being not part of the colored region in Figure 1.2. However most of targeted events for this project are well included in the colored region. The choice of thresholds is of great importance in POT1 approach, details will be discussed in Section 4.

[3] which is the red region in Figure (1.2)

The copula approach is more straight-forward. In theory if the correct marginal distributions of proximity and consequential SMoS are supplied, we can compute the risk for all conflicts. Suppose that F_1, F_2 are the underlying distribution of proximity and consequential SMoSes, C is some copula that exactly fits to the provided data, then we compute (2) by:

$$P(X \leq 0, Y > y) = P(X \leq 0) - P(X \leq 0, Y \leq y) = F_1(0) - C(F_1(0), F_2(y)) \quad (4)$$

In the copula approach, the challenge occurs in identifying suitable marginal distribution for proximity SMoS. For large number of data we can replace the exact distribution with empirical distribution for consequential SMoS. But this is not possible for distribution proximity SMoS. Since our data do not contain any collision, the requirement $x_F < 0$ will not be fulfilled. We discuss the technical parts of copula approach in Section 5.

The probability (1) is computed based on the two approaches mentioned above and is used in the estimation of collision frequency. Denote the number of collisions that have $\Delta v > y$ in a period Y to be N_T, then $N_T \sim Bin(n_T, P(X \leq 0, Y > y))$ where n_T is the number of traffic encounters in the given time period T and the expected number of such collision is equal to $E(N_T) = n_T \cdot P(X \leq 0, Y > y)$. Usually the data are collected from a short (relative to T) observational period t, in this case $N_T \sim Bin\big(\frac{T}{t} \cdot n_t, P(X \leq 0, Y > y)\big)$ and the expected value is changed accordingly:

$$E(N_T) = \frac{T}{t} \cdot n_t \cdot P(X \leq 0, Y > y) \quad (5)$$

In Section 1 we provided motivations and infrastructure for bivariate modeling. Section 2 contains the necessary bivariate extreme value theory for the POT1 approach. More mathematical details regarding EVT are presented in appendix. The actual modelling starts from Section 3 and ends at Section 5. In Section 3 we discuss the data preparation before the analysis. We present the implementation [4] and the results from POT1 models in Section 4. The theory and implementation of copula approach are included together in Section 5. At last we will reflect on the two approaches and suggest future researches on the modeling of collision risks in Section 6.

2 Extreme value theory

In single variate case there exist two ways in modeling the extremes of i.i.d random variables $X_1, \cdots, X_n$: Block Maxima (BM) and Peaks Over Threshold (POT). BM models the extremes using the limiting distribution of extreme order statistics $X_{(n)}$. This approach is based on the Fisher–Tippett–Gnedenko theorem, which was

[4] which is done in `R` R Core Team [2021]. The package `ggplot2` Wickham [2016] is used for plotting

generalized in de Haan [1970] as "there exist sequences of normalizing constants $\{a_n > 0\}_n, \{b_n \in \mathbb{R}\}_n$ such that $\lim_{n\to\infty} F^n(a_n x + b_n) \xrightarrow{d} G(x)$ for some non-degenerate distribution G." Moreover the original distribution F is said to belong to the domain of attraction of a generalized extreme value distribution G. Suppose the observations are divided into m blocks and $\hat{x}_1, \cdots, \hat{x}_m$ are the block maxima, then according to the theorem the maximal order statistics has GEV distribution with parameters estimated from the block maxima.

Peaks over threshold models the upper tail by filtering out observations that do not exceed a certain threshold i.e we are modeling the conditional random variable $X|X > u$. It was shown in section 3 of Pickands [1975] that the distribution of the upper tail $P(X \leq x|x > u)$ has a generalized pareto distribution as u goes to the right end point if and only if F belongs to the domain of attraction of an extreme value distribution G. The result is known as Pickands–Balkema–De Haan theorem, which together with Fisher–Tippett–Gnedenko theorem are often called the first and the second theorem of extreme value theory. We will discuss the Fisher–Tippett–Gnedenko theorem in more details, as it is also a cornerstone in multivariate extreme value theory.

For univariate distributions, the Fisher–Tippett–Gnedenko theorem (also known as the extremal type theorem) provides the complete characterization of the family of max-stable distributions using the three extremal type distributions. This applies even to higher dimensions, but the max-stable distributions in higher dimension are characterized somewhat more complicated. In Appendix we show that the non-degenerate limiting distribution of extreme order statistics must be max-stable.We start by stating one of the most important theorem in extreme value theory. A further step is to derive the expression for univariate max-stable distributions

Theorem 2.1 (Fisher–Tippett–Gnedenko)**.** *Let $\{F_n := F^n\}_n$ be the sequence of distribution functions of extreme order statistics. If $\exists\{\alpha_n > 0\}_n, \{\beta_n \in \mathbb{R}\}_n$ sequences of normalizing constants s.t $F_n(\alpha_n x + \beta_n) \to G$ where G is non-degenerate, then G belongs to one of the following three types:*

$$\begin{cases} \Phi_\alpha(x) = \exp\big(-x^{-\alpha}\big),\ x \geq 0\,,\ \alpha > 0 & (\textit{Fréchet}) \\ \Psi_\alpha(x) = \exp\big(-(-x)^{\alpha}\big), x \geq 0\,,\ \alpha > 0 & (\textit{inverse Weibull}) \\ \Lambda(x) = \exp -(e^{-x}), x \in \mathbb{R} & (\textit{Gumbel}) \end{cases} \tag{6}$$

Putting Lemma 7.5 and 7.6, we know that the non-degenerate limiting distribution of extreme order statistics is max-stable and satisfies the functional equation $G^s(\alpha(s) + \beta(s)) = G(x)\,, \forall s > 0$. By solving the equation we get the three extremal type distributions.

Remark 2.1.1. *The three extremal type distributions can be represented by the family called Generalized Extreme Value distributions (GEV)*

$$\left\{ G(x; \mu, \sigma, \gamma) = \exp\left(-\big((1 + \gamma\frac{x-\mu}{\sigma}\big)_+^{-1/\gamma}\right) : \mu, \gamma \in \mathbb{R}, \sigma \in \mathbb{R}^+ \right\}$$

By Theorem 6 (in Appendix) we know that if the non-degenerate limiting distribution of order statistics belongs to one of the extremal types,it is denoted as $F \in \mathcal{D}(G)$ i.e F belongs to the **Domain of Attraction** of (DoA) of G. The following lemma gives an equivalent definition for DoA.

Lemma 2.2. *Let F be a single variate distribution function, then $F^n(\alpha_n x + \beta_n) \to G(x)$ ($F \in \mathcal{D}(G)$)where G is non-degenerate iff*

$$\lim_{n\to\infty} n(1 - F(\alpha_n x + \beta_n)) \to -\ln G(x)$$

Proof. We rewrite F^n as $F^n = \big(1 - n^{-1} \cdot n(1-F)\big)^n$, then by a standard limit we have $\lim_{n\to\infty} F^n = \exp\big(-n(1-F)\big)$.

$$\lim_{n\to\infty} F^n(\alpha_n x + \beta_n) = \exp\big(-n(1 - F(\alpha_n x + \beta_n))\big)$$

Suppose $\lim_{n\to\infty} n(1 - F(\alpha_n x + \beta_n)) \to -\ln G(x)$, then

$$F^n(\alpha_n x + \beta_n) = \exp\big(-n(1 - F(\alpha_n x + \beta_n))\big) \to \exp\big(\ln(G)\big)$$

Conversely, suppose $F \in \mathcal{D}(G)$, then

$$\exp\big(-n(1 - F(\alpha_n x + \beta_n))\big) \to G(x) \implies \lim_{n\to\infty} n(1 - F(\alpha_n x + \beta_n)) \to -\ln G(x)$$

□

2.1 Approximation of upper tail distribution

An application of Lemma 2.2 is that we can show if $F \in \mathcal{D}(G)$, then the conditional tail distribution of $X \sim F$ is approximately Generalized Pareto distributed (GPD)[5]. Since $F(\alpha_n x + \beta_n)$ and $F(x)$ are of the same type,

$$\begin{aligned}
\lim_{n\to\infty} n(1 - F(\alpha_n x + \beta_n)) \to -\ln G(x) &\implies 1 - F(\alpha_n x + \beta_n) \to n^{-1}\big(1 + \gamma\frac{x-\mu}{\sigma}\big)_+^{-1/\gamma} \\
&\implies 1 - F(x) \simeq n^{-1}\big(1 + \gamma\frac{x-\mu}{\sigma}\big)_+^{-1/\gamma} \\
\implies P(X > x + u | X > u) = \frac{1 - F(x+u)}{1 - F(u)} &\simeq \left(\frac{1 + \gamma\frac{x+u-\mu}{\sigma}}{1 + \gamma\frac{u-\mu}{\sigma}}\right)_+^{-1/\gamma} = \left(1 + \gamma\frac{x}{\underbrace{\sigma + \gamma(u-\mu)}_{\sigma_*}}\right)^{-1/\gamma}
\end{aligned} \tag{7}$$

which we recognize to be the survival function of GPD. We have shown that $F \in \mathcal{D}(G) \implies$ the conditional tail distribution is approximately $GPD(\sigma_*, \gamma)$. The GPD is stable under the filtering-excesses operator s.t if $X \sim GPD(\sigma_*, \gamma)$, then $P(X > x + u | X > u)$ is again GPD. By writting out the conditional probability

[5]The converse also holds

we immediately obtained $P(X > x+u|X > u) = \left(1+\gamma\frac{x}{\sigma_*+\gamma u}\right)^{-1/\gamma}$, which is the survival function for $GPD(\sigma_*+\gamma u, \gamma)$. This property is used for selecting thresholds. At last we can obtain the unconditional tail distribution easily s.t

$$P(X \le x) = P(X \le x|X > u) \cdot P(X > u) = 1 - \eta_u\left(1+\gamma\frac{x}{\sigma+\gamma(u-\mu)}\right)^{-1/\gamma} \quad (8)$$

2.2 DoA and point processes

Suppose $X_1, \cdots, X_n$ are random samples from F and $F \in \mathcal{D}(G)$, G is a GEV. Consider a region $A \subset [0,1] \times [x_G, x^G] := E$. Define a simple point process $N_n(A) := \sum_{i=1}^n \mathbf{1}\{\frac{x_i-\beta_n}{\alpha_n} \in A\}$ [6], in which $\{\alpha_n\}_n, \{\beta_n\}_n$ are the sequences of normalizing constants as prescribed in Theorem 6. We shall establish the characterization of G from the distribution of the limiting point process $N_n(A)$. $N_n(A)$ is $Bin(n, p_n)$ distributed, for small probability p_n, it will converge in distribution to a Poisson distribution.

Lemma 2.3. *Suppose $X_n \sim Bin(n, p_n)$ and $n \cdot p_n \to \lambda < \infty$, then $\lim_{n\to\infty} X_n \xrightarrow{d} Y \sim Poi(\lambda)$.*

Proof.

$$\begin{aligned} P(X_n = k) &= \binom{n}{k}(1-p_n)^{n-k}p_n^k = \binom{n}{k}(1-np_n/p_n)^{n-k}(np_n/n)^k \\ &= \frac{n(n-1)\cdots(n-k+1)}{n^k} \cdot \frac{np_n}{k!} \cdot (1-np_n/n)^{n-k} \\ &\to \frac{\lambda^k}{k!} \cdot \exp(-\lambda) = P(Y = k) \ \forall k \in \mathbb{N} \text{ as } n \to \infty \end{aligned}$$

□

Theorem 2.4. *Let $X_1, \cdots, X_n$ be random samples from $F \sim \mathcal{D}(G)$, $G \sim GEV(\mu, \sigma, \gamma)$ and $N_n(A)$ be a point process defined as above. Moreover choose $A := [t_1, t_2] \times [x, x^G]$, then $N_n(A)$ converge in distribution to a non-homogeneous Poisson process with intensity measure $\Lambda(A) = (t_2 - t_1)\big(1+\gamma\frac{x-\mu}{\sigma}\big)_+^{-1/\gamma}$*

First we use Lemma 2.2, $F \in \mathcal{D}(G)$ implies

$$P(\frac{X-\alpha_n}{\beta_n} \le x) = F(\alpha_n x + \beta_n) \equiv -\ln(G) = \big(1+\gamma\frac{x-\mu}{\sigma}\big)_+^{-1/\gamma}$$

[6] i.e the number of points that lies in A

Intuitively the proof of Theorem 2.4 will follow from Lemma 2.3 and Lemma 2.2, but the convergence of the point process $N_n(A)$ distributed as $Bin(n, (t_2 - t_1)P(\frac{X-\alpha_n}{\beta_n} \leq x))$ to a non-homogeneous Poisson process with intensity function $\Lambda(A) = (t_2 - t_1)\big(1+\gamma\frac{x-\mu}{\sigma}\big)_+^{-1/\gamma}$ turns out to be more complicated. Instead of showing the convergence directly, one can show the convergence of Laplace functional, which uniquely determines the law of the point process (Proposition 3.5 Resnick [2008]). A rigorous proof of Theorem 2.4 is given in Proposition 3.21 Resnick [2008].

In the univariate case, if we let $A := [0,1] \times E \setminus (-\infty, x]$, then the event of no occurrence of a point $x_i > x\,, \forall i = 1, \cdots, n$ is equivalent to the event $N_n(A) = 0$. So

$$P(\frac{M_n - \alpha_n}{\beta_n} \leq x) \approx P(N(A) = 0) = \exp\bigg(- \big(1 + \gamma\frac{x-\mu}{\sigma}\big)^{-1/\gamma}\bigg)$$

2.3 Characterization of multivariate max-stable distributions

Discussion of multivariate max-stability in this section follows closely to A. A. and S. I. [1977],de Haan and Resnick [1977] and Resnick [2008].

Def 2.3.1 (Component-wise operations)**.** *Let $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^d$, then we define:*

1. $\boldsymbol{x} < \boldsymbol{y}$ *if* $x_i < y_i, \forall i = 1, \cdots, d$ *(Component-wise ordering).*
2. $\boldsymbol{x} \in [\boldsymbol{a}, \boldsymbol{b}]$ *if* $\boldsymbol{x} \in \cap_{i=1}^d \{x_i \in [a_i, b_i]\}$ *(Hyper rectangle).*
3. $\boldsymbol{x} \vee \boldsymbol{y} = (x_1 \vee y_1, \cdots, x_d \vee y_d)$*, where* $x_i \vee y_i := \max(x_i, y_i)$ *(Component-wise max operation).*

Let $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n$ be random samples from $\mathbb{R}^d$, then the component-wise maxima is defined by $\boldsymbol{M}_n := \max_{1 \leq j \leq n} \boldsymbol{X}_j = (M_n^{(1)}, \cdots, M_n^{(d)})$.

The multivariate max-stable distribution is defined similarly as in the univariate case, since the multivariate max-operation is component-wise. We provide a proof for conditions of univariate max-stable distribution in Appendix, which will be used as a definition for the multivariate case.

Def 2.3.2 (multivariate max-stability)**.** *A d dimensional distribution G is max-stable if there exists functions defined on $(0, \infty)$, $\alpha^{(i)}(s) > 0\,, \beta^{(i)}(s)$ $s.t\, \forall s > 0\,, 1 \leq i \leq d\,, G^s(\alpha^{(1)}(s)x_1 + \beta^{(1)}(s), \cdots, \alpha^{(d)}(s)x_i + \beta^{(d)}(s)) = G(\boldsymbol{x})$.*

Generally speaking the max-stable distribution is a subclass of **operator stable distributions**. Michael [1969] provided in-depth measure-theoretical and group-theoretical arguments for conditions in order for a distribution to be operator stable and demonstrated the theory on sum-stable distributions. A simplified version of Sharpe's result is that: For a sequence of i.i.d random variables $\{X_i\}_{i=1}^n, X_i \sim F$, suppose $\mathcal{A}$ is a function of $X_1, \cdots, X_n$, if F is $\mathcal{A}$-stable, then there exists an

automorphism of F H such that $H(F)$ is the distribution of $\mathcal{A}\{X_i\}_{i=1}^n$ and $H(F)$ is a non-degenerate distribution function that is different from F only up to a location-scale transformation. To obtain max-stable distributions we let

$$\mathcal{A} := \frac{\max_{1\leq i\leq n}(X_1,\cdots,X_n)-\mu_n}{\sigma_n},\quad H(\cdot) := \cdot^n$$

where $\{\mu_i\}_i, \{\sigma_i > 0\}_i$ are sequences of normalizing constants. Without the normalizing sequences $\max_{1\leq i\leq n}(X_1,\cdots,X_n)$ collapses to degenerate distribution with a point mass at the right end point of X.

The class of operator stable distributions belongs to be class of **operator infinite divisible distributions**. A non-degenerate distribution G is said to be $\mathcal{A}$-infinitely divisible if $\exists\{X_i\}_{i=1}^{\infty}$ sequence of i.i.d random variables s.t $X_i \sim F$ and H is an automorphism of F such that $H^{-1}(G) \stackrel{d}{=} F$. In particular we are interested in max-infinite divisible distributions (max i.d). The key aspect of characterizing multivariate max i.d distribution is the exponent measure $\mu(\cdot)$, which is shown in A. A. and S. I. [1977], a multivariate max i.d distribution takes the form:

$$F(\boldsymbol{x}) = \exp\left(-\mu(\mathbb{R}^d \setminus (-\infty, \boldsymbol{x}])\right) \tag{9}$$

(9) can be interpreted as the probability of having no occurrence in a d dimensional simplex given a non-homogeneous Poisson process. Extra restriction is imposed on the intensity measure μ.

Def 2.3.3 (exponent measure)**.** *A positive non-decreasing function $\mu : \mathcal{B}(\mathbb{R}^d) \mapsto \mathbb{R}^+$ is called an exponent measure if it satisfies:*

1. *$\mu(\mathbb{R}\times A) = \infty$, for any $A \subseteq \mathbb{R}^{d-1}$, i.e the exponent measure for any cube in $\mathbb{R}^d$ is infinte.*
2. *For any $i = 1,\cdots,d$, $\lim_{x_i\to\infty}\mu(\mathbb{R}^d\setminus(-\infty,x_1]\times\cdots\times(-\infty,x_i]\times\cdots\times(-\infty,x_d]) = 0$*

Since max-stable distribution is max i.d, the class of multivariate max-stable distributions also takes the form (9). We will first show that the class of multivariate max-stable distribution has max-stable (GEV margins). The next theorem is an extension of Theorem 7.4 in higher dimensions. The proof is analogous.

Theorem 2.5. *Suppose $(M_n^{(1)},\cdots,M_n^{(d)})$ is the component-wise maximum of i.i.d r.vt $\boldsymbol{X_1},\cdots,\boldsymbol{X_n}$. If $\forall i = 1,\cdots,d\,\exists$ sequences of normalizing constants $\{\alpha_n^{(i)}\}_n$, $\{\beta_n^{(i)}\}_n$ s.t $\frac{M_n^{(i)}-\beta_n^{(i)}}{\alpha_n^{(i)}} \xrightarrow{d} G_i$ non-degenerate, then G is max-stable.*

Def 2.3.4 (MEVD)**.** *A d dimensional max-stable distribution function is called Multivariate Extreme Value distribution if its margins are GEV distributed.*

Hence the class of MEVD takes the form (9). Extra restriction on the exponent measure $\mu(\cdot)$ is needed to preserve multivariate max-stability, s.t $G^s(\boldsymbol{\alpha}(s)\boldsymbol{x}+\boldsymbol{\beta}(s)) = G(\boldsymbol{x})$ holds for some vector valued function $\boldsymbol{\alpha}(s), \boldsymbol{\beta}(s)$. Fortunately we can transform any MEVD $G(\boldsymbol{x})$ to MEVD with unit Frechet margins $G_*(\boldsymbol{x})$, for which the max-stability holds with $G_*^s(s\boldsymbol{x}) = G_*(\boldsymbol{x})\,, \forall s > 0$. Essentially we have to extend Theorem 7.7 to higher dimensions using Theorem 2.5, such that:

Theorem 2.6. *Let G, G_* be d dimensional distributions in which $G_i\,, i = 1, \cdots, d$ are margins of G and G_* has unit Frechet margins, $\phi^{-1}(\boldsymbol{x}) := \big(-\frac{1}{\ln(G_1(x_1))}, \cdots, -\frac{1}{\ln(G_d(x_d))}\big)$, then $G(\boldsymbol{x}) = G_*(\phi^{-1}(\boldsymbol{x}))$ and $G_*(\boldsymbol{x}) = G^{-1}(\exp(-\boldsymbol{x}^{-1}))$. In particular G_* is MEVD $\iff$ G is MEVD.*

Assume that G is MEVD and $F \in \mathcal{D}(G)$, G is max-stable, let $V(\boldsymbol{x}) := \frac{1}{1-F(\boldsymbol{x})}$ and $F_*(\boldsymbol{x}) := F(V^{-1}(\boldsymbol{x}))$, then $F_* \in \mathcal{D}(G_*)$.

The consequence of Theorem 2.6 is that any MEVD can be represented by MEVD with unit Frechet margins s.t

$$\begin{aligned} G(x_1, \cdots, x_d) &= G_*\Big(\underbrace{\big(1+\gamma_1\frac{x_1-\mu_1}{\sigma_1}\big)^{1/\gamma_1}, \cdots, \big(1+\gamma_d\frac{x_d-\mu_d}{\sigma_d}\big)^{1/\gamma_d}}_{\tilde{\boldsymbol{x}}} \Big) \\ &= \exp\big(-\mu_*(\mathbb{R}^{d+} \setminus (\boldsymbol{0}, \tilde{\boldsymbol{x}}])\big) \end{aligned} \tag{10}$$

where $\mu_*(\cdot)$ is the exponent measure associated with G_*. Hence we can w.l.o.g use a MEVD with unit Frechet margins, whose max-stability is fulfilled if

$$\forall s > 0\,,\ \mu_*(\mathbb{R}^{d+} \setminus [0, \boldsymbol{x}]) = s \cdot \mu_*(\mathbb{R}^{d+} \setminus [0, s \cdot \boldsymbol{x}]) \tag{11}$$

(de Haan and Resnick [1977])

2.4 The dependence function

The main result of this part is Proposition 5.11 in Resnick [2008]. We will go through the proposition without formally proving it and state a popular characterization of 10 in two dimensions. The procedure is based on the point process in higher dimensions. We first extend Theorem 2.4 with Lemma 2.2 and Theorem 2.5, that a d dimensional distribution $F \in \mathcal{D}(G)$ for some MEVD G if and only if for the point process

$$\begin{aligned} N_n([0,1] \times E \setminus (-\infty, \boldsymbol{x}]) &= \sum_{i=1}^{n} \mathbf{1}\Big\{ \big(\frac{y_i^{(1)} - \beta_n^{(1)}}{\alpha_n^{(1)}}, \cdots, \frac{y_i^{(d)} - \beta_n^{(d)}}{\alpha_n^{(d)}}\big) > \boldsymbol{x} \Big\} \\ &\to N([0,1] \times E \setminus (-\infty, \boldsymbol{x}]) \end{aligned}$$

where $E := [\boldsymbol{x}_G, \boldsymbol{x}^G] \subseteq \mathbb{R}^d$ is the support of G, $N([0,1] \times E \setminus (-\infty, \boldsymbol{x}])$ is a d dimensional non-homogeneous Poisson process with intensity function $\mu(E \setminus (-\infty, \boldsymbol{x}])$.

For MEVD with unit Frechet margins, we have $E = \mathbb{R}^{d+} \setminus \{0\}$, $\beta_n^{(1)} \equiv \cdots \equiv \beta_n^{(d)} \equiv 0$, $\alpha_n^{(1)} =, \cdots, \alpha_n^{(d)} = n$.

For the intensity function $\mu_*(A)$ of a non-homogeneous Poisson process on A, we use the notation $\mu_*(\mathrm{d}a)$ to denote the infinitesimal intensity measure s.t $\mu_*(A) = \int_{a \in A} \mu_*(\mathrm{d}a)$. When μ_* is also the exponent measure associated with G_*, by Definition 2.3.3 μ_* is equal to 0 for any $x_i = \infty$, thus the exponent measure is concentrated on a compact subset $\mathcal{N} \subset E$. Take $\|\cdot\|$ to be a norm in $\mathbb{R}^d$ and define $\mathcal{N} := \{\boldsymbol{x} \in E : \|\boldsymbol{x}\| \leq c\}$ for some finite constant c. As all norms on finite dimensional Euclidean space are equivalent, we can w.l.o.g define $\mathcal{N} := \{\boldsymbol{x} \in E : \|\boldsymbol{x}\| \leq 1\}$

Moreover the exponent measure μ_* for any cube in $\mathbb{R}^d$ is infinite. Meaning that $G_*(\boldsymbol{x}) = 0$, $, \boldsymbol{x} \in B_r(0) := \{\boldsymbol{x} : \|x\| < r, r > 0\}$ for some arbitrary neighbourhood of the origin. We are only interested in the region $\mathcal{N} \cap B_r(0)^c$, for which $0 < G_*(\boldsymbol{x}) < 1$ and its the expression for $x \in int(E)$ thus becomes:

$$G_*(\boldsymbol{x}) = \exp\big(-\mu_*(\mathbb{R}^{d+} \setminus [\boldsymbol{0}, \boldsymbol{x}])\big) = \exp\big(-\int_{\mathcal{N} \cap B_r(0)^c} \mu(\mathrm{d}\boldsymbol{x})\big) \tag{12}$$

Furthermore there exists a subset $\mathcal{N} \supseteq \tilde{\mathcal{N}} := \{\boldsymbol{x} \in E : \boldsymbol{x}/\|\boldsymbol{x}\| = 1\}$, we can define a new measure S on $\tilde{\mathcal{N}}$ s.t for any $A \subset \tilde{\mathcal{N}}$:

$$S(A) := \mu_*(\{\|\boldsymbol{x}\| > 1, \boldsymbol{x}/\|\boldsymbol{x}\| \in A\}) \tag{13}$$

Note that $\tilde{\mathcal{N}} = \{\boldsymbol{x} \in E : \|\boldsymbol{x}\| = 1\}$ is the unit sphere in $\mathbb{R}^{d+}$, thus $S(\tilde{\mathcal{N}}) = \mu_*(B_1(\boldsymbol{0})^c \cap \tilde{\mathcal{N}}) \leq \mu_*(B_1(\boldsymbol{0})^c \cap \mathcal{N}) < \infty$ by the definition of exponent measure. Next we will show that it is possible to represent (12) in terms of S.

To begin with let us consider a pseudo polar transformation $T : E \mapsto (0, \infty] \times \tilde{\mathcal{N}}$ s.t $T(\boldsymbol{x}) = (\|\boldsymbol{x}\|, \boldsymbol{x}/\|\boldsymbol{x}\|) = (r, \boldsymbol{a})$.

$$\begin{aligned} T(E \setminus (0, \boldsymbol{x}]) &= T(\{\boldsymbol{y} \in E : y^{(i)} > x^{(i)}\,, \text{for some}\, 1 \leq i \leq d\}) \\ &= \{(r, \boldsymbol{a}) \in (0, \infty] \times \tilde{\mathcal{N}} : r \cdot a^{(i)} > x^{(i)}\,, \text{for some}\, 1 \leq i \leq d\} \\ &= \{(r, \boldsymbol{a}) \in (0, \infty] \times \tilde{\mathcal{N}} : r > \bigwedge_{i=1}^{d} \frac{x^{(i)}}{a^{(i)}}\} \end{aligned} \tag{14}$$

and its pre-image is defined as:

$$T^{-1}(r, \boldsymbol{a}) = \{\boldsymbol{y} \in E : \|\boldsymbol{y}\| > r, \boldsymbol{y}/\|\boldsymbol{y}\| \in \tilde{\mathcal{N}}\} \tag{15}$$

Now we can establish the relation between exponent measure μ_* and the measure

S:

$$\begin{aligned}
\mu_* \circ T^{-1}(r, \boldsymbol{a}) &= \mu_*(\{\boldsymbol{y} \in E : \|\boldsymbol{y}\| > r, \boldsymbol{y}/\|\boldsymbol{y}\| \in \tilde{\mathcal{N}}\}) \\
&= r^{-1}\mu_*(\{r^{-1}\boldsymbol{y} \in E : \|\boldsymbol{y}\| > r, \boldsymbol{y}/\|\boldsymbol{y}\| \in \tilde{\mathcal{N}}\}) \qquad (16) \\
&= r^{-1}\mu_*(\{r^{-1}\boldsymbol{y} \in E : r^{-1}\|\boldsymbol{y}\| > 1, r^{-1}\boldsymbol{y}/\|r^{-1}\boldsymbol{y}\| \in \tilde{\mathcal{N}}\}) \qquad (17) \\
&= r^{-1}\mu_*(\{\boldsymbol{x} \in E : \|\boldsymbol{x}\| > 1, \boldsymbol{x}/\|\boldsymbol{x}\| \in \tilde{\mathcal{N}}\}) = r^{-1}S(\tilde{\mathcal{N}}) \qquad (18)
\end{aligned}$$

(16) is due to (11) and we obtain the infinitesimal measure:

$$\mu_* \circ T^{-1}(\mathrm{d}r, \mathrm{d}\boldsymbol{a}) = r^{-2}\mathrm{d}rS(\mathrm{d}\boldsymbol{a}) \tag{19}$$

Next we combine (12), (14), (18) and (19) to arrive at the characterization of MGEV with unit Frechet margins:

$$\begin{aligned}
G_*(\boldsymbol{x}) &= \exp\big(-\mu_*(E \setminus (\boldsymbol{0}, \boldsymbol{x}])\big) \\
&= \exp\big(-\mu_* \circ T^{-1} \circ T(E \setminus (\boldsymbol{0}, \boldsymbol{x}])\big) \\
&= \exp\Big(-\int_{T(E\setminus(\boldsymbol{0},\boldsymbol{x}])} \mu_* \circ T^{-1}(\mathrm{d}r, \mathrm{d}\boldsymbol{a})\Big) \\
&= \exp\Big(-\int_{\tilde{\mathcal{N}}}\int_{r>\bigwedge_{i=1}^d \frac{x^{(i)}}{a^{(i)}}} r^{-2}\mathrm{d}rS(\mathrm{d}\boldsymbol{a})\Big) \\
&= \exp\Big(-\int_{\tilde{\mathcal{N}}} \bigvee_{i=1}^d \frac{a^{(i)}}{x^{(i)}} S(\mathrm{d}\boldsymbol{a})\Big) \qquad (20)
\end{aligned}$$

The only requirement for the measure S is

$$\int_{\tilde{\mathcal{N}}} a^{(i)} S(\mathrm{d}\boldsymbol{a}) = 1, \ , \forall 1 \le i \le d \tag{21}$$

so that $G_*(\boldsymbol{x})$, has unit Frechet margins:

$$G(\infty, \cdots, x^{(i)}, \infty) = \exp\left(-\int_{\tilde{\mathcal{N}}} \frac{a^{(i)}}{x^{(i)}} S(\mathrm{d}\boldsymbol{a})\right) = \exp\left(-1/x^{(i)}\right) \tag{22}$$

In the bivariate case, we may choose $\|\boldsymbol{x}\| = x + y$. Thus the set $\tilde{\mathcal{N}}$ becomes $\{\boldsymbol{w} : \sum_{i=1}^2 w^{(i)} = 1\}$ and $\boldsymbol{x}/\|\boldsymbol{x}\| = \left(\frac{x}{x+y}, \frac{y}{x+y}\right)$ and (20) instead becomes:

$$\begin{aligned}
G_*(\boldsymbol{x}) &= \exp\left(-\int_0^1 \max\left((1-w)/x, w/y\right) S(\mathrm{d}w)\right) \\
&= \exp\left(-\frac{x+y}{xy}\int_0^1 \max\left((1-w)\frac{y}{x+y}, w\frac{x}{x+y}\right) S(\mathrm{d}w)\right) \\
&= \exp\left(\frac{x+y}{xy} \cdot A(q)\right) \qquad (23)
\end{aligned}$$

where $\mu_*(\mathbb{R}^{2+} \setminus (0,x] \times (0,y]) = (x^{-1} + y^{-1}) \cdot A(q)$ is a popular form of exponent measure and

$$A(q) = \int_0^1 \max\left((1-w)q, w(1-q)\right) S(\mathrm{d}w)\,,\ q = y/(x+y)$$

is called the Pickand dependence function Pickands [2010]. Based on (22) we can derive the properties of $A(q)$.

$$A(0) = \int_0^1 wS(\mathrm{d}w) = \int_0^1 (1-w)S(\mathrm{d}w) = A(1) = 1 \tag{24}$$

Since $w, q \in [0,1]$, we can use (22) again to establish the lower and upper bound for $A(q)$, s.t

$$\max(w, 1-w) \leq A(q) \leq 1 \tag{25}$$

The last but also the most important property of $A(q)$ is the convexity. This is due to the positiveness of the density of (23), which is fulfilled if $A''(q) \geq 0$. To sum up briefly, $A(q)$ is the class of function that is convex in $[0,1]$ s.t the boundary conditions (24), (25) are satisfied. We gives some examples of MEVD model based on (23) in Table 1. These parametric models will be used in fitting the data in Section 4. We adopt the parametrization in accordance with the manual of Stephenson [2002] (`evd` package) where the dependence functions and their properties are well summarized from the original papers.

Table 1: Some parametric models of dependence functions based on BEV with unit Frechet margins (23), if $x_i \sim GEV(\mu_i, \sigma_i, \gamma_i)$, then $y_i = \big(1 + \gamma_i \frac{x_i - \mu_i}{\sigma_i}\big)^{1/\gamma_i}$

Model	Distribution function	Independence	complete dependence
Husler-Reiss	$G_*(y_1, y_2) = \exp\Big(- y_1 \cdot \Phi\big(r^{-1} + 0.5 \cdot r \log \frac{y_1}{y_2}\big) - y_2 \cdot \Phi\big(r^{-1} + 0.5 \cdot r \log \frac{y_2}{y_1}\big)\Big)$	$r \to 0$	$r = \infty$
Logistic	$G_*(y_1, y_2) = \exp\Big(- \big(y_1^{1/r} + y_2^{1/r}\big)^r\Big)$	$r = 1$	$r \to 0$
Asymmetric logistic	$G_*(y_1, y_2) = \exp\Big(- (1-\alpha)y_1 - (1-\beta)y_2 - \big((\alpha y_1)^{1/r} + (\beta y_2)^{1/r}\big)^r\Big)$	$r = 1$ $\alpha = 0$ or $\beta = 0$	α, β fixed $r \to 0$
Negative logistic	$G_*(y_1, y_2) = \exp\Big(- y_1 - y_2 + \big(y_1^{-r} + y_2^{-r}\big)^{-1/r}\Big)$	$r \to 0$	$r \to \infty$
Negative asymmetric logistic	$G_*(y_1, y_2) = \exp\Big(- y_1 - y_2 + \big((\alpha y_1)^{-r} + (\beta y2)^{-r}\big)^{-1/r}\Big)$	one of $r, \alpha, \beta \to 0$	α, β fixed $r \to \infty$
Bilogistic	$G_*(y_1, y_2) = \exp\Big(- y_1 q^{1-\alpha} - y_2 (1-q)^{1-\beta}\Big)$	$\alpha = \beta = 1$	One of $\alpha, \beta \to 0$ when the other is fixed
Negative Bilogistic	$G_*(y_1, y_2) = \exp\Big(- y_1 - y_2 + y_1 q^{1+\alpha} + y_2 (1-q)^{1+\beta}\Big)$	Either one or both of $\alpha, \beta \to \infty$	Either one or both of $\alpha, \beta \to 0$
Coles-Tawn	$G_*(y_1, y_2) = \exp\Big(- y_1\big(1 - beta(q; \alpha+1, \beta)\big) - y_2 \cdot beta(q; \alpha, \beta+1)\Big)$	Either one or both of $\alpha, \beta \to 0$	Either one or both of $\alpha, \beta \to \infty$

2.5 Statistical models based on MGEV

There are three statistical models for multivariate EV analysis, which are component-wise block maxima, POT1 and type 2 peaks over threshold (POT2). The component-wise BM method takes the r largest order statistics component-wisely and approximate $(M_r^{(1)}, \cdots, M_r^{(d)})$ with (23). The POT1 model is a analogue of single variate peaks over threshold with exceedances in all margins. The model including the probability distribution and inference were discussed by for example (Joe et al. [1992]). It can be shown that the exceedances in both margins $X, Y|X > u_1, Y > u_2$ are approximated by a truncated version of (12). The POT2 is more general than POT1, as it deals with exceedances in at least one margin. Tajvidi [1995] showed that the limiting distribution of $X, Y|(X, Y) \not\leq (u_1, u_2)$ is bivariate generalized Pareto.

We will show that the joint tail distribution is approximately BEVD (Table 1). The procedure follows from chapter 8.3 of (Beirlant et al. [2004]) and will involve convergence of copulas, which is discussed in Section 5. Suppose a r.vt (X_1, X_2) has marginal distributions that belong to the DoA of GEV and joint distribution F, then according to discussion in Section 2.1, the unconditional tail distribution of the margins satisfy (8), which we denote them as F_1, F_2. The next Lemma is useful for transforming r.vs to unit Frechet distributed.

Lemma 2.7. *Let $X \sim F$ be a continuous random variable, then $-\frac{1}{\ln(F(X))} \sim G_*(x)$ where $G_*(x) = \exp(-x^{-1})$ is unit Frechet distributed.*

Proof. Let $X \sim F$ and $G_*(x) = \exp(-x^{-1})$, then $G_*^{-1}(u) = \frac{-1}{\ln(u)}$.

By theorem 5.1 we know that $F(X) \sim Un[0, 1]$ and $G_*^{-1}(U) \sim G_*$, so $\frac{-1}{\ln(F(X))} \sim G_*$ □

The first step is to transform all margins to unit Frechet. Suppose $V(\boldsymbol{x})$ and $F_*(\boldsymbol{x})$ are defined as in Theorem 2.6, the unconditional joint tail distribution $F(\boldsymbol{x})$ is equal to:

$$F(\boldsymbol{x}) = F_*\big(-\frac{1}{\ln(F_1(x_1))}, \cdots, -\frac{1}{\ln(F_d(x_d))}\big), \ \forall \boldsymbol{x} > \boldsymbol{u} \tag{26}$$

where the marginal tail distribution follows (8). Since $F \in \mathcal{D}(G)$, by Theorem 2.6 $F_* \in \mathcal{D}(G_*)$. Let $z_i := -\frac{1}{\ln(F_i(x_i))}$, then $F_*(\boldsymbol{z})$ has unit Frechet margins (by Lemma 2.7). Next we will show $F_*(\boldsymbol{z}) \approx G_*(\boldsymbol{z})$ for large $\boldsymbol{z}$.

Let $C_{F_*}(\boldsymbol{u}) := F_*(\boldsymbol{z})$ be a copula of F_*, by Theorem 5.2:

$$F_*(\boldsymbol{z}) = C_{F_*}(\exp(-\boldsymbol{z}^{-1})) = C(\boldsymbol{u}) = F_*(-\frac{1}{\ln(\boldsymbol{u})}) \tag{27}$$

Let $\tilde{\boldsymbol{u}} := F(n\boldsymbol{z})$, then $\tilde{\boldsymbol{u}} = \exp(-1/(n\boldsymbol{z})) = \boldsymbol{u}^{1/n}$

$$F^n(n\boldsymbol{z}) = C_{F_*^n}(\tilde{\boldsymbol{u}}) = F_*^n\big(-\frac{n}{\ln(\boldsymbol{u})}\big) = C_{F_*}^n(\boldsymbol{u}^{1/n}) \tag{28}$$

$F_* \in \mathcal{D}(G) \implies \forall s > 0,\ F^s(s\boldsymbol{z}) \to G_*(\boldsymbol{z}) \implies C_{F_*^n}(\tilde{\boldsymbol{u}}) = C_{F_*}^s(\boldsymbol{u}^{1/s}) \to C_{G_*}(\boldsymbol{u})$, where C_{G_*} is EV copula that satisfies the relation $C^s(\boldsymbol{u}^{1/s}) = C(\boldsymbol{u})$[7].

$$C_{F_*}(\boldsymbol{u}^{1/s}) \to C_{G_*}^{1/s}(\boldsymbol{u}) = C_{G_*}(\boldsymbol{u}^{1/s}) \tag{29}$$

$s \to \infty$ in (29) corresponds the upper tail $\boldsymbol{z} \to \boldsymbol{z}^F$, hence we can establish

$$F(\boldsymbol{x}) \approx G_*(-\frac{1}{\ln(F_1(x_1))}, \cdots, -\frac{1}{\ln(F_d(x_d))}),\ \forall \boldsymbol{x} > \boldsymbol{u} \text{ and } \boldsymbol{u} \to \boldsymbol{x}^F \tag{30}$$

with $F_i(x_i) = 1 - \eta_i\left(1 + \gamma_i \frac{x_i}{\sigma_i + \gamma_i(u - \mu_i)}\right)^{-1/\gamma_i},\ i = 1, \cdots, d$

3 Preparing the Data

3.1 Description

The data was formerly used in the analysis of Laureshyn et al. [2017], Borsos et al. [2020] and Borsos [2021]. It consists of video footage of encounters at a non-signalized intersection where traffic conflicts occur between left turning vehicles and straight moving vehicles. The weight of each vehicle was approximately grouped by its appearance: whether it is a car, a minivan or a truck. Measurements were taken between 6:00 to 21:00 of two consecutive days. In total there were 1512 vehicles, 756 encounters. We divided the vehicles into two groups according to the type of maneuvers, because the distance is measured per encounter while the Δv is measured per vehicle. It is noteworthy that there exists one encounter whose proximity SMoS attains 0; this was due to the position of the camera, in fact collision did not happen. So we set the distance to 0.2 meters instead of 0 meters.

3.2 Two ways of measuring

The video footage was first processed by the software T-analyst T-A. Different measurements of proximity and consequential SMoS are recorded. Once when the distance between two vehicles were minimal; once at the last moment before the termination of a conflict (i.e one vehicle has left the colliding course, which is also defined as the post encroachment distance). The observations we obtained under minDistance measurements tend to have smaller values in proximity SMoS, but it does not always imply a near-collision since two vehicles may not be on the colliding course any longer. On the other hand, observations obtained under PETDistance tend to have larger values in both proximity and consequential SMoS which is more difficult to make inference on the collisions, but the results are more convincing from a practical point of view.

[7] see Section 5.2 for details of EV copula

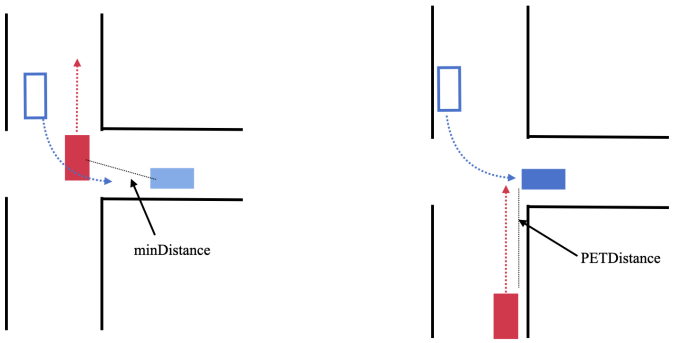


Figure 6: Illustration of the SMoS measurements at two moments, dotted arrows indicate the direction of the vehicles

We choose the SMoS pair to be **Distance** and $\boldsymbol{\Delta v}$. The Δv for the first vehicle is computed by

$$\Delta v_0 = \frac{m_1}{m_1 + m_2}\sqrt{v_1^2 + v_2^2 - 2v_1 v_2 \cdot \cos\alpha}$$

where v_1, v_2 are the speed; m_1, m_2 are the mass; α is the angle of collision. We assume the mass are the same for all vehicles in the same category, thus the only difference in Δv measurements between the two vehicles occur when they are not the same type class vehicles. In total there will be four data sets based on which we compute the probability of traffic conflicts:

Table 2: the name abbreviation for the four data set

measurement / maneuver	minDistance	PET Distance
Left turning	LminD	LPETD
Straight moving	SminD	SPETD

Note that the same measurement of consequential SMoS may result in different levels of injury. This is due to the difference in the contact point. The left-turning drivers are more vulnerable than straight-moving drivers. Given the same Δv, the chance of passenger injury in left-turning vehicles are greater. Hence we decide to make inference using only the left-turning data.

The plots of Distance and Δv against record time are listed in the Appendix (Figure 19,20). No obvious trends in time are spotted and thus we may assume the samples are i.i.d. The difference between two measurements are reflected on the scale of the axis.

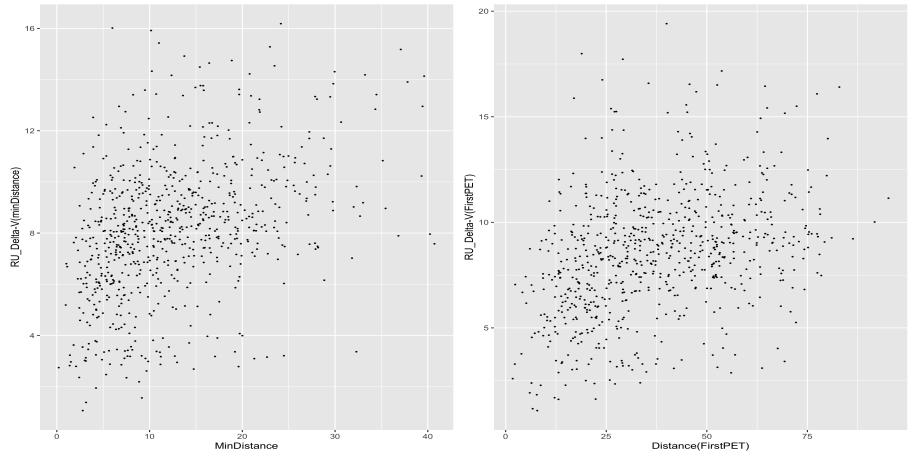


Figure 7: Scatter plots of the left turning data under two measurements, left: LminD; right: LPETD

3.3 Transformation of the Data

As mentioned in Section 1.4, the measurements of distance need transforming decreasingly. Moreover in order for (3) to be non-zero, we demand that $P(T(X) > T(0)|T(x) > T(u_1))$ to be non-zero. We know from Section 2 that if $T(X)$ belongs to the domain of attraction of some $GEV(\mu, \sigma, \gamma)$, $T(X)|T(X) > T(u_1)$ is asymptotically generalized Pareto distributed with scale $\tilde{\sigma} = \sigma + \gamma(T(u_1) - \mu)$ and shape γ as $T(u_1) \to T(x_F)$. The distribution $P(T(X) \leq x|T(X) > T(u_1))$ is supported on the positive real axis if $\gamma \geq 0$; it is supported on the interval $[0, \mu - T(u_1) - \sigma/\gamma]$. Hence it is required for the monotone decreasing transformation T s.t $T(0) \in [0, \mu - T(u_1) - \sigma/\gamma)$.

The simplest possible transformation T is the negation $T(X) = -X$. But in practice we found that under negation collision events do not always have non-zero probabilities in **LminD, LPETD** situations. Besides the negation, we have also considered the shifted reciprocal transformation $T(X) = \frac{1}{X+\delta}$ for some constant δ. Under this transformation, we identified (3) to be

$$P(Y > y, X > T(0)) = P(Y > y, \frac{1}{X + \delta} > 1/\delta) \tag{31}$$

This transformation is adopted from Zheng et al. [2014b] and the parameter estimation is simplified. In the original approach, the threshold and δ was treated as parameters. The estimation of the model was carried out by Monte Carlo method. Though such approach is slightly more computationally expensive, it helps to eliminate the subjectivity in selecting the threshold value and the value of δ. Moreover the distribution of the tail is different, the joint distribution will change accordingly. Considering the difficulty in characterizing POT1 models, we decided not to treat the threshold value u_1 and shift constant δ as parameters, but as prescribed constants instead. For example we choose $\delta := \min_i X_i$.

4 The Peaks Over Threshold Type 1 approach

In the study of collisions of certain severity we consider only the POT1 models. This is because bivariate block maxima is taken component-wisely. Even though bivariate block maxima is indeed suitable for modeling the dependence between two extreme order statistics, the events does not have to happen simultaneously. Hence it does not satisfy our need of computing the collision probability of certain severity. In this section we present the detailed implementation for the left turning vehicles (refer to Table 2), the implementation procedure for straight moving vehicles is identical.

4.1 The notations

We follow the same notation as in the introduction. In addition we define $\tilde{X} := T(X) = 1/(X+\delta)$ as the shifted reciprocal proximity SMoS. Based on our discussion in Section 2, the tail distribution of any bivariate distribution, which is $P(Y \leq y, \tilde{X} \leq \tilde{x})$, is approximately BEVD distributed. The scheme of computing the collision frequencies goes as follows: we fit all models presented in Table 1 to the chosen tail region by maximizing the censored log-likelihood and select the best model(s) according to likelihood ratio test, AIC and confidence interval of collision probabilities. Then we will compute the probability (1) and expected frequency (5) using the fittest model(s).

4.2 Selection of thresholds

We begin the analysis from threshold selection. To achieve a balance of bias and variance of the parameter estimation, $T(u_1) := \tilde{u}_1, u_2$ should be large enough such that the conditional marginal distribution of $P(Y > y, \tilde{X} > 1/\delta)$ can be approximated by GPD while the semi-open rectangular region $[u_1, \infty) \times [u_2, \infty)$ (which corresponds to the tail of a bivariate distribution) contains a decent amount of observations. Such semi-open rectangular region in which BEVD approximates tail distribution well is obtained by selecting the thresholds for the two margins separately, each as in the setting of univariate peaks over threshold. The package `extRemes` Gilleland and Katz [2016] is used univariate analysis.

The threshold selection in single variate POT is aided by the mean residual plot and the threshold range plot. Both plots exploit the stability of GPD s.t

$$\forall u > 0\,,\, X \sim GPD(\sigma_0, \gamma_0) \implies X - u | X > u \sim GPD(\sigma_1, \gamma_1)\,,\, \sigma_1 = \sigma_0 + \gamma u\,,\, \gamma_1 = \gamma_0$$

For the mean residual plot, the first moment $E(X - u|X > u) = \frac{\sigma+\gamma u}{1-\gamma}$, which in theory is linear in threshold u. We plot its empirical version, that is,

$$\Big\{(u, n_u^{-1} \sum_{i=1}^{n} (x_i - u)_+) : u \in I \subset \mathbb{R}\Big\}$$

in which n_u denotes the number observations that are greater than u. We should select the threshold from a range where the mean residual plot appears to be linear. For the threshold range plot, we plot $\sigma_1 - \gamma_0 u$ and γ_1 against u, which should be a constant. In practice the parameters are replaced with their maximum likelihood estimates, for which we can also add confidence intervals. We should select the thresholds as large as the variances of the estimates are acceptable.

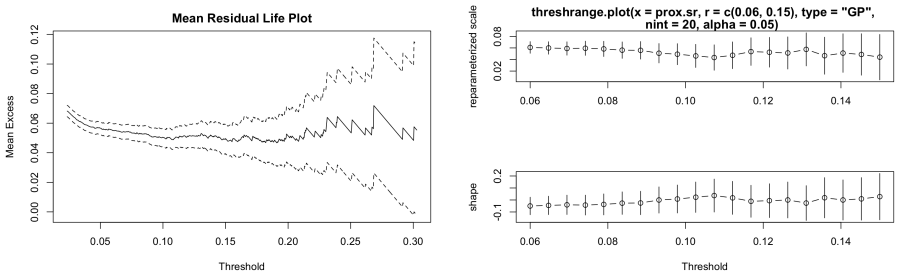


Figure 8: Threshold selection for the proximity SMoS in LminD

We combine both mean residual plot and threshold range plot for threshold selection. From the mean residual plot we see that the value for a good threshold $\tilde{u}_1$ (under LminD) is between 0.06 and 0.15 as the mean residual is approximately linear in this range. Then we turn to the threshold range plot to locate the threshold more precisely. The threshold range plot plots the width of the confidence intervals when a POT model is fitted to the data at some given thresholds. The principle is to choose the threshold along the threshold range plot as large as possible, however we should also bear in mind that the variance may affect the fitting of bivariate model later. According to Figure 8 we select $\tilde{u_1}$ to be 0.1 for LminD.

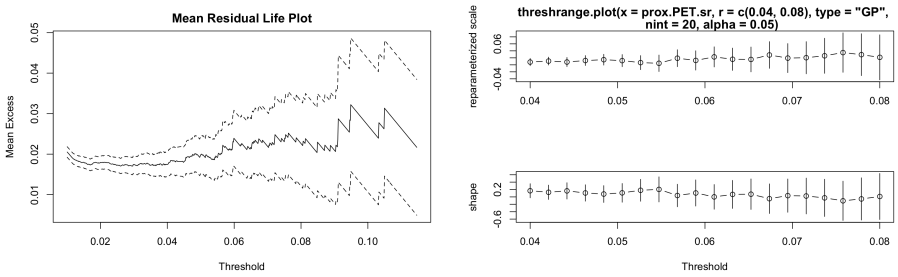


Figure 9: Threshold selection for the proximity SMoS in LPETD

$\tilde{u}_1$ LPETD: A threshold that yields acceptable variance is chosen between 0.04 and 0.06. Combining with the threshold range plot and the number of exceedences in two dimension, we choose $\tilde{u}_1$ to be 0.04.

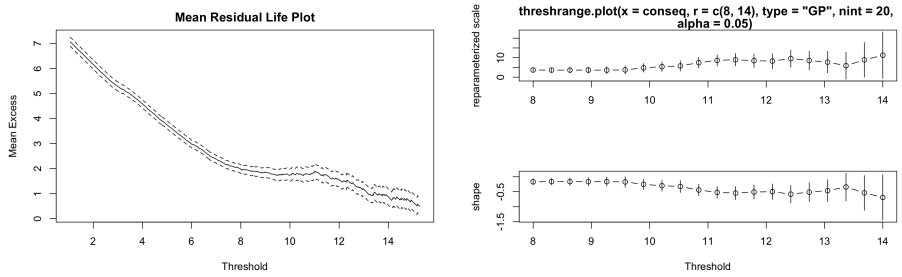


Figure 10: Threshold selection for the consequential SMoS in LminD

u_2 LminD: threshold that yields acceptable variance is chosen between 2 and 10, combining with the threshold range plot and the number of exceedences in two dimension, we choose u_2 to be 10.

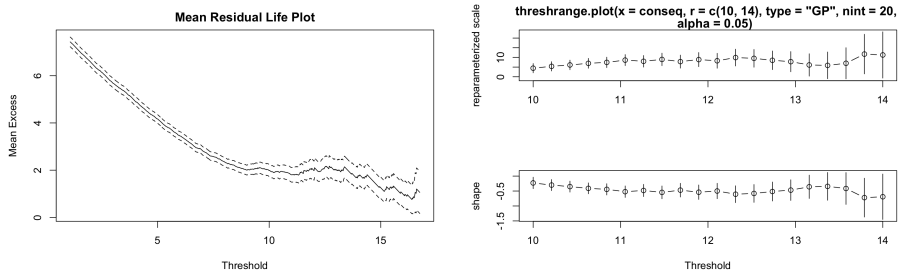


Figure 11: Threshold selection for the consequential SMoS in LPETD

u_2 LPETD: A threshold that yields acceptable variance is chosen between 10 and 14, combining with the threshold range plot and the number of exceedences in two dimension, we choose u_2 to be 10. The exceedences over $\tilde{u}_1, u_2$ for LminD and LPETD are plotted below:

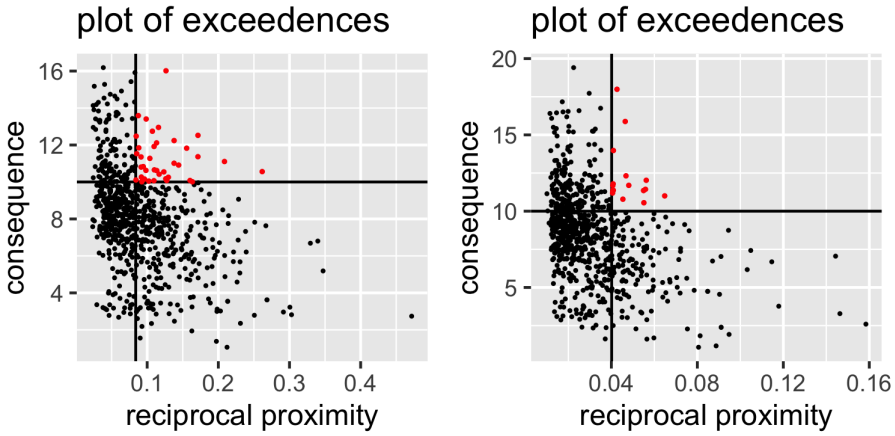


Figure 12: Type 1 exceedences (Left:LminD; Right:LPETD)

4.3 Parameter estimation

We use Full Maximum Likelihood (FML) to estimate the parameters[8]. The log-likelihood of tail distribution $P(\tilde{X} \leq x, Y \leq y), \forall (x,y) \geq (\tilde{u}_1, u_2)$ censors the observations below either thresholds. When applying the FML to distributions from Table 1, we should adjust the observations that do not exceed both thresholds accordingly. Suppose G is a distribution from Table 1, $l(\boldsymbol{\theta}, x, y) = \ln(g(x, y|\boldsymbol{\theta}))$, then we have the censored log-likelihood

$$l(\boldsymbol{\theta}, x, y) = \begin{cases} \ln \dfrac{\partial G_*}{\partial x \partial y}|_{(x,y)}, & x > \tilde{u}_1, y > u_2 \\ \ln \dfrac{\partial G_*}{\partial y}|_{(\tilde{u}_1,y)}, & x \leq \tilde{u}_1, y > u_2 \\ \ln \dfrac{\partial G_*}{\partial x}|_{(x,u_2)}, & x > \tilde{u}_1, y \leq u_2 \\ G_*(\tilde{u}_1, u_2), & x \leq \tilde{u}_1, y > u_2 \end{cases} \tag{32}$$

Table 3: Estimates of model parameters (LminD)

Model	scale1	shape1	scale2	shape2	dep	asy1	asy2	alpha	beta
neglog	0.053	-0.022	2.218	-0.264	0.051	NA	NA	NA	NA
alog	0.053	-0.031	2.149	-0.228	0.999	0.752	0.776	NA	NA
aneglog	0.054	-0.032	2.147	-0.227	0.075	0.999	0.401	NA	NA
bilog	0.054	-0.022	2.149	-0.237	NA	NA	NA	0.816	0.998
negbilog	0.054	-0.031	2.151	-0.229	NA	NA	NA	16.611	16.634
hr	0.054	-0.031	2.151	-0.229	0.201	NA	NA	NA	NA
ct	0.057	0.016	2.211	-0.143	NA	NA	NA	0.002	0.329
log	0.060	-0.052	1.732	-0.122	0.999	NA	NA	NA	NA

Table 4: Estimates of model parameters (LPETD)

	scale1	shape1	scale2	shape2	dep	asy1	asy2	alpha	beta
ct	0.015	0.189	2.251	0.033	NA	NA	NA	0.128	0.002
neglog	0.015	0.158	2.167	-0.091	0.064	NA	NA	NA	NA
negbilog	0.015	0.164	2.148	-0.070	NA	NA	NA	6.639	7.647
hr	0.015	0.155	2.172	-0.094	0.265	NA	NA	NA	NA
aneglog	0.015	0.135	2.148	-0.056	0.050	0.814	0.813	NA	NA
alog	0.015	0.163	2.161	-0.084	1.000	0.797	0.841	NA	NA
log	0.016	0.161	1.932	-0.188	0.999	NA	NA	NA	NA
bilog	0.017	0.161	2.170	-0.052	NA	NA	NA	0.999	0.832

The FML estimates of dependence parameters in the models given by Table 1 suggest that proximity SMoS (transformed) and consequential SMoS are asymptotically independent in the upper tail. Meaning that the consequential SMoS does not depend on the collisions in serious traffic conflicts. But there are still differences between bivariate POT1 models and the product of two independent GPD margins, as we have have used FML to fit the bivariate models.

[8]See Appendix for details of FML.

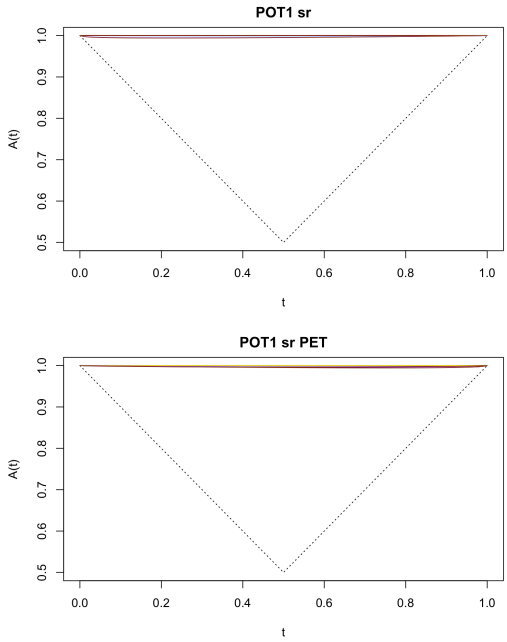


Figure 13: The dependence function of models in Table 1 estimated from the data

4.4 Model selection

In all eight models provided by Table 1, the logistic model is nested in the asymmetric logistic model; negative logistic model is nested in asymmetric negative logistic model (when the $\alpha = \beta = 1$). For near independence we may regard logistic model and negative logistic models being nested in bilogistic and negative bilogistic models. The first step in model selection is to determine whether more complicated models result in a significant increase of the likelihood. This is done by the likelihood ratio test.

Table 5: Likelihood ratio test for nested models in LminD

Test	Deviance statistics	Critical value	Result
log against alog	5.63	5.99	do not reject log
log against bilog	2.86	5.99	do not reject log
neglog against aneglog	0.35	5.99	do not reject neglog
neglog against negbilog	0.38	5.99	do not reject neglog

Table 6: Likelihood ratio test for nested models in LPETD

Test	Deviance statistics	Critical value	Result
log against alog	28.77	5.99	reject log
log against bilog	23.57	5.99	reject log
neglog against aneglog	0.46	5.99	do not reject neglog
neglog against negbilog	2.99	5.99	do not reject neglog

Based on the results from the likelihood ratio test, we discard the asymmet-

ric logistic, bilogistic, asymmetric negative logistic and negative bilogistic models in LminD; we discard logsitic, asymmetric negative logistic and negative bilogistic models in LPETD.

The next selection is based on whether the right end point of the proximity SMoS based on the FML estimation significantly includes the collisions (i.e $\tilde{x}_{G_1} > 1/\delta$), which can be done by constructing a confidence level. We will construct the confidence interval by Delta method.

$$I_{\bar{G}_1(\hat{\sigma}_1,\hat{\gamma}_1;1/\delta)} = \left[\bar{G}_1(\hat{\sigma}_1,\hat{\gamma}_1;q/\delta) \pm 1.96 \cdot \sqrt{\frac{\nabla\bar{G}_1(\hat{\sigma}_1,\hat{\gamma}_1;1/\delta)^T I^{-1}(\hat{\sigma}_1,\hat{\gamma}_1)\nabla\bar{G}_1(\hat{\sigma}_1,\hat{\gamma}_1);1.\delta}{n}}\right]$$

where

$$\bar{G}_1(\hat{\sigma}_1,\hat{\gamma}_1;x) = \left(1+\hat{\gamma}_1\frac{x-\tilde{u}_1}{\hat{\sigma}_1}\right)^{-1/\hat{\gamma}_1}\ ,\ \nabla\bar{G}_1 = \left(\frac{\partial\bar{G}_1}{\partial\sigma_1},\frac{\partial\bar{G}_1}{\partial\gamma_1}\right)$$

$$\frac{\partial\bar{G}_1}{\partial\sigma_1} = \frac{1/\delta-\tilde{u}_1}{\sigma_1^2}\cdot\left(1+\gamma_1\frac{1/\delta-\tilde{u}_1}{\sigma_1}\right)^{-\frac{\gamma_1+1}{\gamma_1}}$$

$$\frac{\partial\bar{G}_1}{\partial\gamma_1} = \left(1+\gamma_1\frac{1/\delta-\tilde{u}_1}{\sigma_1}\right)^{-\frac{\gamma_1+1}{\gamma_1}}\cdot\left(\left(1+\gamma_1\frac{1/\delta-\tilde{u}_1}{\sigma_1}\right)\cdot\ln\left(1+\gamma_1\frac{1/\delta-\tilde{u}_1}{\sigma_1}\right)-\gamma_1\cdot\frac{1/\delta-\tilde{u}_1}{\sigma_1}\right)\cdot\gamma_1^{-2}$$

Note that we do not construct confidence interval for the joint probability (3). As it requires additional treatments. The first and the most important reason is that the maximization of (32) using gradient-based algorithm such as Newton-Raphson does not always work. Thus the presence of covariance matrix is not guaranteed. As in our case some models are estimated through the Nelder-Mead algorithm where the covariance matrix is not available. We may obtain the variance of the estimates through bootstrapping. Another problem is the gradient of $G_*(\boldsymbol{x};\boldsymbol{\theta})$ w.r.t $\boldsymbol{\theta}$ requires more effort in terms of computation. However we may construct only confidence interval for the proximity margin.

Table 7: Criteria for model selection (LminD)

	df	AIC	CI of collision prob (lower)	CI of collision prob (upper)
hr	5	984.0221	0.0001002	0.0001003
log	5	990.0884	0.0001059	0.0001061
neglog	5	984.4095	0.0001067	0.0001068
bilog	6	989.2230	0.0001428	0.0001430
ct	6	995.1402	0.0007201	0.0007218

From Table 7 and 8 the confidence intervals of all collision probabilities do not include 0. More accurate confidence interval may be constructed based on profile

Table 8: Criteria for model selection (LPETD)

	df	AIC	CI of collision prob (lower)	CI of collision prob (upper)
hr	5	1381.887	0.0011168	0.0011255
neglog	5	1381.839	0.0011589	0.0011679
alog	7	1386.000	0.0014137	0.0014256
bilog	6	1389.187	0.0022946	0.0023194
ct	6	1392.099	0.0017176	0.0017318

likelihood. For details of CI for GPD probability. one may refer to Mach [2022], who have construct CI for GPD probabilities using profile likelihood for the same data set. Finally we compare the AIC for models that are left in LminD and LPETD, we decide to keep **Hustler-Reiss and negative logistic models** as the candidates for computing probabilities and expected frequencies.

4.5 Probabilities and frequencies

Suppose $\hat{\sigma}_1, \hat{\sigma}_2, \hat{\gamma}_1, \hat{\gamma}_2$ are the estimated parameters of the GPD margins using FML, then the probability (3) is computed by (30):

$$P(\tilde{X} > 1/\delta, Y > 12) = \bar{G}_*\bigg(-\frac{1}{\ln\left(G_1(1/\delta)\right)}, -\frac{1}{\ln\left(G_2(12)\right)}\bigg) \cdot \eta \tag{33}$$

where $G_i(\cdot) = 1 - \eta\big(1 + \hat{\gamma}_i \frac{\cdot - u_i}{\hat{\sigma}_i}\big)\big)^{-1/\hat{\gamma}_i}, i = 1, 2$. The expected crash frequency is computed by (5)

$$E(N_{10yrs}) = \frac{T}{t} \cdot n_t, P(X \leq 0, Y > 12) = \frac{10 \cdot 365}{2} \cdot 758 \cdot P(\tilde{X} > 1/\delta, Y > 12)$$

Table 9: The probability and frequencies (in 10 years) of having collision(s) that has $\Delta v > 12$

Model	Prob (minD)	Freq (minD)	Prob (First PET)	Freq (First PET)
hr	3.212352e-06	4.443807	2.529293e-05	34.98897
neglog	2.577869e-06	3.566095	3.133501e-05	43.34729
indep	3.2e-06	4.4640637	2.65e-05	36.6663307

Under both LminD and LPETD, we note that the bivariate POT1 models fails to make distinct difference from the product of independent marginal distributions. More importantly, for side collisions, a Δv of 10 m/s is sufficient to constitute a serious injury. However in the POT1 approach we have selected the margin u_2 to be 10, which makes the models less flexible for evaluating risk of left turning vehicles. Hence POT1 models are not always the best approach.

5 Copula approach

The major theoretical limitation of POT1 approach in modeling SMoS is that high thresholds are required for the sake of assumptions, yet the consequential SMoS measurement of a severe collision can be much lower than the thresholds. The idea of copula approach is that we study the marginal distribution of proximity and consequential SMoS separately, then compute probability (4) directly from the joint distribution and this is made possible by Sklar's theorem. Conceptually the copula approach is more straightforward than the POT1 approach, as there is no longer restriction on the events for which probability (4) is evaluated. But it is challenging to find suitable marginal distributions.

To mitigate the difficulties of finding parametric margins, we may instead use empirical distributions. Unfortunately it is not possible for proximity margins. As we discussed in Section 1, we wish to extend the support of temporal SMoS below 0 and it can not be done by using empirical distribution, since there is no observation of proximity SMoS that is less or equal to 0 in our data. However we may filter the data in specific ways so that the marginal distribution of proximity SMoS becomes known, or can be approximated by some known distributions.

The section is structured as the following: we first go through the basics of copula theory, including properties and the copula functions that we selected for this data set. Then we will fit the selected copulas to the data with two ways of transforming the margins: semi-parametric case and parametric case. In the first case we use some parametric distribution to model proximity SMoS and empirical distribution to model consequential SMoS; in the latter case we use some parametric distributions to model both margins.

5.1 Basic copula theory

Def 5.1.1. *Let U_1, U_2 be identically $Un[0,1]$ distributed random variables on probability space $(\Omega, \mathcal{F}, P)$, then a function $C : [0,1]^2 \mapsto [0,1]$ is called a copula if $C(U_1, U_2) = P\big(\cap_{i=1}^2 \{U_i \leq u_i\}\big)$*

Def 5.1.2. *A function $F : \mathbb{R}^d \mapsto [0,1]$ is a d dimensional distribution function if F satisfies the following properties:*

1. *$F(x_1, \cdots, x_d)$ is increasing w.r.t all arguments, moreover and $\lim_{x_i \to -\infty} F(\boldsymbol{x}) = 0$ for any $i = 1, \cdots, d$, $\lim_{\boldsymbol{x} \to \infty} F(\boldsymbol{x}) = 1$.*
2. *Let x_{-i} denote the arguments that is not x_i, then $\lim_{x_{-i} \to \infty} F(\boldsymbol{x}) = F_i(x_i)\, \forall i$, where F_i is a one dimensional distribution function.*
3. *For any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$, $F(\boldsymbol{x} \wedge \boldsymbol{y}) \geq F(\boldsymbol{x}) \cdot F(\boldsymbol{y})$. (d increasing)*

A bivariate copula is by definition a joint distribution function of r.vt of uniformly distributed r.vs. Hence it has the properties of distribution function in two dimen-

sions. According to the Kolmogorov existence on $\mathbb{R}^d$, for each probability measure on the unit square there exists unique copula associated with the measure s.t for any $B \in \mathcal{B}([0,1]^2)$, $P((U_1, U_2) \in B) = \int\int_B \mathrm{d}C(u,v)$, hence it can be shown that for any bivariate random vector (X, Y) that has marginal distributions $F_X(x), F_Y(y)$ and joint distribution $\boldsymbol{F}(x,y)$, there exist a copula such that $C(F_X(x), F_Y(y)) \overset{d}{=} \boldsymbol{F}(x,y)$ uniquely.

Lemma 5.1 (Inverse method)**.** *Let X a continuous r.v, F be the distribution function of X, $F^{-1}(u) := \inf\{x \in \mathbb{R} : F(x) \geq u\}$ be the general inverse function, then*

1. *$F(X) \sim Un[0,1]$*
2. *$F^-1(U) \sim F$ for some $U \sim Un[0,1]$*

Proof. Let $U := F(X)$, $q_u := \sup_x\{x : F(x) < u\}$, we wish to show $P(U \leq u) = u$, which is the same as showing $V \leq v \iff X \leq q_u$.

$$U \leq u \iff F(X) \leq u \iff X \leq F^{-1}(u) = \inf\{x \in \mathbb{R} : F(x) \geq u\} = q_u$$

So $P(U \leq u) = P(F(X) \leq u) = u \iff F(X) \sim Un[0,1]$, it follows immediately that $F^{-1}(U) = F^{-1}(F(X)) = X \sim F$. □

Remark 5.1.1. *Lemma 5.1 can be extended for all distribution functions in which case the proof becomes more complicated. Since we only deal with continuous r.v in this project, we adapt to a simpler version.*

Theorem 5.2 (Sklar's theorem)**.** *Let $(X_1, \cdots, X_d) \sim F$ be r.vt on probability space $(\Omega, \mathcal{F}, P)$, $F_1, \cdots, F_d$ are marginal distributions of F, then there exists a d-dimensional copula s.t*

$$C(F_1(x_1), \cdots, F_d(x_d)) = F(x_1, \cdots, x_d) \ \forall (x_1, \cdots, x_d)$$

Conversely, if C is a copula and there is a function

$$F(x_1, \cdots, x_d) = C(F_1(x_1)), \cdots, F_d(x_d)) \ \forall (x_1, \cdots, x_d)$$

then F is a joint distribution function whose margins are $F_1, \cdots, F_d$.

Proof. Let $(X_1, \cdots, X_d) \sim F$, $X_i \sim F_i$, then

$$F(x_1, \cdots, x_d) = P\big(\cap_{i=1}^d \{X_i \leq x_i\}\big) = P\big(\cap_{i=1}^d \{\underbrace{F_i(X_i)}_{\sim Un[0,1]} \leq F_i(x_i)\}\big) = C(F_1(x_1), \cdots, F_d(x_d)) \tag{34}$$

The third equality is followed from Lemma 5.1 and the fourth equality is followed from the definition of copula. Conversely, suppose $X_i \sim F_i$, $U_i = F_i(X_i)$, by Lemma 5.1 we know $U_i \sim Un[0,1]$. Assume that $(U_1, \cdots, U_d) \sim C$ for some copula C, then

$$C(u_1, \cdots, u_d) = P\big(\cap_{i=1}^d \{U_i \le u_i\}\big) = P\big(\cap_{i=1}^d \{F^{-1}(F(X_i)) \le F^{-1}(u_i)\}\big) = P\big(\cap_{i=1}^d \{X_i \le x_i\}\big)$$

□

Theorem 5.3 (Frechet-Hoeffding bounds)**.** *For every copula* $C(u_1, \cdots, u_d)$,

$$\max_{1\le i\le d}\big(\sum_{i=1}^{d}(u_i) - d + 1, 0\big) \le C(u_1, \cdots, u_d) \le \min_{1\le i\le d}(u_1, \cdots, u_d)$$

Proof. We prove the case for $d = 2$. Let $(U_1, U_2) \sim C$, then the upper bound is simply obtained from the definition of copula:

$$C(u_1, u_2) = P(U_1 \le u_1, U_2 \le u_2) \le \min(P(U_1 \le u_1), P(U_2 \le u_2)) = \min(u_1, u_2)$$

The lower bound comes from the De-Morgan's law:

$$\begin{aligned} C(u_1, u_2) = P(U_1 \le u_1, U_2 \le u_2) &= 1 - P(U_1 > u_1, U_2 > u_2) \\ &\ge 1 - P(U_1 \ge u_1) - P(U_2 \ge u_2) = u_1 + u_2 - 1 \end{aligned}$$

□

In the case of $d = 2$, the upper and lower Frechet-Hoeffding bounds are also copulas for the strongest positive and negative dependence structure respectively. The lower bound is called *counter monotonicity copula* (denoted as $W(u_1, u_2)$) and the upper bound is called *co-monotonicity copula* (denoted as $M(u_1, u_2)$). Note that not every copula can reach the Frechet-Hoeffding bounds.

The dependence structure of r.vs in r.vt is modelled by copula via the rank correlation:

Def 5.1.3 (Concordance)**.** *Let* $(x_1, y_1), (x_2, y_2)$ *be observations of r.vt* (X, Y)*, then* $(x_1, y_1), (x_2, y_2)$ *are said to be concordant if* $(x_2 - x_1) \cdot (y_2 - y_1) > 0$*, otherwise the pair is discordant.*

Kendall's τ is one popular measurement of rank correlation. It measures the mean difference between the number of concordant pairs and discordant pairs in a data set. Suppose a data set has n observations from a r.vt whose joint distribution is $H(x, y)$, then the sample τ is computed by:

$$\begin{aligned} \tau &= \frac{\#\text{concordant pairs} - \#\text{discordant pairs}}{\binom{n}{2}} \\ &= P((X_i - X_j) \cdot (Y_i - Y_j) > 0) - P((X_i - X_j) \cdot (Y_i - Y_j) > 0) \\ &= 2 \cdot P((X_i - X_j) \cdot (Y_i - Y_j) > 0) - 1 \text{ , for any } i \ne j \end{aligned}$$

Since $(X_i, Y_i), (X_j, Y_j)$ are drawn from the same r.vt

$$\begin{aligned} P((X_i - X_j) \cdot (Y_i - Y_j) > 0) &= 2 \cdot P(X_i < X_j, Y_i < Y_j) \\ &= 2 \cdot \int\int_{\mathbb{R}^2} P(X_i \leq x, Y_i \leq y | X_j = x, Y_j = y) \mathrm{d}H(x, y) \\ &= 2 \cdot \int\int_{\mathbb{R}^2} H(x, y) \mathrm{d}H(x, y) = 2 \cdot \int\int_{[0,1]^2} C(u, v) \mathrm{d}C(u, v) \end{aligned}$$

The last equality is due to Theorem 5.2. So the Kendall's τ can be written as a function of the copula C.

$$\tau = 4 \int\int_{[0,1]^2} C(u, v) \mathrm{d}C(u, v) - 1 \tag{35}$$

It is possible to verify that the Frechet-Hoeffding bounds indeed give the greatest and lowest τ:

$$\tau = 4 \int\int_{[0,1]^2} \min(u_1, u_2) \mathrm{d}M(u_1, u_2) - 1 = 4 \int_0^1 u_1 \mathrm{d}u_1 - 1 = 1$$

$$\tau = 4 \int\int_{[0,1]^2} \max(u_1 + u_2 - 1, 0) \mathrm{d}W(u_1, u_2) - 1 = -1$$

We use the `Copula` package Yan [2007] for parameter estimation and evaluation of copula probabilities. We will test the following one-parameter copula whose parametrzation is adopted from the manual of Yan [2007].

Normal: $C(u, v) = F(\sigma_1 \cdot \Phi^{-1}(u), \sigma_2 \cdot \Phi^{-1}(v))$ where F is bivariate normal distribution with covariance matrix Σ and σ_1, σ_2 are the diagonal entries of Σ.

Frank: $C(u, v) = -\frac{1}{\theta} \log \left(1 + \frac{(e^{-u\theta} - 1)(e^{-v\theta} - 1)}{e^{-\theta} - 1} \right)$

Clayton: $C(u, v) = \max \left(u^{-\theta} + v^{-\theta} - 1, 0 \right)^{-1/\theta}$

Gumbel: $C(u, v) = \exp \left(- \left((-\log(u))^\theta + (-\log(v))^\theta \right)^{1/\theta} \right)$

Hustler-Reiss: $C(u, v) = \exp \left(- \ln(u) \cdot \Phi\left(\theta^{-1} + 0.5 \cdot \theta \ln \frac{\ln(u)}{\ln(v)} \right) - \ln(v) \cdot \Phi\left(\theta^{-1} + 0.5 \cdot \theta \ln \frac{\ln(v)}{\ln(u)} \right) \right)$

Galambos: $C(u, v) = uv \cdot \exp \left(\left(|\ln(u)|^{-\theta} + |\ln(v)|^{-\theta} \right)^{-1/\theta} \right)$

Tawn: $C(u, v) = uv \cdot \exp \left(- \theta \frac{\ln(u) \ln(v)}{\ln(uv)} \right)$

5.2 Connection with MEVD

Suppose $G_*(\boldsymbol{x})$ is a MEVD (12), then the max-stability condistion for G_* yields $\forall s > 0$, $G^s(s\boldsymbol{x}) = G(\boldsymbol{x})$. This condition must be fulfilled by copula function that is equal in distribution to MEVD. Let $\boldsymbol{u} := \exp\left(-\boldsymbol{x}^{-1}\right)$, $\tilde{u} := \exp\left(-(s\boldsymbol{x})^{-1}\right)$, note that $\tilde{u} = u^{1/s}$, by Theorem (5.2):

$$G_*(\boldsymbol{x}) = C(\boldsymbol{u}) = G_*(-1/\ln(\boldsymbol{u})) = \exp\Big(- \mu_*\big(\mathbb{R}^{d+} \setminus [\mathbf{0}, \frac{-1}{\ln(\boldsymbol{u})}]\big)\Big) \tag{36}$$

$$\begin{aligned} G_*^s(s\boldsymbol{x}) = C^s(\tilde{\boldsymbol{u}}) = G_*^s(-1/\ln(\tilde{\boldsymbol{u}})) &= \exp\Big(- s \cdot \mu_*\big(\mathbb{R}^{d+} \setminus [\mathbf{0}, \frac{-1}{\ln(\tilde{\boldsymbol{u}})}]\big)\Big) \\ &= \exp\Big(- s\mu_*\big(\mathbb{R}^{d+} \setminus [\mathbf{0}, \frac{-s}{\ln(\boldsymbol{u})}]\big)\Big) \\ &= \exp\Big(- \mu_*\big(\mathbb{R}^{d+} \setminus [\mathbf{0}, \frac{-1}{\ln(\boldsymbol{u})}]\big)\Big) \end{aligned} \tag{37}$$

The first equality in (37) is due to (11). Putting (36) and (37) together, we obtain the relation $C^s(\boldsymbol{u}^{1/s}) = C(\boldsymbol{u})$. The class of copulas that satisfies this property is called **Extreme Value copula**. In bivariate case the EV copula can adapt to the MEVD with Pickand dependence function (23), s.t

$$C_G(u, v) = G_*\big(\frac{-1}{\ln(u)}, \frac{-1}{\ln(v)}\big) = \exp\big(\ln(uv) \cdot A\big(\frac{\ln(u)}{\ln(uv)}\big)\big) \tag{38}$$

The EV copula can be seen as a characterization of MEVD with uniform margins. The dependence structure will not change. In the 7 selected copulas of ours, Gumbel (logistic), Hustler-Reiss, Galambos (negative logistic) and Tawn are EV copulas which has same dependence structure as the MEVD in Table 1.

5.3 Goodness of Fit of copula

The empirical distribution plays an important role in testing goodness of fit of distribution function as the test statistics are continuous functional of the empirical distribution. In this subsection we introduce the one level parametric bootstrap test which is used both in testing GoF of copulas as well as the GoF of their marginal distributions. Main references of this section are Stute et al. [1993], Genest and Rémillard [2008] and chapeter 19,23 of Vaart [2000].

Def 5.3.1 (Empirical distribution)**.** *Let $X_1, \cdots, X_n$ be random sample from $X \sim F$, then the empirical distribution of X is defined by*

$$\mathbb{F}_n(x) = n^{-1} \sum_{i=1}^{n} \mathbf{1}\{X_i \leq x\}$$

By the law of large number, we have $\mathbf{F}_n(x) = n^{-1}\sum_{i=1}^n \mathbf{1}\{X_i \le x\} \xrightarrow{a.s} E(\mathbf{1}\{X \le x\}) = F(x)$. The empirical distribution will converge almost surely to the underlying distribution.

Theorem 5.4 (Glivenko-Cantelli)**.** *Let $X_1, \cdots, X_n$ be random sample from $X \sim F$, $\mathbb{F}_n$ be the empirical distribution of F, then* $\sup_x |\mathbb{F}_n(x) - F(x)| \to 0$

Proof. Let $F(x-)$ denote the limit from the left, $\varepsilon > 0$, $-\infty = x_0 < x_1 < \cdots < x_n = \infty$ be a partition of the real line s.t $\forall i = 1, \cdots, n$, $\sup_i F(x_i) - F(x_{i-1}) < \varepsilon$. By definition we have $\mathbb{F}_n(x_i-) = \mathbb{F}_n(x_{i-1}) < \mathbb{F}_n(x_i)$ and $F(x_i-) \le F(x_i)$,for $x \in [x_{i-1}, x_i]$, we have

$$\mathbb{F}_n(x_{i-1}-) - F(x_{i-1}-) - \varepsilon \le \mathbb{F}_n(x) - F(x) \le \mathbb{F}_n(x_i-) - F(x_i-) + \varepsilon$$

$\mathbb{F}_n(x) \to F \implies \sup_x |\mathbb{F}_n(x) - F(x)| \le \varepsilon$ for arbitrary ε. □

Suppose $\mathcal{P}_0 := \{F_\theta : \theta \in \Theta\}$ is a parametric family of distributions, we form the following hypothesis for testing GoF of F.

$$H_0 : F \in \mathcal{P}_0$$
$$H_1 : F \notin \mathcal{P}_0$$

With the test statistics:

$$\sqrt{n} \cdot \sup_x |F_n(x) - F_0(x)| \quad \text{Kolmogorov–Smirnov}$$
$$\int_{\mathbb{R}} |F_n(x) - F_0(x)|^2 \mathrm{d}F_0(x) \quad \text{Cramér–von Mises}$$

Under certain regularity conditions the test statistics possess asymptotic normality. But we do not use normal distribution to determine the p value of the test, instead the p value is computed from the ranks of the bootstrap samples. We loosen the confidence level from 0.95 to 0.9 i.e we reject the null hypothesis if the p-value is less 0.1. The procedure of one level bootstrap Cramér–von Mises test (for copula) goes as the following:

Suppose $C = C_\theta$ is the null hypothesis, $\hat{\theta}_n$ be a estimate of θ, $U_1, \cdots, U_d$ are samples between $[0, 1]$. $C_n(\boldsymbol{u}) = n^{-1}\sum_{i=1}^n \mathbf{1}\{\boldsymbol{u}_i \le \boldsymbol{u}\}$ is the empirical copula, then the one level bootstrap test is proceed:

- Compute the sample Cramér–von Mises test statistics:

$$S_n = n \cdot \int \left(C_n(\boldsymbol{u}) - C_{\hat{\theta}_n}(\boldsymbol{u})\right)^2 \mathrm{d}C_n(\boldsymbol{u}) = n \cdot E_{C_n}\left(\left(C_n(\boldsymbol{U}) - C_{\hat{\theta}_n}(\boldsymbol{U})\right)^2\right)$$
$$= \sum_{i=1}^n \left(C_n(\boldsymbol{u}_i) - C_{\hat{\theta}_n}(\boldsymbol{u}_i)\right)^2 \qquad (39)$$

- The bootstrap step (N samples): For each bootstrap trial $j = 1, \cdots, N$,
  1. simulate n bootstrap samples $u^*_{j,1}, \cdots, u^*_{j,n}$
  2. Compute the bootstrap Cramér–von Mises test statistics $S_{n,j}$ using (39) and $u^*_{j,1}, \cdots, u^*_{j,n}$.
- Compute the p value by $p^* = \frac{1}{N} \sum_{j=1}^{N} \mathbf{1}\{S_{n,j} > S_n\}$

The same procedure is applied to test the GoF of marginal distributions, except the Cramér–von Mises statistics is replaced with Kolmogorov-Smirnov statistics and the empirical copula is replaced with empirical distribution in one variable.

5.4 Fitting the marginal distributions

Let (X, Y) denote the r.vt of the SMoS pair (distance and ΔV) and $H(x, y)$ be its joint distribution. In order to apply Sklar's theorem, first we need to identify the marginal distributions. Note that for the margin of consequential SMoS we may use the empirical distribution since we are not interested in the extreme quantile of consequential SMoS, where observations are usually lacking.

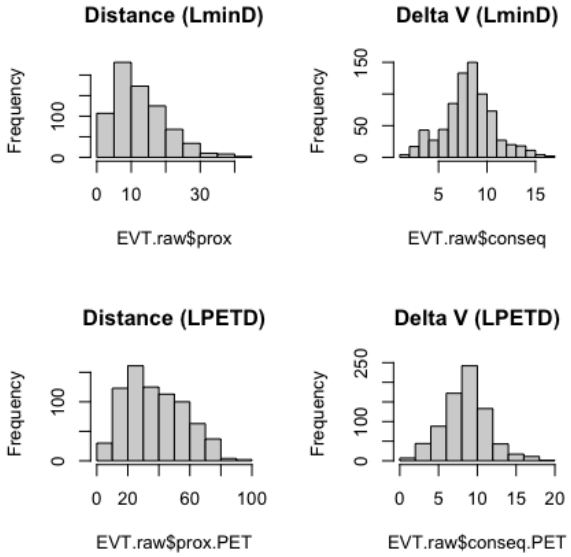


Figure 14: The histograms of margins in LminD and LPETD

Figure 14 shows the histogram of the marginal densities. In particular the marginal distribution of proximity SMoS should be an asymmetric distribution whose support strictly includes 0. Intuitively a gamma distribution with an extra shift parameter satisfied our demands, but the shift parameter turned out to be insignificant according to likelihood ratio test. Natural continuation is to test

other parametric distributions for proximity SMoS until we achieve satisfactory result from the one level bootstrap Kolmogorov-Smirnov test. The best candidates for proximity margin are Frechet distribution and bimodal distribution. Take bimodal distribution as an example, the normal copula and Tawn copula passes the Cramér–von Mises test (with p values 0.25 and 0.5). But the frequency computed from normal copula is 0.03 while the frequency computed from Tawn copula 312 [9]. The difference between two frequencies are huge and the later case is not within our acceptable range. This is either because poor choice of proximity margin, or the selected copulas are not suitable for this data set.

We believe that the marginal distribution of proximity SMoS is the most challenging part in this approach. Another idea is to filter the data such that the marginal distribution of proximity SMoS after filtering will follow certain distribution, or at least can be approximated by some known distributions. We introduce the block maxima (BM) filter and peaks over threshold filter (POT). Instead of taking the component-wise block maxima or exceedences in both margins, we restrict the filters only to the proximity SMoS. The filtered proximity is then able to be approximated by univariate extreme value distribution (GEV or GPD).

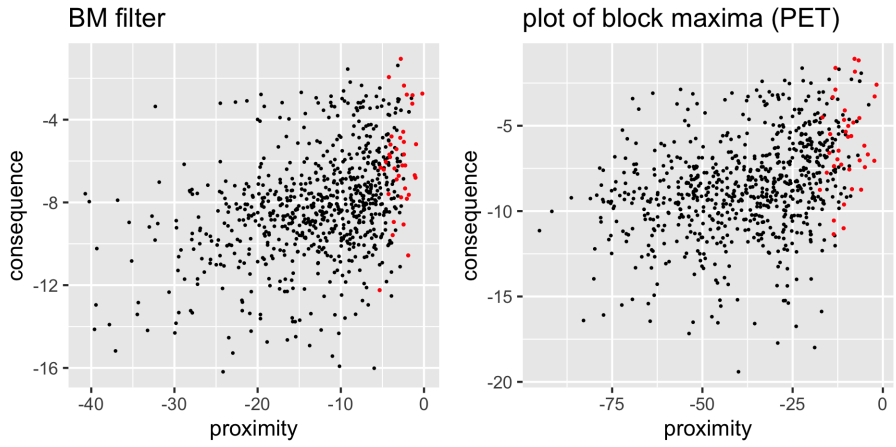


Figure 15: LminD and LPETD + BM filter

For the BM filter, we select the block maxima of proximity SMoS and keep the corresponding consequential SMoS in the pair. After applying the BM filter the number of data left is equal to the number of original data divided by the choice of block size. In Figure 15, the red dots are the data we used for fitting copulas.

Similarly for the POT filter, we select the pairs of SMoS in which the proximity SMoS are greater than a given threshold. Note that we have transformed both margins decreasingly, the decreasing transformation of proximity SMoS was explained in Section 1 and 3. The decreasing transformation of consequential SMoS is because some of the selected copulas are available only for data that have positive

[9]Same problems are observed when using Frechet margin, that the difference among frequencies are huge

Figure 16: LminD and LPETD + POT filter

rank correlation [10]. In the case of BM filter the decreasing transformation for both margins are negation while in the case of POT filter we use the shifted reciprocal for proximity SMoS and negation for consequential SMoS.

The disadvantage of using the extreme value filter is that there are much less observations after filtering. Implying that the use of empirical distribution in modeling consequential SMoS will be impaired, especially in LPETD, where no filtered observations have Δv greater than 12. Thus we must fit a parametric distribution in the second margin as well. Bimodal distributions are chosen to model the consequential SMoS, as it passes the bootstrap Kolmogorov-Smirnov test with p values 0.32.

Table 10: Parameters in the GEV proximity margins (BM)

Parameters (LminD)	95% lower CI	Estimate	95% upper CI
μ	-3.76	-3.31	-2.87
σ	0.93	1.25	1.57
γ	-0.53	-0.31	-0.09
Parameter (LPETD)	95% lower CI	Estimate	95% upper CI
μ	-12.82	-11.40	-9.98
σ	2.80	3.84	4.88
γ	-0.51	-0.23	0.05

Note that the confidence interval for the shape parameters in GPD margins include 0. Hence we fit also exponential distribution to the exceedences and found that the estimates for σ is the same up to the fourth digit. Furthermore the likelihood ratio test indicates we can not reject the exponential models with deviance statistics 0.002 and 0.5 (the critical value is $\chi^2(1) = 3.84$).

[10] Originally the proximity SMoS and consequential SMoS have positive rank correlations

Table 11: Parameters in the GPD proximity margins (POT)

Parameters (LminD)	95% lower CI	Estimate	95% upper CI
σ	0.04	0.05	0.06
γ	-0.170	0.004	0.179
Parameter (LPETD)	95% lower CI	Estimate	95% upper CI
σ	0.01	0.02	0.03
γ	-0.24	0.12	0.48

Under LPETD, bimodal distribution is used for the filtered consequential SMoS, the maximum likelihood estimation and plots use the functions from `mixtools` package Benaglia et al. [2009]. Suppose X_1, X_2 are two normal r.v with different mean and standard deviation, then its affine combination $\lambda X_1 + (1-\lambda)X_2$, $\lambda \in [0,1]$ is bimodal distributed. The bimodal distribution is used in modeling asymmetric density. The estimates of bimodal parameters are listed below:

Table 12: Estimates of bimodal distribution parameters

Case	μ_1	μ_2	σ_1	σ_2	λ
LPETD (BM)	-7.43	-4.03	1.88	1.88	0.64
LPETD (POT)	-7.69	-3.57	1.76	1.76	0.52

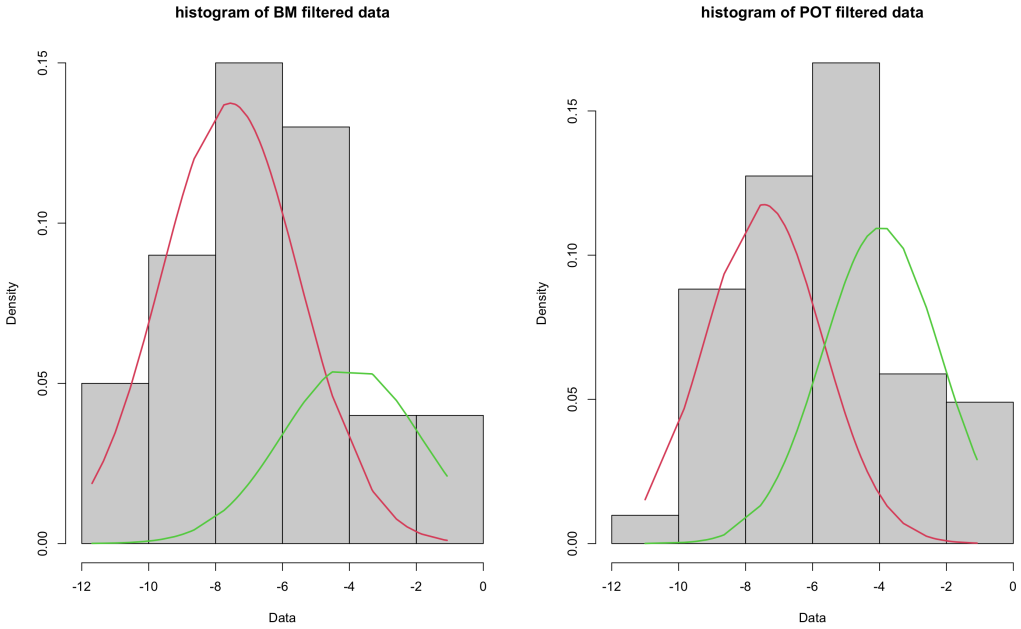


Figure 17: Histogram and kernel estimate of density of consequential SMoS (Left:BM filter; right: POT filter)

5.5 An intuitive formulation of the EV filters

Instead of taking the whole span of the data, as we did in the "copula + whole data" approach (equation (5)), the "copula + EV filter" approach focus on restricted outcome space. Consider the partition of outcome space based on the proximity SMoS as discussed in Section 1.3, that $\Omega = \cup_{i=1}^{n} \Omega_i$.

$$\begin{aligned} P(X \leq 0, Y > y) &= E(\mathbf{1}\{X \leq 0, Y > y\}) = \int_{\Omega} \mathbf{1}\{X \leq 0, Y > y\} \mathrm{d}P(\omega) \\ &= \int_{\cup_i \Omega_i} \mathbf{1}\{X \leq 0, Y > y\} \mathrm{d}P(\omega) \\ &= \int_{\Omega_1} \mathbf{1}\{X \leq 0, Y > y\} \mathrm{d}P(\omega) \\ &= E(\mathbf{1}\{X \leq 0, Y > y\} | \Omega_1) \cdot P(\Omega_1) \end{aligned} \tag{40}$$

The last equality is essentially the more general definition of conditional probability in terms of conditional expectation. Now let P_1 be a probability measure restricted on Ω_1, s.t $P_1(A) = P(A \cap \Omega_1)$, then we may rewrite

$$\begin{aligned} E(\mathbf{1}\{X \leq 0, Y > y\} | \Omega_1) &= \int_{\Omega_1} \mathbf{1}\{X \leq 0, Y > y\} \mathrm{d}P_1(\omega) \\ &= \int_{\Omega_1} \mathbf{1}\{F_1(X) \leq F_1(0), F_2(Y) > F_2(y)\} \mathrm{d}C(F_1(x), F_2(y)) \\ &= F_1(0) - C(F_1(0), F_2(y)) \end{aligned} \tag{41}$$

The second last equality is due to Theorem 5.2 that we can find a copula on probability space $(\Omega_1, \mathcal{F}_1, P_1)$. If X and Y are independent, then copula takes no part in the computation of (40) s.t:

$$\begin{aligned} P(X \leq 0, Y > y) &= P(Y > y) \cdot E(\mathbf{1}\{X \leq 0\}) \\ &= P(Y > y) \cdot \int_{\Omega} \mathbf{1}\{X \leq 0\} \mathrm{d}P(\omega) \\ &= \int_{\cup_i \Omega_i} \mathbf{1}\{X \leq 0, Y > y\} \mathrm{d}P(\omega) \\ &= (1 - F_{(n)}(y)) \cdot P_1(X \leq 0) \cdot P(\Omega_1) \end{aligned} \tag{42}$$

5.6 Estimation of the dependence parameter

The parameter estimation is mostly done by maximum likelihood estimation. But there are models in which the optimization procedure fails. As an alternative we can exploit the relation between copula and rank correlations, such as Kendall's τ.

For any copula with parameter θ, the rank correlation τ 35 is defined as

$$\tau = 4 \cdot E(C_\theta(U,V)) - 1 = 4 \int \int_{[0,1]^2} C_\theta(u,v) \mathrm{d}C_\theta(u,v) - 1 =: g(\theta)$$

Since we have only selected one-parameter copula, $g : \mathbb{R} \mapsto \mathbb{R}$. Moreover if the function g is bijective s.t its pre-image $g^{-1}(\tau)$ is continuously differentiable w.r.t τ, then $\hat{\theta}_n = g^{-1}(\hat{\tau})$ is an estimator, where we can use the sample estimates $\hat{\tau}$

It can be shown that [11]

$$\hat{\tau} = \frac{2n}{n-1} \sum_{i=2}^{n} \sum_{j=2}^{n} \big(C_n(u_i, v_j) \cdot C_n(u_{i-1}, v_{j-1}) - C_n(u_i, v_{j-1}) \cdot C_n(u_{i-1}, v_j) \big)$$

where $C_n(u,v) = n^{-1} \sum_{i=1}^{n} \mathbf{1}\{u_{(i)} \leq u, v_{(i)} \leq v\}$ is the empirical copula.

Table 13: Estimates of the dependence parameter in selected copulas

Copula	BM (LminD)	BM (LPETD)	POT (LminD)	POT (LPETD)
Gumbel	1.131	1.115	1.201	1.179
Clayton	0.262	0.362	0.402	0.531
Frank	1.053	1.405	2.000	1.829
normal	0.181	0.233	0.260	0.241
Husler-Reiss	0.693	0.732	0.803	0.794
Galambos	0.366	0.377	0.451	0.437
Tawn	0.324	0.421	0.456	0.356

5.7 Probabilities and frequencies

We will compute (1) the expected frequencies of collisions with $\Delta v > 12$, which is used in comparison with results from POT1, then with $\Delta v > 9$, which is to test the copula models validity in predicting collisions of lower severity. The expected frequencies are computed by (40),(41),(42) and (5). But the ordering is changed during the decreasing transformation of the data, (41) is changed accordingly, instead the frequency is computed by:

$$E(N_{10yrs}) = \frac{T}{t} \cdot n_t, P(X \leq 0, Y > y) = \frac{10 \cdot 365}{2} \cdot 758 \cdot (F_2(\tilde{y}) - C(F_1(\tilde{0}), F_2(\tilde{y}))) \cdot P(\Omega_1) \tag{43}$$

for some copula C defined on the smaller probability space $(\Omega_1, \mathcal{F}_1, P_1)$. The frequencies based on (43) will be compared with the frequencies assuming X, Y are independent where copula plays no role in the computation.

$$E(N_{10yrs}) = \frac{T}{t} \cdot n_t, P(X \leq 0, Y > y) = \frac{10 \cdot 365}{2} \cdot 758 \cdot (1 - F_{(n)}(y)) \cdot P_1(X \leq 0) \cdot P(\Omega_1) \tag{44}$$

[11] For example, in chapter 5.6 of Nelsen [2006]

Under the assumption that X, Y are independent, the BM filter and the POT filter is different in $P_1(X \leq 0)$, which is either GEV or GPD. $P(\Omega_1)$ is also different between BM filter and POT filter:

$$\begin{aligned}\text{BM filter: } E(N_{10yrs}) &= \frac{10 \cdot 365}{2} \cdot 758 \cdot P_1(\tilde{X} > 0, \tilde{Y} \leq -y) \cdot P(\Omega_1) \\ &= \frac{10 \cdot 365}{2} \cdot 758 \cdot (F_2(-y) - C(F_1(0), F_2(-y)))/\text{blocksize}\end{aligned} \tag{45}$$

$$\begin{aligned}\text{POT filter: } E(N_{10yrs}) &= \frac{10 \cdot 365}{2} \cdot 758 \cdot P_1(\tilde{X} > 1/\delta, \tilde{Y} \leq -y) \cdot \eta \\ &= (F_2(-y) - C(F_1(0), F_2(-y))) \cdot \frac{\text{number of exeedances}}{\text{number of observations}}\end{aligned} \tag{46}$$

Table 14: the Frequency of collisions ($\Delta v > 12$) computed based on selected copulas

Copula \ Data set	BM filter (LminD)	BM filter (LPETD)	POT filter (LminD)	POT filter (LPETD)
Gumbel	0.5518144	1.3075681	0.1577956	0.0147283
Clayton	0.6135732	0.4969233	0.1740107	0.0031587
Frank	1.7714681	1.5411765	0.3602215	0.0269669
Normal	0.1305214	0.6769440	0.0459019	0.0025400
Hulser-Reiss	0.0282599	0.2614957	0.0052122	0.0003423
Galambos	0.1966677	0.5897830	0.0441941	0.0035147
Tawn	2.4437302	1.9840091	0.6197687	0.0558500
Independent	20.2086	43.50807	8.84	2.9086

Table 15: the Frequency computed based on selected copulas ($\Delta v > 9$)

Copula \ Data set	BM filter (LminD)	BM filter (LPETD)	POT filter (LminD)	POT filter (LPETD)
Gumbel	3.3014931	32.70260	2.4074726	0.6770924
Clayton	8.5555891	37.88434	6.3459871	0.8863747
Frank	9.9036418	37.96380	5.2513336	1.1711494
Normal	1.9878885	31.43132	1.9449360	0.4212750
Hustler-Reiss	0.2966294	10.71848	0.1604585	0.0356338
Galambos	1.3523754	18.04795	0.8124120	0.2018897
Tawn	12.2340224	43.49555	8.0580916	2.2195737
Independent	93.69458	200.6205	41.01055	13.4121

The frequencies in Table 15 can not be computed by means of POT1 approach, since the thresholds we chose were 10 in both LminD and LPETD. Note that in Table 15, the independent case is computed from (42), (45) and (46). Compare Table 15 with Table 14, we can see the differences in frequencies as stepping away slightly from high-severity collisions. We found that the Hustler-Reiss copula, Galambos copula and Gumbel copula produce the most reasonable frequency estimates when Δv is greater than 9. Furthermore we compute even the frequencies using these three copulas with even smaller Δv, to gain some ideas on how good are these three copulas in modeling collisions of low severity.

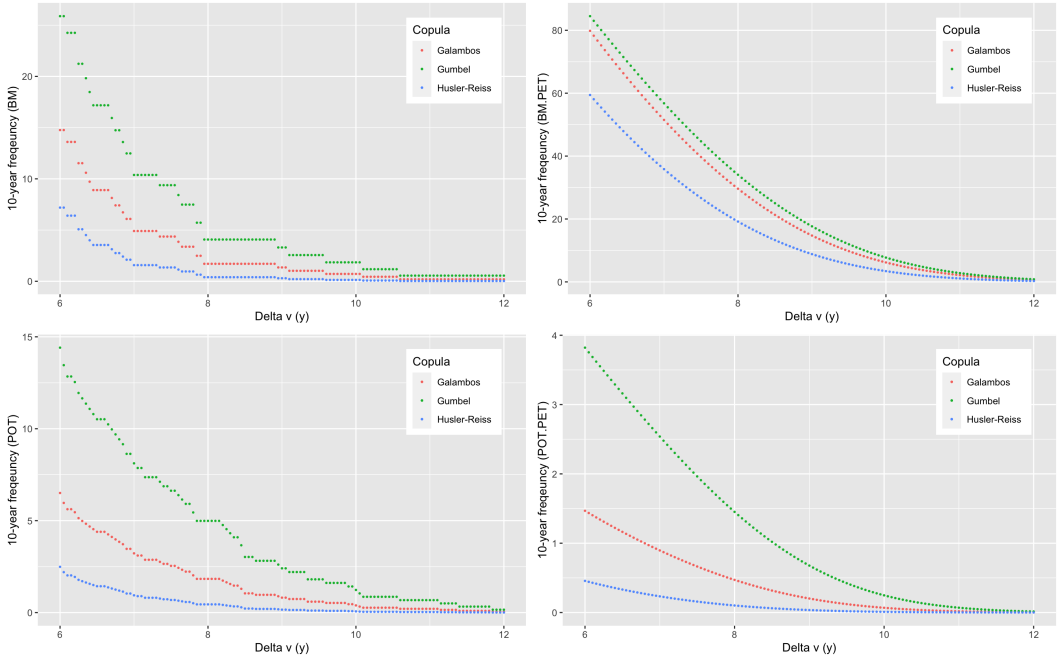


Figure 18: The frequency of collisions that have Δv¿y

Figure 18 shows the change of estimated frequency of collisions of $\Delta v > y$, $6 \leq y \leq 12$. Under LminD there are steps in the frequencies, which are due to the use of empirical distribution to model consequential SMoS. The POT filter with PET measurement perform best in predicting low-severity crash frequency, for the 10-year frequencies are less than 10 when $\Delta v > 6$, which are the closest to the police report.

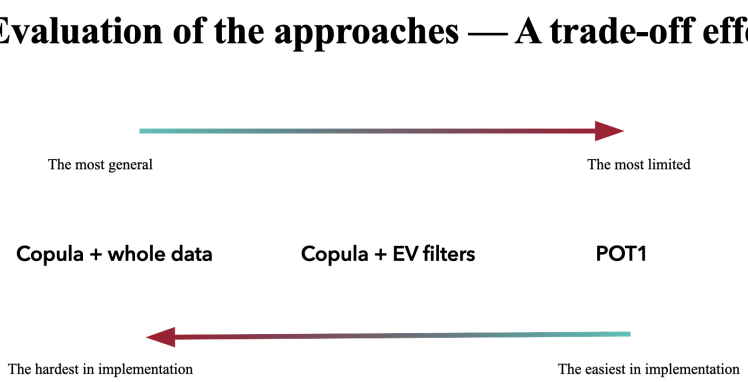
6 Conclusions and Discussions

In the data of 758 of left-turning vehicles, each X_i and Y_i does not appear to be dependent in time. We can assume for this intersection that the each sample (X_i, Y_i) are independent. Distance and Δv are weakly dependent (in terms of rank correlation τ) in the total data and asymptotically independent. The results on the data are consistent with Borsos [2021], where different choices of SMoSes were applied. Without any statistical models we may say the consequential SMoS in a near-collision scenario does not exhibit relatable patterns. However as shown in Section 4 and 5, that modeling the joint distribution independently will result in much greater frequencies.

To compute probability (2) and expected 10-year frequency (5), we used three approaches where POT1 and copula + EV fillers were discussed in details. We found that Hustler-Reiss and Galambos copula produced the most reasonable estimates of 10-year frequencies, which coincides with the fittest from Section 4, that the dependence structure of the upper tail distribution is best described by Hustler-

Reiss and negative logistic (which is also called Galambos) models. We do not know yet whether this is a coincidence, since we have selected the BEVD models and copula models in different ways.

The POT1 models for this data set is not as satisfactory as we anticipated. First of all, BEVD models make no difference from the product of two marginal extreme value distribution if the data exhibit asymptotic independence [12]. Secondly the threshold choice for consequential margin is too high. We have shown that the new approach, copula + EV filter, gave us more flexibility to analyze collisions of of lower severity. Within the two EV filters, we prefer the POT filter rather than BM filter. To begin with, from a theoretical perspective, (40) is designated for POT filter. If the BM filter is applied, then $P(\Omega_1)$ in (40) becomes trickier. In the computation we made, 1/block size is used as $P(\Omega_1)$, which is less theoretically justified since (40) should instead become $E(\mathbf{1}\{\max_i X \leq 0, Y > y\}|\Omega_1) \cdot P(\Omega_1)$. This is reflected in the 10-year frequency: when Δv tends to smaller value, the estimated frequencies by BM filter (especially under LPETD), are much larger than those from POT filter.



If we were to compare the three approaches, the aspects that intrigues us are the flexibility in analyzing traffic conflicts and the difficulties in implementation. The "copula + whole data" approach is the most general from a theoretical point of view, but it perform poorly, at least for this data set. The biggest problem is the distribution of proximity SMoS, which was and still is an interesting topic where more researches are needed.

Both the POT1 approach and the copula + EV filter approach are means of bypassing the assumption on the distribution of proximity SMoS. The POT1 has the simplest implementation, as the assumptions on the marginal distributions are defined by the model, but the events that can be studied by POT1 is limited to severe

[12]There are POT1 models based on multivariate slow varying functions for asymptotic independent data Ledford and Tawn [1996]. For asymptotic dependent data it is enough to require the tail to be regular varying (chapter 5.4 Resnick [2008])

collisions. The copula + EV filters combines the advantages from POT1 approach and copula + whole data approach – it is more flexible for analyzing collisions of lower severity, though the marginal distribution of consequential SMoS is the price to pay. But under some circumstances we might be able to use the empirical distribution, which makes the implementation almost as simple as the POT1 approach. Another concern with the copula + EV filter is that we are extrapolating the joint distribution with the distribution defined on a much smaller outcome space. Hence the deviation in the joint probability/frequency may be huge as we keep decreasing Δv, as shown in Figure 18. Nevertheless the copula + EV filters approach is a good starting point for investigation of less severe collisions, which might rise interest in further researches such as insurance claims of traffic accidents.

There are some technical details from this report which may be studied more thoroughly in future studies. For instance, the probability distribution of proximity SMoS, which is not a new yet still challenging topic. Intuitively negative binomial regression is a standard model for counts. It has been used in modeling crash frequency with proximity SMoS and other covariates Tarko [2019](Chapter 2). The proximity SMoS is suggested to be gamma-like distributed. Combining with our discussion in Section 1, we may approximate the gamma distributed margins with some distribution (for example, the mixture of some parametric distributions) whose support strictly includes 0. Alternatively, one may consider sampling from the distribution of proximity SMoS directly using Monte Carlo instead of going after the parametric distribution of proximity SMoS that only makes sense for some specific crash sites. Better understanding of the distribtuion of proximity SMoS can enhance the algorithm for micro-behaviour simulation, as well as the "copula + whole data" approach. Another example is the construction of confidence intervals for BEVD probabilities, as we discussed in Section 4.4.

To sum up briefly, we applied two approaches to estimated long term crash frequency of certain severity given a relatively short observational data, in which the "copula + EV filter" is developed by us. More precisely, we have only looked at a parameter (the crash probability) in a simple crash frequency model (with binomial distribution), which assumes constant traffic volume. The next step for research in this direction is to calibrate the crash probabilities in more realistic models where more parameters are involved. The approach should be tested with different SMoS, in different crash sites, for its validity.

References

Christer Hydén. *The development of a method for traffic safety evaluation :.* Lund,Sweden, 1987.

Praprut Songchitruksa and Andrew P. Tarko. The extreme value theory approach to safety estimation. *Accident Analysis and Prevention*, 38(4):811–822, 2006. ISSN 0001-4575. doi:https://doi.org/10.1016/j.aap.2006.02.003. URL `https://www.sciencedirect.com/science/article/pii/S0001457506000236`.

Andrew P. Tarko. Use of crash surrogates and exceedance statistics to estimate road safety. *Accident Analysis and Prevention*, 45:230–240, 2012. ISSN 0001-4575. doi:https://doi.org/10.1016/j.aap.2011.07.008. URL `https://www.sciencedirect.com/science/article/pii/S000145751100193X`.

Lai Zheng, Karim Ismail, and Xianghai Meng. Freeway safety estimation using extreme value theory approaches: A comparative study. *Accident Analysis and Prevention*, 62:32–41, 2014a. ISSN 0001-4575. doi:https://doi.org/10.1016/j.aap.2013.09.006. URL `https://www.sciencedirect.com/science/article/pii/S000145751300359X`.

Attila Borsos, Haneen Farah, Aliaksei Laureshyn, and Marjan Hagenzieker. Are collision and crossing course surrogate safety indicators transferable? a probability based approach using extreme value theory. *Accident Analysis and Prevention*, 143:105517, 2020. ISSN 0001-4575. doi:https://doi.org/10.1016/j.aap.2020.105517. URL `https://www.sciencedirect.com/science/article/pii/S0001457519310735`.

Haneen Farah and Carlos Lima Azevedo. Safety analysis of passing maneuvers using extreme value theory. *IATSS Research*, 41(1):12–21, 2017. ISSN 0386-1112. doi:https://doi.org/10.1016/j.iatssr.2016.07.001. URL `https://www.sciencedirect.com/science/article/pii/S0386111216300218`.

Jenny K. Jonasson and Holger Rootzén. Internal validation of near-crashes in naturalistic driving studies: A continuous and multivariate approach. *Accident Analysis and Prevention*, 62:102–109, 2014. ISSN 0001-4575. doi:https://doi.org/10.1016/j.aap.2013.09.013. URL `https://www.sciencedirect.com/science/article/pii/S0001457513003667`.

Chen Wang, Chengcheng Xu, and Yulu Dai. A crash prediction method based on bivariate extreme value theory and video-based vehicle trajectory data. *Accident Analysis and Prevention*, 123:365–373, 2019. ISSN 0001-4575. doi:https://doi.org/10.1016/j.aap.2018.12.013. URL `https://www.sciencedirect.com/science/article/pii/S0001457518304275`.

Aliaksei Laureshyn. *Application of automated video analysis to road user behaviour. [Elektronisk resurs].* Lund University, 2010.

Lai Zheng, Karim Ismail, Tarek Sayed, and Tazeen Fatema. Bivariate extreme value modeling for road safety estimation. *Accident Analysis and Prevention*, 120:83–91, 2018. ISSN 0001-4575.

doi:https://doi.org/10.1016/j.aap.2018.08.004. URL https://www.sciencedirect.com/science/article/pii/S0001457518304007.

Lai Zheng, Tarek Sayed, and Mohamed Essa. Validating the bivariate extreme value modeling approach for road safety estimation with different traffic conflict indicators. *Accident Analysis and Prevention*, 123:314–323, 2019. ISSN 0001-4575. doi:https://doi.org/10.1016/j.aap.2018.12.007. URL https://www.sciencedirect.com/science/article/pii/S0001457518311138.

Joana Cavadas, Carlos Lima Azevedo, Haneen Farah, and Ana Ferreira. Road safety of passing maneuvers: A bivariate extreme value theory approach under non-stationary conditions. *Accident Analysis and Prevention*, 134, 2020. ISSN 0001-4575.

Aliaksei Laureshyn, Åse Svensson, and Christer Hydén. Evaluation of traffic safety, based on micro-level behavioural data: Theoretical framework and first implementation. *Accident Analysis and Prevention*, 42(6):1637–1646, 2010. ISSN 0001-4575. doi:https://doi.org/10.1016/j.aap.2010.03.021. URL https://www.sciencedirect.com/science/article/pii/S0001457510001041.

Aliaksei Laureshyn, Tim De Ceunynck, Christoffer Karlsson, Åse Svensson, and Stijn Daniels. In search of the severity dimension of traffic events: Extended delta-v as a traffic conflict indicator. *Accident Analysis & Prevention*, 98:46–56, 2017. ISSN 0001-4575. doi:https://doi.org/10.1016/j.aap.2016.09.026. URL https://www.sciencedirect.com/science/article/pii/S0001457516303566.

Omar Bagdadi. Estimation of the severity of safety critical events. *Accident Analysis and Prevention*, 50:167–174, 2013. ISSN 0001-4575. doi:https://doi.org/10.1016/j.aap.2012.04.007. URL https://www.sciencedirect.com/science/article/pii/S0001457512001406.

Attila Borsos. Application of bivariate extreme value models to describe the joint behavior of temporal and speed related surrogate measures of safety. *Accident Analysis and Prevention*, 159:106274, 2021. ISSN 0001-4575. doi:https://doi.org/10.1016/j.aap.2021.106274. URL https://www.sciencedirect.com/science/article/pii/S0001457521003055.

Douglas J. Gabauer and Hampton C. Gabler. Comparison of roadside crash injury metrics using event data recorders. *Accident Analysis & Prevention*, 40(2):548–558, 2008. ISSN 0001-4575. doi:https://doi.org/10.1016/j.aap.2007.08.011. URL https://www.sciencedirect.com/science/article/pii/S000145750700139X.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2021. URL https://www.R-project.org/.

Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016. ISBN 978-3-319-24277-4. URL https://ggplot2.tidyverse.org.

L. de Haan. *On Regular Variation and Its Application to the Weak Convergence of Sample Extremes, by L. de Haan*. Mathematical Centre tracts. 1970.

James III Pickands. Statistical inference using extreme order statistics. *The Annals of Statistics*, 3(1):119 – 131, 1975. ISSN 00905364.

Sidney I. Resnick. *Extreme values, regular variation and point processes.* Springer series in operations research and financial engineering. Springer, 2008. ISBN 9780387759524.

Balkema A. A. and Resnick S. I. Max-infinite divisibility. *Journal of Applied Probability*, 14(2):309 – 319, 1977. ISSN 00219002.

2) de Haan, L. (1 and 2) Resnick, S.I. (1. Limit theory for multivariate sample extremes. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 40 (4):317–337, 1977. ISSN 00443719.

Sharpe Michael. Operator-stable probability distributions on vector groups. *Transactions of the American Mathematical Society*, 136:51 – 65, 1969. ISSN 00029947.

James III Pickands. Multivariate extreme value distributions. In *Proc. 43rd Sess. International Statistical Institute*, pages 859–878, Buenos Aires, 2010. Amsterdam: International Statistical Institute.

A. G. Stephenson. evd: Extreme value distributions. *R News*, 2(2):0, June 2002. URL `https://CRAN.R-project.org/doc/Rnews/`.

Harry Joe, Richard L. Smith, and Ishay Weissman. Bivariate threshold methods for extremes. *Journal of the Royal Statistical Society. Series B (Methodological)*, 54(1):171–183, 1992. ISSN 00359246. URL `http://www.jstor.org/stable/2345953`.

Nader Tajvidi. *Multivariate generalised Pareto distributions.* Department of Mathematics, Chalmers University of Technology and the University of Göteborg: 1995:17. 1995.

Jan Beirlant, Yuri Goegebur, Jozef Teugels, and Johan Segers. *Statistics of extremes. [Elektronisk resurs] theory and applications.* Wiley series in probability and statistics. J. Wiley, 2004. ISBN 9780471976479.

T-analyst,software for semi-automated video processing. URL `https://bitbucket.org/TrafficAndRoads/tanalyst/wiki/Manual`.

Lai Zheng, Karim Ismail, and Xianghai Meng. Shifted gamma-generalized pareto distribution model to map the safety continuum and estimate crashes. *Safety Science*, 64:155–162, 2014b. ISSN 0925-7535. doi:https://doi.org/10.1016/j.ssci.2013.12.003. URL `https://www.sciencedirect.com/science/article/pii/S0925753513003056`.

Eric Gilleland and Richard W. Katz. extRemes 2.0: An extreme value analysis package in R. *Journal of Statistical Software*, 72(8):1–39, 2016. doi:10.18637/jss.v072.i08.

Heidi Mach. Traffic safety analysis by surrogate measures:an extreme value approach. Master's thesis, Lund University, 2022. URL `https://www.lu.se/lup/publication/9076348`.

Jun Yan. Enjoy the joy of copulas: With a package copula. *Journal of Statistical Software*, 21(4):1–21, 2007. doi:10.18637/jss.v021.i04. URL https://www.jstatsoft.org/index.php/jss/article/view/v021i04.

W. (1) Stute, W.G. (2) Manteiga, and M.P. (3) Quindimil. Bootstrap based goodness-of-fit-tests. *Metrika: International Journal for Theoretical and Applied Statistics*, 40(1):243–256, 1993. ISSN 1435926X. URL https://link.springer.com/article/10.1007/BF02613687.

Christian Genest and Bruno Rémillard. Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 44(6):1096 – 1127, 2008. doi:10.1214/07-AIHP148. URL https://doi.org/10.1214/07-AIHP148.

A. W. van der Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics: 3. Cambridge University Press, 2000. ISBN 9780521496032.

Tatiana Benaglia, Didier Chauveau, David Hunter, and Derek Young. mixtools: An r package for analyzing finite mixture models. *Journal of Statistical Software*, 32, 10 2009. doi:10.18637/jss.v032.i06.

Roger B. Nelsen. *An introduction to copulas.* Springer series in statistics. Springer, 2006. ISBN 9780387286594.

Anthony W. Ledford and Jonathan A. Tawn. Statistics for near independence in multivariate extreme values. *Biometrika*, 83(1):169–187, 1996. ISSN 00063444. URL http://www.jstor.org/stable/2337440.

Andrew Tarko. *Measuring Road Safety Using Surrogate Events*. ELSEVIER, 2019. ISBN 978012810504. URL https://www.sciencedirect.com/science/book/9780128105047.

Allan Gut and SpringerLink. *Probability: A Graduate Course.* Springer Texts in Statistics: 75. Springer New York, 2013. ISBN 9781461447078.

M. R. Leadbetter and Georg Lindgren. *Extremes and related properties of random sequences and processes.* Springer series in statistics. Springer-Vlg, 1983. ISBN 9781461254515.

jbowman (https://stats.stackexchange.com/users/7555/jbowman). Goodness of fit test for a mixture in r. Cross Validated. URL https://stats.stackexchange.com/q/28878. URL:https://stats.stackexchange.com/q/28878 (version: 2012-05-21).

7 Appendices

7.1 Univariate max-stable distributions

The proof of *convergence to types* is adapted from chapter 7.4 of Gut and Springer-Link [2013]. The development of *Extremal Type Theorem* is referred to chapter 1.3, 1.4 of Leadbetter and Lindgren [1983]. The univariate version of some important results in Section 2 are also proved.

First we would like to introduce the *type of laws* of random variables:

Def 7.1.1 (types)**.** *The law of types a r.v $X \sim F$ is defined by the family of distributions $\{\sigma X + \mu : \sigma > 0, \mu \in \mathbb{R}\}$. Another random variable $Y \sim G$ is said to be the same type as X if $Y = \sigma X + \mu$, or equivalently, $F(x) = G(\sigma x + \mu)$.*

A distribution of degenerate type is defined when $\sigma = 0$.

Lemma 7.1 (Extended continuous mapping theorem)**.** *Let $(D, d), (E, d)$ be two metric spaces with the same metric, $\{X_n\}_n$ be a sequence of random variables s.t X_n takes value in $D_n \subset D, \forall n$, $\{g_n : D_n \mapsto E\}_n$ be a sequence of continuous map s.t $g_n \to g$ and $X_n \xrightarrow{d} X$, then $g_n(X_n) \xrightarrow{d} g(X)$ where $g(X)$ is a random variable that takes values in E*

Proof. The proof relies on Bolzano-Weierstrass theorem and portmanteau lemma, one may refer to chapter 18 of Vaart [2000] □

We introduce extended continuous mapping theorem as a technical lemma for proving the *convergence to types theorem*.

Lemma 7.2. *1. Suppose $\alpha_n \to A = \infty$ then $X_n/\alpha_n \xrightarrow{d} 0$ for any sequence of r.v $\{X_n\}_n$*

2. Suppose $X_n \xrightarrow{p} X$, $\{\beta\}_n$ is an unbounded sequence with $\sup_n |\beta_n| = \infty$, then $X_n - \beta_n$ does not converge in distribution.

Proof. 1. Since $\lim_{x\to\infty} P(|X_n| > x) = 0$, $\forall n = 1, 2, \cdots$, $\exists\{c_n > 0 : P(|X_n| > c_n) < 1/n\}_n$. Let $\{\alpha_n = n \cdot c_n\}$, then for any $\varepsilon > 0$ and $\varepsilon > n^{-1}$:

$$P\big(\big|\frac{X_n}{\alpha_n}\big| > \varepsilon\big) = P(|X_n| > n \cdot c_n \cdot \varepsilon) \leq P(|X_n| > c_n) < n^{-1} \to 0, \text{ , as } n \to \infty$$

2. Suppose $\{\beta_n\}_n$ is unbounded, then $\exists$ sub-sequences $\{\beta_{n_j}\}_j, \{\beta_{n_k}\}_k$ s.t $\lim_j \to \infty \beta_{n_j} \to -\infty$, $\lim_k \to \infty \beta_{n_k} \to \infty$, but this means that $F_{X_n - \beta_n}$ has two different limits, namely

$$\lim_{j\to\infty} P(X_{n_j} - \beta_{n_j} \leq x) = 0 \ , \ \lim_{k\to\infty} P(X_{n_k} - \beta_{n_k} \leq x) = 0$$

□

Theorem 7.3. *Suppose $\{X_n\}_n$ is a sequence of random variables s.t $X_n \xrightarrow{d} U$ and $\exists$ sequences of normalizing constants $\{\alpha_n > 0\}_n, \{\beta_n \in \mathbb{R}\}_n$ s.t*

$$\frac{X_n - \beta_n}{\alpha_n} \xrightarrow{d} V \quad \textit{where} U, V \textit{ are non-degenerate.}$$

then $\alpha_n \to A \in \mathbb{R}^+ \setminus \{0\}$, $\beta_n \to B \in (-\infty, \infty)$ and $U \overset{d}{=} A \cdot V + B$, i.e U, V are of the same type.

Proof. Assume that $\alpha_n \to A$, $\beta_n \to B$, $\{g_n(\cdot) : \frac{(\cdot)-\beta_n}{\alpha_n}\}$, then $g_n(\cdot) \to g(\cdot) := \frac{(\cdot)-B}{A}$.

Denote $V_n := g_n(X_n)$, $V := g(U)$, by the extended continuous mapping theorem (7.1) we obtain immediately that

$$X_n \xrightarrow{d} U \implies V_n \xrightarrow{d} V = \frac{U - B}{A}$$

Next we want to show that if U, V are non-degenerate, the sequences $\{\alpha_n\}_n, \{\beta_n\}_n$ must be bounded, meaning that:

$$0 < \inf_n \alpha_n \leq \sup_n \alpha_n < \infty \,, \; \sup_n |\beta_n| < \infty \tag{47}$$

Let $\tilde{X}_n := X_n/\alpha_n$, $\tilde{\beta}_n := \beta_n/\alpha_n$ and suppose first $\sup_n |\beta_n| < \infty$ then

$$\tilde{X}_n - \tilde{\beta}_n = V_n \xrightarrow{d} V$$

By (2) of Lemma 7.2, $\sup_n |\tilde{\beta}_n| = \sup_n |\beta_n/\alpha_n| < \infty \implies \inf_n \alpha_n > 0$.

Boundedness of $\{\alpha_n\}_n$: Now assume that $\sup_n \alpha_n = \infty$, then we can find a sub-sequence $\{\alpha_{n_k}\}_k \subset \{\alpha_n\}_n$ s.t $\lim_{k\to\infty} \alpha_{n_k} = \infty$. By (1) of Lemma 7.2 we know that $X_{n_k}/\alpha_{n_k} \xrightarrow{p} 0$. In addition we have $sup_n |\beta_{n_k}/\alpha_{n_k}| < \infty$, thus the sequence $\{\beta_{n_k}/\alpha_{n_k}\}_k$ is bounded and there exists a convergent sub-sequence $\{\{\beta_{n_{k_j}}/\alpha_{n_{k_j}}\}_j\}$ which has finite limit point c. We use (1) of Lemma 7.2 again that

$$\lim_j \to \infty \frac{X_{n_{k_j}} - \beta_{n_{k_j}}}{\alpha_{n_{k_j}}} \xrightarrow{d} V = -c$$

which contradicts that V is non-degenerate.

Boundedness of $\{\beta_n\}_n$: Assume that $\sup_n |\beta_n| = \infty$, then $\exists$ sub-sequence $\{\beta_{n_k}\}_k$ which converges to ∞. By assumption of the theorem we have $V_n \xrightarrow{d} V$, thus

$$\lim_{k\to\infty} V_{n_k} \lim_{k\to\infty} \tilde{X}_{n_k} - \tilde{\beta}_{n_k} \xrightarrow{d} V$$

which is not possible according to (2) of Lemma 7.2. Hence $\sup_n |\beta_n| < \infty$

It remains to show that $U \overset{d}{=} AV + B$. We start with the iteration:

$$U \overset{d}{=} \frac{\frac{U-b}{a} - b}{a} \overset{d}{=} \frac{\frac{\frac{U-b}{a}-b}{a} - b}{a} \overset{d}{=} \cdots \overset{d}{=} \frac{U}{a^n} - b\sum_{k=1}^{n} a^{-k} = \frac{U - b \cdot \sum_{k=1}^{n} a^{n-k}}{a^n}$$

Choose $\{\alpha_n = a^n\}_n, \{\beta_n = b \cdot \sum_{k=1}^{n} a^{n-k}\}_n$, the boundedness of $\{\alpha_n\}_n, \{\beta_n\}_n$ as we showed above implies $a = 1, b = 0$. So $\lim_{n\to\infty} \alpha_n = 1 =: A$, $\lim_{n\to\infty} \beta_n = 0 =: B$ Basically we showed that $U \overset{d}{=} \frac{U-B}{A} \implies A = 1, B = 0$.

Finally suppose $\{\alpha_n\}_n, \{\beta_n\}_n$ that satisfy the assumption of the theorem and the conditions in the proof, let $\{\alpha_{n_k}\}_k, \{\alpha_{n_j}\}_j, \{\beta_{n_k}\}_k, \{\beta_{n_j}\}_j$ be convergent sub-sequences such that

$$\lim_{k\to\infty} V_{n_k} \xrightarrow{d}= \frac{U - B}{A} = V = \lim_{j\to\infty} V_{n_j} \xrightarrow{d} \frac{U - B^*}{A^*}$$

$$\iff \frac{A}{A^*} = 1\,,\; B - B^* = 0$$

□

Corollary 7.3.1 (Convergence to types)**.** *Suppose $\{X_n\}_n$ is a sequence of random variables and $\exists$ sequences of normalizing constants $\{\sigma_n > 0\}_n\,,\; \{\mu_n \in \mathbb{R}\}_n\,,\; \{\alpha_n > 0\}_n, \{\beta_n \in \mathbb{R}\}_n$ s.t*

$$\frac{X_n - \mu_n}{\sigma_n} \xrightarrow{d} U\;,\; \textit{and}\; \frac{X_n - \beta_n}{\alpha_n} \xrightarrow{d} V \quad \textit{where} U, V \textit{ are non-degenerate.}$$

then $\alpha_n/\sigma_n \to A < \infty\,,\; (\beta_n - \mu_n)/\sigma_n \to B \in (-\infty, \infty)$ and $U \overset{d}{=} A \cdot V + B$, i.e U, V are of the same type.

Proof. Let $\tilde{X}_n = \frac{X_n - \mu_n}{\sigma_n} \xrightarrow{d} U$, by assumption $\frac{X_n - \beta_n}{\alpha_n} \xrightarrow{d} V$, then

$$\implies \frac{\sigma_n \tilde{X}_n + \mu_n - \beta_n}{\alpha_n} \xrightarrow{d} V \implies \frac{\tilde{X}_n - \sigma_n^{-1}(\beta_n - \mu_n)}{\frac{\alpha_n}{\sigma_n}} \xrightarrow{d} V$$

then We apply Theorem 7.3 to $\tilde{X}$. □

Remark 7.3.1. *For later use, it is more convenient to state the result in terms of the corresponding distributions of X_n: Suppose the conditions in corollary (7.3.1) is fulfilled and $U \sim G_1\,,\; V \sim G_2$, then its result may be alternatively stated as: $F_n(\sigma_n x + \mu_n) \to G_1(x)\,,\; F_n(\alpha_n x + \beta_n) \to G_2(x) \implies$*

$$F_n\big(\frac{\alpha_n}{\sigma_n} x + \frac{\beta_n - \mu_n}{\sigma_n}\big) \to G_2(x) = G_1(Ax + B)$$

The interpretation of theorem 7.3 is that if there exists normalizing sequences s.t $F_n(\sigma_n x + \mu_n)$ converges to some non-degenerate G, then F_n can not converge to other distributions that are of different type from G. Under this circumstance, F_n **is said to be in the domain of attraction of** G denoted as $F \in \mathcal{D}(G)$. Next we need to introduce the concept of max-stability. It can be shown that if the limiting distribution of extreme order statistics converges to some non-degenerate distribution, then the limiting distribution must be max-stable. The max-stability can be extended to higher dimensions as disussed in Section 2.

Def 7.1.2 (max-stability)**.** *Suppose G is a distribution function, then G is said to be max-stable if $\exists$ normalizing sequence $\{\alpha_n\}_n$, $\{\beta_n\}_n$ s.t $\lim_{n\to\infty} G^n(\alpha_n x+\beta_n) = G(x)$*

Lemma 7.4. *A distribution G is max-stable iff $\exists\{F_n\}_n$ a sequence of distribution functions, $\{\alpha_n > 0\}_n, \{\beta_n \in \mathbb{R}\}_n$ are sequences of normalizing constants s.t $F_n(\alpha_n x + \beta_n) \to G$ and $\forall k = 1, 2, \cdots$, $\lim_{n\to\infty} F_n(\alpha_{nk} x + \beta_{nk}) \to G^{1/k}(x)$ where $G, G^{1/k}$ are non-degenerate.*

Proof. For the necessary condition, assume that $F_n(\alpha_{nk} x + \beta_{nk}) \to G^{1/k}(x)$ and $G^{1/k}$ is non-degenerate $\forall k = 1, 2, \cdots$, then we apply Remark 7.3.1 to $G, G^{1/k}$

$$\implies \lim_{k\to\infty} G\Bigg(\underbrace{\frac{\alpha_{nk}}{\alpha_n}}_{\tilde{\alpha}_k} x + \underbrace{\frac{\alpha_n(\beta_{nk} - \beta_n)}{\alpha_{nk}}}_{\tilde{\beta}_k} \Bigg) = G^{1/k}(x)$$

$$\implies G^k(\tilde{\alpha}_k x + \tilde{\beta}_k) = G(x)$$

For the sufficient condition, suppose G is max-stable. Consider $\{F_n := G^n\}$, by the definition of max-stability, $\exists\{\alpha_n\}_n$, $\{\beta_n\}_n$ s.t $\lim_{n\to\infty} G^n(\alpha_n x + \beta_n) = \lim_{n\to\infty} = G(x)$. Moreover,

$$F_n(\alpha_{nk} x + \beta_{nk}) = \big(G^{nk}(\alpha_{nk} x + \beta_{nk})\big)^{1/k} \to G^{1/k}(x)\,,\ \forall k = 1, 2, \cdots$$

□

7.1.1 Limiting distributions of extreme order statistics

Def 7.1.3 (Order statistics)**.** *Let $\{X_i\}_{i=1}^n$ be a sequence of i.i.d random variables, then the sequence of its order statistics is $\{X_{(i)}\}_{i=1}^n$ where $X_{(j)} < X_{(k)}$ $\forall 1 \le j < k \le n$. In particular we denote the largest order statistics $M_n = X_{(n)} = \max_{1\le i\le n} X_i$*

Suppose $\{X_i\}_i$ is a sequence of i.i.d random variables, then the distribution of the maximum order statistics is given by:

$$F_{X_{(n)}}(x) = P\big(\cap_{i=1}^n \{X_i \le x\}\big) = F^n(x)$$

Hence we can apply convergence to types to the sequence $\{X_n = X_{(n)}\}_n$.

Lemma 7.5. *Let $\{F_n := F^n\}_n$ be the sequence of distribution functions of extreme order statistics. If $\exists\{\alpha_n > 0\}_n, \{\beta_n \in \mathbb{R}\}_n$ sequences of normalizing constants s.t $F_n(\alpha_n x + \beta_n) \to G$ where G is non-degenerate, then G is max-stable.*

Proof.

$$\lim_{n\to\infty} F^n(\alpha_n x + \beta_n) \to G(x) \implies \lim_{n\to\infty} F^{nk}(\alpha_{nk} x + \beta_{nk}) \to G(x)$$

$$\implies \lim_{n\to\infty} F^n(\alpha_{nk} x + \beta_{nk}) \to G^{1/k}(x)\,, \forall k = 1, 2, \cdots$$

By Lemma 7.4, G is max-stable. □

Lemma 7.6. *Let $\{F_n := F^n\}_n$ be the sequence of distribution functions of extreme order statistics. If $\exists\{\alpha_n > 0\}_n, \{\beta_n \in \mathbb{R}\}_n$ sequences of normalizing constants s.t $F_n(\alpha_n x + \beta_n) \to G$ where G is non-degenerate, then $\exists\alpha(s) > 0, \beta(s)$ s.t $G^s(\alpha(s)x + \beta(s)) = G(x), \forall s > 0$.*

Proof. From assumption we have $F_n(\alpha_n x + \beta_n) = F^{[ns]}(\alpha_{[ns]} x + \beta_{[ns]}) \to G$. Denote $[ns]$ the integer part of the ns, then

$$F^{[ns]}(\alpha_n x + \beta_n) = \big(F^n(\alpha_n x + \beta_n)\big)^{[ns]/n} \to G^s(x)$$

G^s is non-degenerate since G is non-degenerate. Thus we can apply Corollary 7.3 to $G^s =: G_1, G =: G_2$ and obtain

$$G^s\left(\lim_{n\to\infty} \frac{\alpha_n}{\alpha_{[ns]}} x + \lim_{n\to\infty} \frac{(\beta_n - \beta_{[ns]})}{\alpha_{[ns]}}\right) = G(x)$$

The proof is concluded by letting $\alpha(s) := \lim_{n\to\infty} \frac{\alpha_n}{\alpha_{[ns]}}$, $\beta(s) := \lim_{n\to\infty} \frac{(\beta_n - \beta_{[ns]})}{\alpha_{[ns]}}$. □

As a consequence of Lemma 7.5 and 7.6, if non-degenerate distribution function G is max-stable, then $\exists\alpha(s) > 0, \beta(s)\forall s > 0\,,\, G^s(\alpha(s)x+\beta(s)) = G(x)$. In particular we are interested in the unit Frechet distribution $G_*(x) = \exp(-1/x)$, whose max-stability is satisfied by $G_*^s(sx) = G_*(x)$. Next we will show for any r.v $X \sim F$ s.t $F \in \mathcal{D}(G)$, there is a transformation s.t the distribution after the transformation $F_*(x) \in \mathcal{D}(G_*)$.

Theorem 7.7. *Let G be a non-degenerate distribution and $\phi^{-1}(x) := \frac{1}{-\ln(G(x))}$, then*

$$G_*(x) = G(\phi(x)) \text{ and } G(x) = G_*(\phi^{-1}(x))$$

where $G_*(x) = \exp(-x^{-1})$ is unit Frechet distributed.

In addition if G is one of the extremal type distribution and suppose a distribution function $F \in \mathcal{D}(G)$, $V(x) := \frac{1}{1-F(x)}$ and define $F_*(x) := F(V^{-1}(x))$, then $F_* \in \mathcal{D}(G_*)$.

Proof. By Lemma 2.2, $F \in \mathcal{D}(G) \implies n(1 - F(\alpha_n x + \beta_n)) \to -\ln G(x)$

$$\implies \frac{V(\alpha_n x + \beta_n)}{n} \to -\frac{1}{\ln(G(x))} = \phi^{-1}(x) \implies \left(\frac{V}{n}\right)^{-1}(x) \to \phi(x)$$

Next we take the inverse of $\phi^{-1}(x)$ and obtain $\phi(x) = G^{-1}\big(\exp(-x^{-1})\big)$. Take F_* as in the statement, then we notice $\exists\{\alpha_n = n\}_n$, $\beta_n \equiv 0$ s.t

$$\begin{aligned}
\lim_{n\to\infty} F_*^n(nx) &= \lim_{n\to\infty} F^n(V^{-1}(nx)) \\
&= \lim_{n\to\infty} P(M_n \le V^{-1}(nx)) \\
&= \lim_{n\to\infty} P\left(\frac{M_n - \beta_n}{\alpha_n} \le \frac{V^{-1}(nx) - \beta_n}{\alpha_n}\right) = G(\phi(x)) = G_*(x)
\end{aligned}$$

So we have shown $F_* \in \mathcal{D}(G_*)$, the second relation also holds since

$$G(x) = \lim_{n\to\infty} F^n(\alpha_n x + \beta_n) = \lim_{n\to\infty} F_*^n(n \cdot V(\alpha_n x + \beta_n)/n) = G_*(\phi^{-1}(x))$$

□

The consequence of Theorem 7.7 is that for a r.v $X \sim GEV(\mu, \sigma, \gamma)$, the distribution of $\phi^{-1}(X)$ is unit Frechet, where

$$\phi^{-1}(x) = \frac{-1}{\ln(G(x))} = \big(1 + \gamma\frac{x-\mu}{\sigma}\big)^{1/\gamma}$$

7.2 Maximum Likelihood Likelihood estimation

Def 7.2.1 (MLE)**.** *Let $X_1, \cdots, X_n$ be random sample of $F \in \{F_\theta : \theta = (\theta_1, \cdots, \theta_p) \in \Theta\}$, f be the density or mass of F, then the maximum likelihood estimator (MLE) of the parameters theta is defined as $\hat{\theta}_n := \arg\max_{\theta\in\Theta} l(\theta)$, where $l(\theta, x) = \sum_{i=1}^n \ln f(x_i|\theta)$ is the log-likelihood.*

Suppose $\hat{\theta}_n$ is the MLE, it can be shown that for large number of sample, we can instead maximize $E(l(\theta, X))$. Note that the expectation is taken under X. The following theorem states that $E(l(\theta, X))$ attains maximum at $\hat{\theta}_n$.

Theorem 7.8. *Suppose $f(x|\theta)$ is the density of X such that f is continuous w.r.t to θ and under some mild regularity conditions, $l(\theta, x) = \ln f(x|\theta)$ is the log-likelihood of f, then*

$$E\left(\frac{\partial l}{\partial \theta}\right)^2 = -E\left(\frac{\partial^2 l}{\partial \theta^2}\right)$$

Proof. By definition $\int f(x|\theta)\mathrm{d}x = 1 \implies \frac{\partial}{\partial\theta}\int f(x|\theta)\mathrm{d}x = 0 (*)$
Use chain rule we get $\frac{\partial l}{\partial \theta} = \frac{\partial f}{\partial \theta}\cdot\frac{1}{f(x|\theta)} \implies \frac{\partial f}{\partial \theta} = \frac{\partial l}{\partial \theta}\cdot f(x|\theta)$

From $(*)$

$$\int \frac{\partial l}{\partial \theta} f(x|\theta)\mathrm{d}x = 0 \implies \frac{\partial}{\partial\theta}\int \frac{\partial l}{\partial \theta} f(x|\theta)\mathrm{d}x = 0$$

$$\implies \frac{\partial}{\partial\theta}\int \frac{\partial l}{\partial \theta} f(x|\theta)\mathrm{d}x = \int \frac{\partial^2 l}{\partial \theta^2} f(x|\theta)\mathrm{d}x + \int \frac{\partial l}{\partial \theta}\frac{\partial f}{\partial \theta}\mathrm{d}x = 0 \implies E\left(\frac{\partial l}{\partial \theta}\right)^2 = -E\left(\frac{\partial^2 l}{\partial \theta^2}\right)$$

□

Theorem 7.8 holds under some regularity conditions. One such condition requires $E(l(\theta, X)) < \infty$, thus we can interchange differentiation and integration under the expectation. Hence $E\left(\frac{\partial^2 l}{\partial \theta^2}\right) = \frac{\partial^2}{\partial\theta^2}E(l(\theta, X)) \leq 0$ implies that the function $H(\theta) = E(l(\theta, X)) \approx 1/n\sum_{i=1}^n \ln(f(x_i|\theta))$ is concave. Hence the MLE $\hat{\theta}_n$ is obtained by solving $\partial H/\partial\theta = 0$, where we can solve $\partial l/\partial\theta = 0$ instead. The arguments for a vector of parameters are analogous. In the multivariate setting we define $I(\boldsymbol{\theta}) := E\left(\frac{\partial l}{\partial \boldsymbol{\theta}}\right)^2$ to be the **Fisher Information Matrix** and the central limit theorem yields that:

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} MVN(0, I^{-1}(\hat{\boldsymbol{\theta}}))$$

In general MLE does not have a closed form. We obtained the MLE by solving the equation $\frac{\partial l}{\partial \boldsymbol{\theta}}$ numerically. For parameter space that has more than one dimensions, the two most common approaches are **Full Maximum Likelihood (FML)** and **Inference For Margins (IFM)**. Suppose $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_p)$, then the MLE through FML is obtained by sovling $\left(\frac{\partial l}{\partial \theta_1}, \cdots, \frac{\partial l}{\partial \theta_p}\right) = 0$ globally. MLE through IFM, on the hand, find the local solutions $\hat{\theta}_1, \cdots, \hat{\theta}_{p-1}$ by solving $\frac{\partial l}{\partial \theta_i} = 0$, $i = 1, \cdots, p-1$. The last estimates is obtained by solving $\frac{\partial l}{\partial \theta_p}\big|_{(\hat{\theta}_1, \cdots, \hat{\theta}_{p-1})} = 0$. Comparing with FML, IFM is more numerically stable but the estimates may deviates from the true parameters.

MLE is invariant under parameter transformations. Meaning that if $\hat{\theta}_n$ is the MLE of θ, $h(\theta) : \Theta \mapsto \mathbb{R}$, then $h(\hat{\theta})$ is the MLE of $h(\theta)$. The asymptotic normality of $h(\theta)$ is also preserved. This is due to the **Delta Method**, in which it is stated that if a estimator $\hat{\theta}_n$ is normal, then $h(\hat{\theta}_n)$ is again normal for some "nice" function h. The variance of Delta method is given in the following theorem:

Theorem 7.9. *Suppose $\hat{\theta}_n$ is the MLE of $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_p)$, $h(\boldsymbol{\theta}) : \Theta \mapsto \mathbb{R}$ is continuously differentiable, then*

$$Var(h(\hat{\theta})) = \nabla h(\boldsymbol{\theta})^T I^{-1}(\boldsymbol{\theta}) \nabla h(\boldsymbol{\theta}) \Big|_{\theta=\hat{\theta}}$$

where $\nabla h(\theta) = (\partial h/\partial \theta_1, \cdots, \partial h/\partial \theta_p)$

Proof. We can expand h in the neighbourhood of $\hat{\theta}_n$ since h is continuously differentiable,

$$\begin{aligned} Var\big(h(\hat{\theta}_n)\big) &= Var(h(\theta) + \nabla h(\theta)^T(\hat{\theta}_n) - \theta) + \wr(\|\theta - \hat{\theta}_n\|)\big) \\ &\approx Var(\nabla h(\theta)^T \hat{\theta}_n) = \nabla h(\boldsymbol{\theta})^T I^{-1}(\boldsymbol{\theta}) \nabla h(\boldsymbol{\theta}) \end{aligned}$$

We can replace all θ with $\hat{\theta}_n$. □

The Delta method yields that:

$$\sqrt{n}(h(\hat{\theta}) - h(\theta_0)) \xrightarrow{d} N(0, \nabla h(\hat{\theta}_n)^T I^{-1}(\hat{\theta}_n) \nabla h(\hat{\theta}_n))$$

And the two-sided confidence interval at level α from Delta method is:

$$I_{h(\hat{\theta}_n)} = \left[h(\hat{\theta}_n) \pm 1.96 \cdot \sqrt{\frac{\nabla h(\hat{\theta}_n)^T I^{-1}(\hat{\theta}_n) \nabla h(\hat{\theta}_n)}{n}} \right]$$

7.3 Extra plots

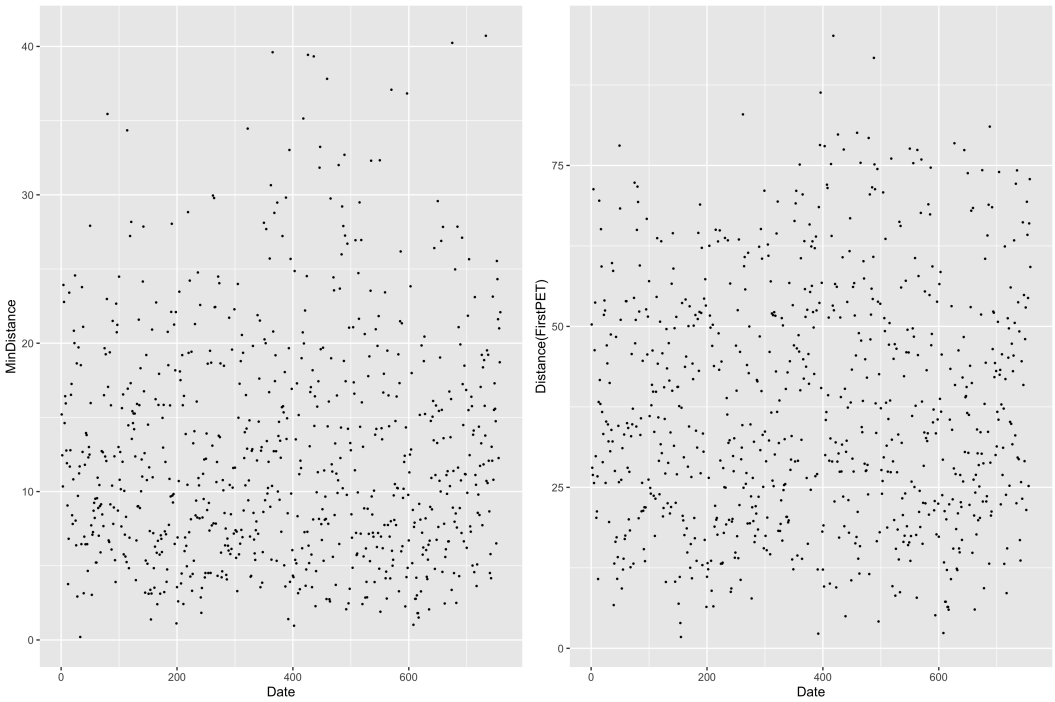


Figure 19: plot of Distance against time (Left: LminD; right: LPETD)

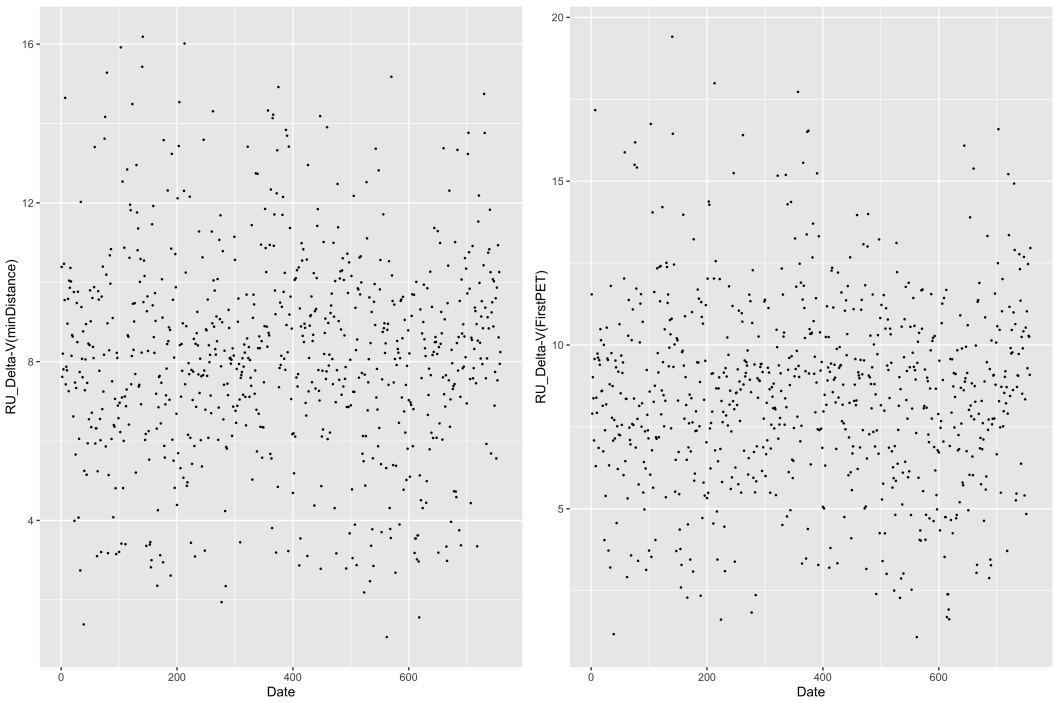


Figure 20: plot of Δv against time (Left: LminD; right: LPETD)

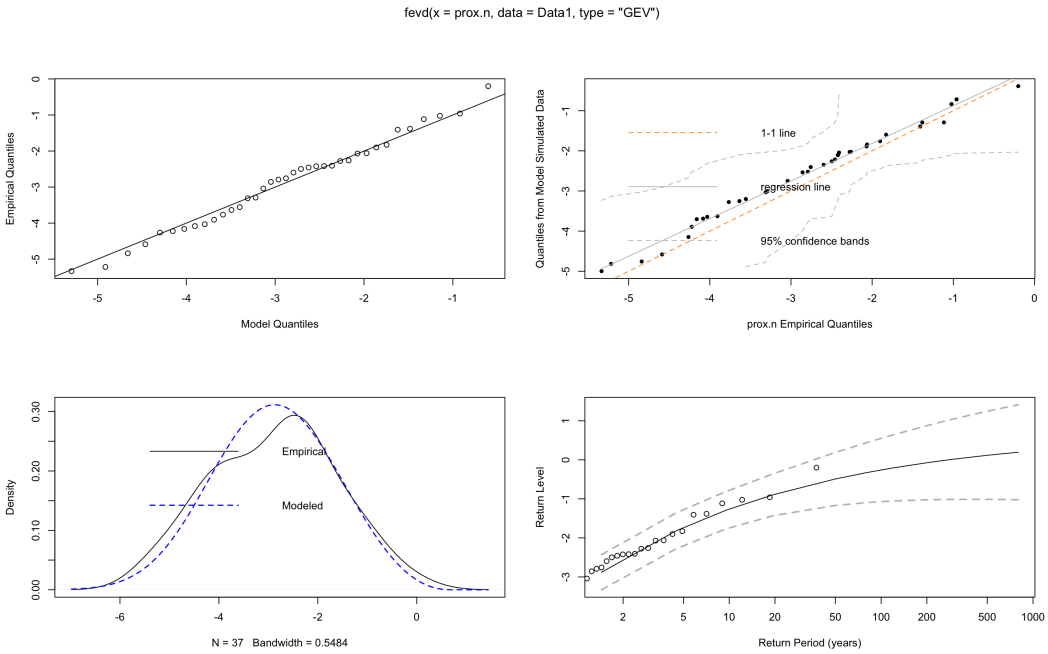


Figure 21: diagnostic plots of temporal SMoS margin in LminD, BM filter

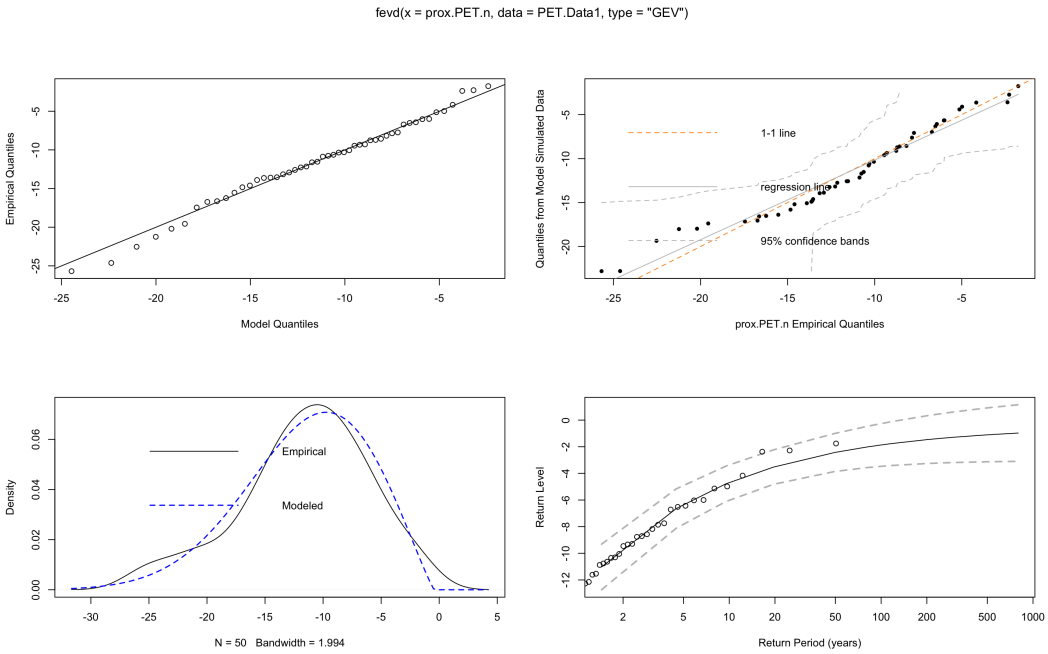


Figure 22: diagnostic plots of temporal SMoS margin in LPETD, BM filter

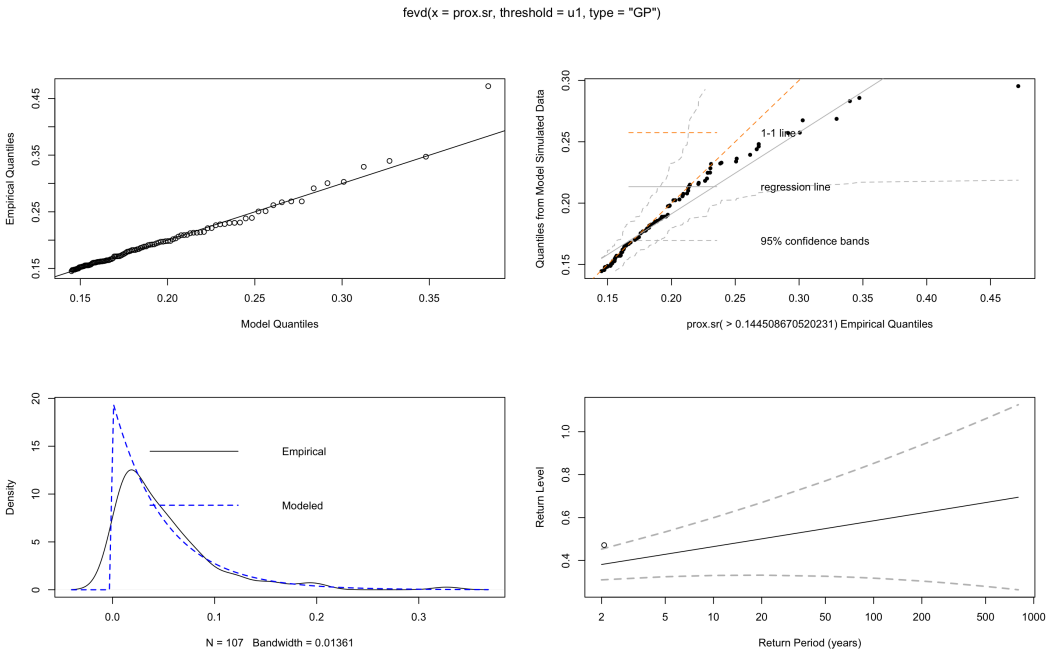


Figure 23: diagnostic plots of temporal SMoS margin in LminD, POT filter

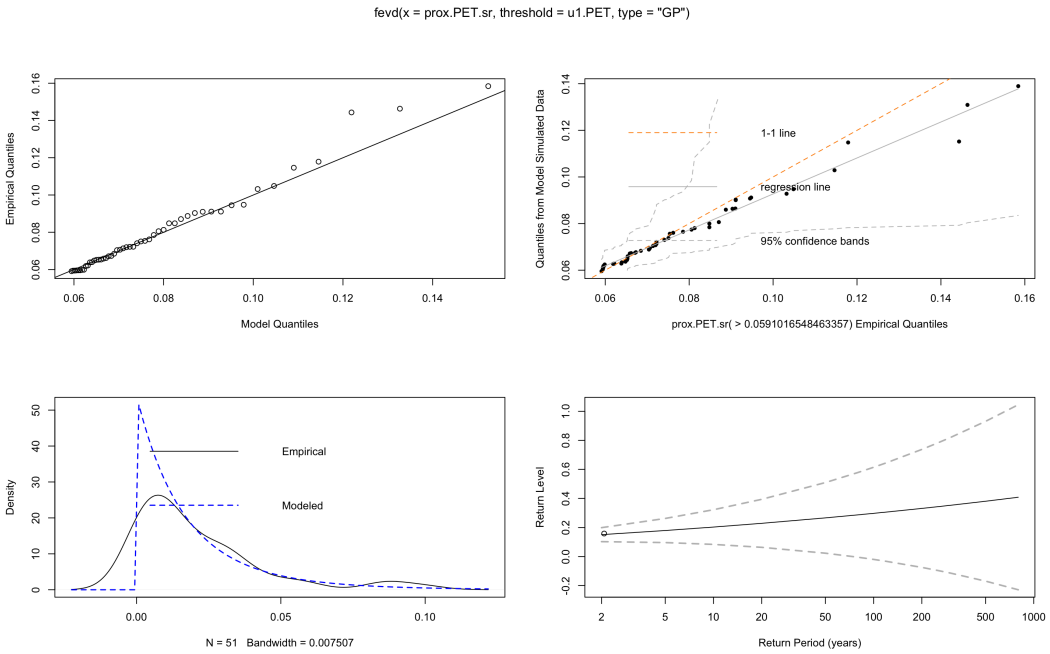


Figure 24: diagnostic plots of temporal SMoS margin in LPETD, POT filter

7.4 R functions

Computing BEVD probabilities for POT1

```
library(evd)
evalProbPOT1 <- function(BPD,x,tail.type){
  # BPD is a fbvpot object, x = (x_1,x_2) is a vector, tail
     = 1 returns dist, =2 return survival function,
  #= 3 is the prob(X<=x_1,Y>x_2), = 4 is the prob(X>x_1,Y<=x
     _2). u1,u2 are thresholds

  param <- BPD$estimate
  scale <- c(param[1],param[3])
  shape <- c(param[2],param[4])
  u1 <- BPD$threshold[1]
  u2 <- BPD$threshold[2]
  x.F <-c(-1/log(1 - BPD$nat[1]/BPD$n*pevd(x[1],scale =
     scale[1],shape = shape[1],lower.tail = F,type = 'GP',
     threshold = u1)),
          -1/log(1 - BPD$nat[2]/BPD$n*pevd(x[2],scale =
             scale[2],shape = shape[2],lower.tail = F,
             threshold = u2,type = 'GP')) )

  if (length(param) == 5){
    #q <- abvevd(x.F =w,dep=param[7],model=BPD$model)
    if (tail.type == 1){
      return(pbvevd(x.F,model = BPD$model, dep= param[5],
         mar1 =c(1,1,1),
                    mar2 = c(1,1,1)))
    }
    else if (tail.type == 2){
      return(pbvevd(x.F,model = BPD$model, dep= param[5],
         mar1 =c(1,1,1),
                    mar2 = c(1,1,1),lower.tail = FALSE))
    }
    else if (tail.type == 4){
      return(pgev(x.F[1],loc = 1, scale = 1, shape = 1) -
         pbvevd(x.F,model = BPD$model, dep= param[5],mar1 =c
         (1,1,1),mar2 = c(1,1,1)) )
    }
    else if (tail.type == 3){
      return(pgev(x.F[2],loc = 1, scale = 1, shape = 1) -
         pbvevd(x.F,model = BPD$model, dep= param[5],mar1 =c
         (1,1,1),mar2 = c(1,1,1)) )
    }}

  else if (length(param) == 6){
    if (tail.type == 1){
```

```
33        return(pbvevd(x.F,model = BPD$model, alpha = param[5],
              beta = param[6],mar1 =c(1,1,1), mar2 = c(1,1,1)))
34      }
35      else if (tail.type == 2){
36        return(pbvevd(x.F,model = BPD$model, alpha = param[5],
              beta = param[6],mar1 =c(1,1,1),mar2 = c(1,1,1),
             lower.tail = FALSE))
37      }
38      else if (tail.type == 4){
39        return(pgev(x.F[1],loc = 1, scale = 1, shape = 1) -
             pbvevd(x.F,model = BPD$model, alpha = param[5],beta
              = param[6],mar1 =c(1,1,1),mar2 = c(1,1,1)) )
40      }
41      else if (tail.type == 3){
42        return(pgev(x.F[2],loc = 1, scale = 1, shape = 1) -
             pbvevd(x.F,model = BPD$model, alpha = param[5],beta
              = param[6],mar1 =c(1,1,1),mar2 = c(1,1,1)) )
43      }}
44    else{
45      if (tail.type == 1){
46        return(pbvevd(x.F,model = BPD$model, asy = param[5:6],
              dep = param[7],mar1 =c(1,1,1),mar2 = c(1,1,1)))
47      }
48      else if (tail.type == 2){
49        return(pbvevd(x.F,model = BPD$model, asy = param[5:6],
              dep = param[7],mar1 =c(1,1,1), mar2 = c(1,1,1),
             lower.tail = FALSE))
50      }
51      else if (tail.type == 4){
52        return(pgev(x.F[1],loc = 1, scale = 1, shape = 1) -
             pbvevd(x.F,model = BPD$model, asy = param[5:6],dep
             = param[7],mar1 =c(1,1,1),mar2 = c(1,1,1)) )
53      }
54      else if (tail.type == 3){
55        return(pgev(x.F[2],loc = 1, scale = 1, shape = 1) -
             pbvevd(x.F,model = BPD$model, asy = param[5:6], dep
              = param[7],mar1 =c(1,1,1),mar2 = c(1,1,1)) )
56      }}
57  }
```

One-level bootstrap KS test for bimodal distribution

```
1  library(mixtools)
2  pmixnorm <- function(x,mixmodel){
3    # distribution function of mixture of two normal r.vs, mu,
       sigma ,lamb are vectors of length 2
4    lamb <- mixmodel$lambda
5    mu <- mixmodel$mu
```

```
6     sig <- mixmodel$sigma
7     return(lamb[1]*pnorm(x,mu[1],sig[1]) + lamb[2]*pnorm(x,mu
          [2],sig[2]))
8   }
9
10  Mixmodelks.test.boot <- function(data,n,mu0,sig0,lamb0,k0){
11    # the bootstrap estimate of ks test statistics. n is the
          number of bootstrap trials; k0 is the test
12    #statistics of ks test when the mixmodel is applied to the
          same the data from which is estimated.
13    N <- length(data)
14    ks.boot <- rep(0,n)
15    mu <- mu0
16    sig <- sig0
17    lamb <- lamb0
18
19    for (i in 1:n) {
20      z <- rbinom(N, 1,lamb)
21      x.boot <- length(z==1)/N*rnorm(N,mu[1],sig) + (1-length(
            z==0)/N)*rnorm(N,mu[2],sig[2])
22      mixmodel.boot <- normalmixEM(x.boot, maxit=500, lambda=
            lamb, mu=mu, sigma=sig)
23      ks.boot[i] <- ks.test(x.boot, pmixnorm,mixmodel.boot)$
            statistic
24      mu <- mixmodel.boot$mu
25      sig <- mixmodel.boot$sig
26      lamb <- mixmodel.boot$lambda[1]
27    }
28
29    return(mean(ks.boot >= k0))
30  }
```

jbowman [https://stats.stackexchange.com/users/7555/jbowman]

Master's Theses in Mathematical Sciences 2022:E47
ISSN 1404-6342

LUNFMS-3110-2022

Mathematical Statistics
Centre for Mathematical Sciences
Lund University
Box 118, SE-221 00 Lund, Sweden

http://www.maths.lu.se/