Lund University

thesis submitted for the degree of Master of science

Centre for mathematical sciences

Analysis and Simulation Study of Stochastic Time-To-Collision as a Severity Measure in Traffic Security

Author:
Per Niklas Waaler

Supervisor:
Nader Tajvidi

May 20, 2020

Contents

Acknowledgements

I would like to thank my supervisor Nader Tajvidi for consistent and reliable support throughout my thesis work, and for giving me the freedom to explore any ideas that I found interesting. I would also like thank Alieksei Laureshyn, Carl Johnsson and Carmelo Dagostino for all their help and cooperation, which was always enthusiastic and friendly.

Abstract

Traffic accidents are extremely rare, creating the need for surrogate methods for safety analysis that makes efficient use of the information provided by traffic conflicts, which also are limited in availability. The severity measure time-to-collision (TTC) in combination with extreme value theory have so far been one of the primary measures used to infer traffic safety levels, but it relies on unrealistic assumptions that results in severity measures that do not always agree well with observed danger. Stochastic TTC has been proposed as an alternative, which replaces the constant velocity trajectories used to define collision course with naturalistic ones, resulting in a distribution of potential TTC values. The main focus of this thesis is to find a way to mathematically model such data. Stochastic TTC is conceptualized within the framework of mixed distributions, and equations allowing for extreme value theory to be used in such a context are derived. A second point of focus is on presenting a new method for estimating the collision probability, allowing for separate estimation of collisions that occurred with and without attempts at evasive action. Also, the effects of data-transforms were investigated in a simulation setting, which proved highly useful in reducing the tendency for underestimation which seems to be a common problem when extreme value theory is applied to traffic data, and overall improving accuracy. Stochastic TTC and the proposed methods were also tested in this simulated traffic environment, which showed that stochastic TTC can work at least as well as regular TTC, but in order for more significant differences to be seen, the simulation would have to contain more variable road user behaviour and more curved trajectories.

Chapter 1

Introduction and theoretical background

1.1 Introduction

Traffic safety research is a field that is concerned with modeling road traffic and identifying relationships between various measurable factors of traffic and rate of accidents in a given traffic location. The task of estimating the expected number of collisions turns out to be a challenging problem, primarily due to the fact that collisions are extremely rare [4], and the fact that the few accidents that do occur sometimes go unreported, according to senior researcher A. Laureshyn at Lund University. Even when it is practically feasible to collect sufficient amounts of crash data to produce a reliable empirical estimate, this reactive approach runs contrary to the main purpose of traffic security research, which is to identify dangerous traffic environments and prevent accidents before they occur. For this reason there is a need to evaluate traffic safety levels based on the more frequently occurring traffic conflicts, which are encounters similar to accidents in that they seem to be caused by similar factors, but do not result in a collision [4]. Research has established that there is indeed a relationship between rates of traffic conflicts and rates of traffic collisions [5]. This suggests that it should be possible to predict the rates of collision based on the rates of conflicts, and much attention has been directed at modeling this relationship. The desire to perform such extrapolation has led to a growing interest in applying Extreme value theory (EVT) [7] to traffic security - an approach made possible due to the apparent continuity along the severity-spectrum of road traffic [6]. One of the advantages to using EVT is that it requires few modelling assumptions compared to other estimation methods, such as the assumption of a fixed conflict to collision ratio [6]. It is supported by mathematical theory that ensures asymptotic convergence under weak assumptions [2]. Also, it allows for severity to be measured by a one-dimensional continuous variable, such as separation in time or space between road users [6].

In order to apply EVT to the problem of estimating crash frequency, a value is assigned to each moment of each traffic conflict, according to some objective rule. This value acts as an indicator of its level of severity, and is known as a **severity measure**, or **safety surrogate measure**. The word surrogate refers to the fact that traffic conflicts acts as a surrogate for observing actual accidents. After a large number of conflicts have been observed, and each has been associated with a value using the severity measure of choice, EVT is then used to model the lower tail of the distribution of the resulting data, which contains information about the frequency of the most dangerous traffic events, including the events corresponding to collisions.

The selection of which severity measure to use is an important aspect of the accident-rate estimation procedure. Two commonly used severity measures are time-to-collision (TTC) and post-encroachmenttime (PET). TTC is defined as the time it would take for two road users to collide under the assumption that they continue moving with constant velocity [4], that is with constant speed and direction. In [4] PET is defined as "the time measured from the moment the first road user leaves the potential collision point to the moment the other road user enters this point". A drawback of PET is that it it fails to capture the severity of the situation when a collision is avoided by a road user breaking hard, since this road user would then have to first accelerate after stopping before crossing the path of the other road user, resulting in a misleadingly high PET value that gives no indication of the danger of the encounter.

TTC does not suffer this drawback, but has instead a disadvantage in that it sometimes assigns a severity rating based on an unrealistic assumption, as road users in many situations cannot by any reasonable model be expected to continue forward with constant velocity. For instance, in an intersection, people will often decelerate out of habit, and have a certain trajectory in mind that they will follow by default should there be no obstacles to react to. In the article [1] the authors demonstrate how this concept fails when the vehicles moves along a curved path, such as in a roundabout. Furthermore, TTC can miss-label certain situations as safe, when in fact they are not. For instance, if two road users pass each other in an intersection with high constant velocities and miss each other only by a few centimeters(for instance as a consequence of not seeing each other), this will be correctly identified by an observer as a dangerous encounter that may very well be indicative of some ill-designed feature of that intersection which contributes to the collision risk, such as a distraction in the traffic environment or an obstacle obscuring vision. Despite this, TTC would indicate no danger, since the drivers were by this definition never on collision course. Therefore, TTC can be somewhat wasteful of the the data available, which is particularly problematic given how rare serious conflicts are.

One way to modify the concept of TTC to remedy this drawback, is to define TTC in terms of the possible intended movements that are likely in a given situation, resulting in a distribution of possible TTC values rather than a single deterministic TTC value. The paper [4] was one of the first to investigate the application of a probabilistic framework to traffic safety analysis. In it, the authors suggests redefining the idea of collision course as an interaction wherein there is a non-zero probability of the interaction resulting in a collision. As an extension of this idea, researches at Lunds Transport and Roads department have considered using the idea of a probability distribution of possible intended paths to define a stochastic severity measure called stochastic TTC, essentially replacing TTC by a probability distribution of potential TTC values. Each possible TTC value is computed by considering every possible combination of potential intended paths for two road users that are in a conflict. The potential intended paths are based on previously observed movements of road users rather than on the assumption of constant movement, and should therefore provide a more realistic indication of a situations safety-level. This approach of discretizing the space of possible future trajectories is similar to that taken in articles [1] and [4].

To the knowledge of the author of this thesis, there has been no previous attempts to combine this sort of data - namely data consisting of a collection of probability distributions rather than a collection of data-points - with EVT. One of the main focuses of this thesis is therefore to derive the equations that allows for this type of data to be used, provide motivation for why such an approach would make sense mathematically, and what could be gained from it, at least in an idealized setting. The thesis also aims to provide proof-of-concept of the presented equations and methods in the form of simulation testing. In this simulation, stochastic TTC is tested in performance against two other severity measures: TTC and minimum-distance, the latter of which is the minimum distance between the hulls of two road-users at time t.

Some attention is also given to the effects of data-transformation, which was inspired by the use of a data-transform $1/(\delta+x)$ in the article [8], where it was used to seemingly beneficial effect by producing fewer zero-estimates of the collision probability. The outcome of obtaining a zero-estimate appears to be a common problem when applying EVT to traffic data (see [5], [6], [7] and [8]), rendering the estimate essentially useless in such cases. This is particularly problematic when the purpose of the collision-rate-estimate is to compare the safety levels of different traffic locations, for instance in order to analyse the impact on safety levels of some aspect of traffic, such as differences in traffic lights, traffic signs or road markings. We will investigate how data transformations could potentially decrease the probability of obtaining a zero estimate.

Finally, we will present what is to the authors knowledge a novel method for estimating the probability of collision using EVT. This approach is based on the notion of TTC-at-first-evasive-action which was first introduced in the article [3], and allows for separate estimation of the collision probability for two types of collision, namely collisions with and without attempts at evasive action. The idea is to separate the set of observed encounters into those with and those without attempts at evasive action, and collect different information from each set. This presents one way to get around an issue associated with stochastic TTC, namely the difficulty of predicting intended movement that occurs when a road user starts engaging in evasive maneuvering.

1.2 Overview of basic extreme value theory

Extreme value theory is a branch of mathematics that is concerned with prediction of the frequency with which rare or extreme events occur. It can be considered a kind of extrapolation method that allows us to make inference about the extreme parts of the probability distribution we have little or few observations. It can be used whenever the data can reasonably be considered to be independent and identically distributed, and sufficiently smooth to justify such extrapolation. This makes EVT useful in applications of accident and disaster prevention, since waiting for sufficient amounts of data to make meaningful empirical estimates is not an option. EVT is widely used in applications where risk management is of importance, such as predicting the risk of floods, risk of structural failure in structural-engineering, or risk of large insurance claims [2].

Two common ways for EVT to be applied are the block maxima and peaks over threshold methods, both of which are motivated by asymptotic arguments. The block maxima method makes use of the fact that the maximum M_n of a sample of size n can - under weak conditions - be well approximated within the three parameter family of distribution functions known as the Generalized extreme value (GEV in short) family. This means that the data can be split up into blocks of size n, after which the the maxima of each block is taken to produce a set of observations of M_n. To these maxima, we may then fit a GEV distribution using the maximum likelihood method. This process involves a variance-bias trade-off, as smaller blocks means more data points (reduction in variance) but also less accurate modelling assumptions (increase in bias) given that n is smaller.

If a random variable X satisfies the conditions allowing for approximation of M_n within the GEV family, then it turns out that the exceedences over a high threshold u are well approximated by the generalized Pareto distribution (GPD in short) [2]. Specifically, if $Y = X - u | X > u$, then

$$P(Y \leq y) \approx 1 - \left(1 - \frac{\sigma}{\xi} y\right)^{-\frac{1}{\xi}} .$$

Note that the above approximation is asymptotically precise, i.e. it tends towards equality as u tends towards the upper endpoint of the distribution of X. The case when $\xi = 0$ is obtained by taking the limit as ξ approaches zero, and yields the exponential distribution with mean $1/\sigma$. When $\xi < 0$, the support of the distribution is $o < y < \xi/\sigma$, and when $\xi > 0$ the support is $0 < y$. A way to estimate the tail of the distribution of X then is to select a high threshold u, collect the data that exceeds this threshold and compute for each exceedence the excess Y, and finally fit a GPD to the excesses. After this is done, we may estimate the probability $P(X > x)$ by using

$$
\begin{aligned}
P(X > x) =& P(X - u > x - u) = P(X - u > x - u | X > u) P(X > u) \\
=& \left(1 - \frac{\sigma}{\xi}(x - u)\right)^{-\frac{1}{\xi}} P(X > u)
\end{aligned}
$$

where $P(X > u)$ can be estimated by counting the number of exceedences, and ξ and σ can be estimated using the maximum likelihood method applied to the excesses. This is the **peaks over threshold** method, or POT in short. Note that the same principle of trade off between variance and bias applies here, as choosing a lower threshold means we get more exceedences and thus more data to estimate the GPD parameters, but also less accurate modelling assumptions.

So far, everything has been formulated in terms of maxima and data points exceeding a threshold, but the methods work equally well for investigating the behaviour of minima and data that falls below a threshold u. To adjust for this, all we need to do is negate the data so that the sample minima becomes the sample maxima, and the above methods can be applied.

In this thesis we will be using only the POT method. This is because we are focusing on application to small data samples, and research has shown that in such cases POT is likely to outperform the block maxima method [7].

1.2.1 Estimation of return levels

Above we described how one may use EVT to estimate the probability of the random variable exceeding a high threshold. Another way to get a sense of the frequency of rare events is to look at the return level associated with some return period m, denoted x_m. x_m is defined in terms of the equation

$$
P(X > x_m) = \frac{1}{m}.
$$

One interpretation of this equation is that the m-observation return level x_m is the level such that it takes on average m observations before it gets exceeded once. Another way to think of it is as the level that on average gets exceeded once per m observations. For example, if we are concerned with the height of a wave, we might want to know the 10 year return level, which is then the wave height that is exceeded on average once every 10 years.

If we are using POT method with threshold u, the m-observation return level is given by

$$
\begin{aligned}
x_m =& u + \frac{\sigma}{\xi}\left[(m\zeta_u)^{\xi} - 1\right], \text{ when } \xi \neq 0 \\
x_m =& u + \sigma \log(m\zeta_u), \qquad \text{when } \xi = 0,
\end{aligned}
$$

where $\zeta_u = P(X > u)$.

1.2.2 Diagnostic plots

As mentioned in the previous section, the selection of threshold represents a trade-off between variance and bias. As we increase the threshold, eventually the reduction in bias is no longer worth the corresponding increase in variance. Likewise, as we lower the threshold, eventually the decrease in variance is is no longer worth the increase in bias. Thus, it is important to have diagnostic methods for getting a sense of which threshold to use in order to get a reasonable balance between the two. An often used method for selecting a threshold, is to estimate the shape and scale parameter for a range of thresholds, and select the lowest threshold for which the model assumptions appears to hold reasonably well. One way to do this is to plot the estimated shape parameter for different thresholds together with confidence intervals, and see for which thresholds the shape parameter looks reasonably constant relative to the uncertainty indicated by the confidence intervals. This is an indicator of model accuracy, because it can be shown that if the excesses are GPD for some threshold u, then increasing the threshold further will not change the shape parameter [2].

Another indicator that the model assumptions are accurate is that the fitted distribution - or model distribution, as it is sometimes called - is close to the empirical distribution function of the data. This can be checked by plotting model and empirical distribution functions or density functions together to see how well they agree. Also, we can estimate the quantiles using the order statistics $x_{(i)}$, and plot these against the model quantiles. The quantile plot consists of the points

$$\left\{ \left(\hat{H}^{-1}(i/(k+1)), x_{(i)} \right), \text{ for } i = 1, 2, ..., k \right\},$$

where

$$\hat{H}(y) = 1 - \left(1 - \frac{\hat{\sigma}}{\hat{\xi}} y \right)^{-\frac{1}{\hat{\xi}}}$$

is the GPD fitted to the k excesses.

1.3 Definitions and modelling assumptions

In this section, the goal is to clarify some of the basic concepts and definitions that will be used frequently throughout this thesis. We will also identify and motivate some of the underlying modeling assumptions that are inevitable when attempting to estimate traffic safety, given the complex nature of road traffic.

Let us begin with formulating the purpose of severity measures. Loosely formulated, the purpose of a severity measure is to objectively measure the potential for an accident to occur in a given situation during an encounter between two road users. In this context, *situation* refers to the state of the traffic encounter at a fixed point in time, and is determined by a set of initial conditions containing information about the state. These initial conditions are what allows the algorithms used in the article [1] to classify each situation, and determine which previously observed situations are similar enough to the current one so that they can provide information about possible future outcomes. The number of initial conditions and what are what they are may vary, but the most important factors include the speed, direction, and position of the road users. Note that initial position may also include information relating to the "past" and "future", such as whether a road user intends make a turn or continue moving forward, and their previous positions. The more conditions that are included, the fewer past observations of "identical" situations will be available.

Mathematically, we define a severity measure to be any function that assigns to each situation a value X in such a way that either $X \leq C$ or $X \geq C$ for some threshold value C is equivalent to the event of a collision. In order for X to be useful in EVT however, we need to impose a further restriction. We need the distribution function of X to be at least continuous, so that extrapolation from less extreme events is justified.

Next, we need to clarify what is meant by "potential for danger". This formulation assumes that an encounter involves some non-zero probability of resulting in an accident, at least if the road users are in some meaningful sense on a collision course. The notion of collision course is an important concept here, and can be defined in several ways that typically rely on some extrapolation hypothesis regarding future movement [4]. In this thesis, we will model traffic by assuming that every driver has at each moment what we will refer to as *intended path* or *default path*, inspired by the article [4] and discussions with A. Lauhrensyn and other researchers at Lunds Transport and Roads department. We define this path as the motion a road user would follow if there were no other road users for them to potentially collide with. Thus, we are assuming that each road user has a "preferred" or "characteristic" way of moving through a given traffic environment, which they are sometimes forced to deviate from due to the presence of other road users. The intended path can be thought of as representing a kind of behavioural momentum, and conscious control and attention must be exerted in order to deviate from this intended path to avoid collision when the intended paths of two road users results in a collision course. A natural assumption to make, is that situations that require a greater amount of conscious decision making in order to avoid an accident are more dangerous than those that require only a small modification of habitual behaviour to avoid an accident.

The above discussion suggests that one component to "danger" in a traffic situation is that there is a collision course, i.e. the road users are in a position where modification of their intended paths is required to avoid collision. Another component is the amount of time available to the road users to take corrective action before a collision occurs, i.e. TTC. Here we are making the additional assumption that the less time available to react to the collision course, the higher the risk for an interaction to result in a collision, all other things being equal. To summarize, danger may be thought of as a function of how much time is available to correct behaviour, and the amount of correction required.

1.3.1 EVT applied to deterministic TTC

A standard way to apply EVT within the context of traffic security, is for each encounter to be reduced down to a single value indicating its degree of danger. To do this, each time-frame may be assigned a severity rating using some severity measure such as TTC, after which we take the minimum of these values as the representative for the encounter as a whole. In other words, the severity level of each encounter is rated on the basis of its most dangerous moment. Once such values have been collected for a large number of conflicts, we may apply for instance the POT method to estimate the probability that the minimum TTC will fall bellow zero, which is equivalent to the event that a conflict resulted in a collision.

1.3.2 Definition of stochastic TTC

The main idea behind stochastic TTC is to define the severity measure of a traffic situation as a probability distribution rather than as a deterministic value. stochastic TTC differs from deterministic TTC in two ways. First, collision course and TTC are defined in terms of trajectories that have been observed, rather than on the assumption of continued constant motion. To determine what the intended path of each road user is likely to be, we look at how road users behave in similar observed situations where they were allowed to move freely. For instance, if we want to estimate the intended path of a left turning vehicle with speed s, direction d, we look at the history of observations of road users that have been observed with such speed, velocity and destination (meaning that the space of possible situations is discretized to produce categories containing more than one observation per situation), but under circumstances where there were no other road users to collide with. This allows us to form an empirical estimate of the distribution of the possible intended paths for road users in such a situation. When the probability distribution has been computed for two road users at some time t, we can compute TTC for each combination of intended paths. This then gives us an empirical estimate of the probability distribution of TTC for that situation, where TTC is defined in terms of intended paths.

The above method seems reasonable when the road users have not yet started interacting. Once they have started interacting however, the idea of predicting intended movement based on the behaviour of road users that are free to move as they want does not make much sense. A way around this is to define stochastic TTC only for the first moment where evasive action occurs, which can be considered as the most dangerous non-interactive part of the encounter. This problem will be discussed in more detail in Chapter 2.

When using deterministic TTC, we reduce the encounter as a whole down to a single value that represents the encounter at its most dangerous moment. In the case of stochastic TTC, it is less obvious how to do this. One suggestion is to use the distribution that is "minimum" in the sense that it supports the lowest TTC value to represent the encounter as a whole. We could attempt to try to compute the distribution of the minimum TTC for the whole encounter, but this raises the question of how to model the dependence between TTC values of each time frame of an encounter. It is not clear how this would be done, even in theory. We will not pursue this question further in this thesis, leave it as an open question to pursue in future research, and choose to get around this problem in a different way which will be explained in Chapter 2.

Chapter 2

Propositions for utilizing and modelling stochastic TTC

2.1 Modeling traffic safety levels as a mixed distribution

In order to model stochastic TTC mathematically, we will consider it in terms of what is known as a mixed distribution. A mixed distribution refers to any random variable X that may be sampled by first drawing a random variable by some random mechanism, and then drawing an observation from the distribution of that random variable. To summarize, we may sample from the distribution of X by following the sampling scheme:

$$\begin{aligned}&\text{draw } Z \text{ from } F_Z\\&\text{draw } X|Z \text{ from } F_{X|Z}.\end{aligned}$$

Another way to think of this, is that we first randomize which distribution to draw from, and then sample from that distribution.

To relate this to the idea of stochastic TTC, Z would correspond the situation. Thus, we imagine that the level of danger in a conflict is the result of two steps of randomization. First, the situation is "drawn" - that is, things like the position, velocity and destination (left turning, right turning etc.) of each vehicle is determined. This then determines a distribution for possible intended movements for each road user, and subsequently also a distribution of possible TTC values. Drawing from this distribution corresponds to the second step of the sampling process outlined above, where we finally obtain a TTC value. To be precise, we need to define which part of the encounter we are referring to when we say that the situation is drawn. The situation may be referring to a predetermined part of a conflict, such as the first moment of evasive action, or to the most dangerous moment (according to some definition) during an interaction. For now we will assume that each encounter can be associated with a single distribution of TTC values, whether it be a distribution of the minimum TTC of the encounter as a whole, or the TTC at some specific part of the encounter.

With this in mind, we return to the general situation. Assume for the sake of simplicity that all random variables are discrete. The situation we want to investigate is the following. Suppose that we can sample from $p_{X|Z}$, and that the p.m.f. of X, $P_X = p_\theta$, has a known expression and depends on an unknown parameter vector θ which we would like to estimate using the maximum likelihood method. Suppose further that the distribution of Z is unknown, and that we have as our data a sample of n

observations of Z: $\boldsymbol{Z} = \{z_1, z_2, ..., z_n\}$. The first question we will investigate is how to make a maximum likelihood estimate of θ which best utilizes the data available to us.

One way to get an estimate is of course to simply generate a sample of X of size n by drawing one sample $x_i^{(1)}$ (the meaning of the index (1) will be become clear later) from each conditional distribution $P_{X|Z_i}$, thus obtaining a sample which we denote $\boldsymbol{X}^{(1)} = \{x_1^{(1)}, ..., x_n^{(1)}\}$. To simplify notation, define $p_{X|i} = p_{X|Z_i}$. We can then plug $\boldsymbol{X}^{(1)}$ into the likelihood function and estimate θ in the usual way. It seems however that we are getting an unnecessarily large variance due to the randomization that occurs in the second sampling step, and that we may reduce this added variance by drawing a second sample $\boldsymbol{X}^{(2)} = \{x_1^{(2)}, ..., x_n^{(2)}\}$ by drawing $x_i^{(2)}$ from $P_{X|i}$, so that we have now observed in total two observations from each conditional distribution $P_{X|i}$. If we treat $\boldsymbol{X}^{(1)}$ and $\boldsymbol{X}^{(2)}$ to be approximately independent, we have that the likelihood function for the combined sample $(\boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)})$ is

$$L(\boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)}|\theta) = \prod_{i=1}^{n} p_\theta(x_i^{(1)}) \prod_{i=1}^{n} p_\theta(x_i^{(2)}).$$

We can continue this procedure, drawing N such samples, which then gives the likelihood function

$$L(\boldsymbol{X}^{(1)}, ..., \boldsymbol{X}^{(N)}|\theta) = \prod_{j=1}^{N} \prod_{i=1}^{n} p_\theta(x_i^{(j)})$$

Since we expect to see variance reduction by increasing the number of samples N, a natural strategy is to let N go to infinity and see if the above expressions converges (after some appropriate normalization) in some sense. With this in mind, we introduce the notation

$$m_{i,j}^{(N)} = \#\text{ times that } x_j \text{ was drawn from } P_{X|i} \text{ after } N \text{ samples drawn.}$$

where x_j is the j'th value of the outcome space of $P_{X|i}$, so that $\sum_{j=1}^{\infty} m_{i,j}^{(N)} = N$, since we have drawn a total of N observations from $P_{X|i}$. Note that this notation assumes that $p_{X|i}$ has the same outcome space for all i. With this notation we can rewrite the likelihood fcn. as

$$L(\boldsymbol{X}^{(1)}, ..., \boldsymbol{X}^{(N)}|\theta) = \prod_{i=1}^{n} \prod_{j=1}^{\infty} p_\theta(x_j)^{m_{i,j}^{(N)}}. \tag{2.1}$$

After normalizing by raising the expression to the power of $1/N$, we get

$$L(\boldsymbol{X}^{(1)}, ..., \boldsymbol{X}^{(N)}|\theta)^{1/N} = \prod_{i=1}^{n} \prod_{j=1}^{\infty} p_\theta(x_j)^{m_{i,j}^{(N)}/N}.$$

Note that $m_{i,j}^{(N)}$ is a sum of N independent Bernoulli trials with probability of success equal to $p_{X|i}(x_j)$, so by the Law of large numbers $m_{i,j}^{(N)}/N$ will converge in probability to $p_{X|i}(x_j)$. Hence we get that

$$\log\left(L(\boldsymbol{X}^{(1)},...,\boldsymbol{X}^{(N)}|\theta)^{1/N}\right) \to \sum_{i=1}^{n}\sum_{j=1}^{\infty} p_{X|i}(x_j)\log\left(p_\theta(x_j)\right)$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{\infty} P(X = x_j|Z = z_i)\log\left(p_\theta(x_j)\right) \tag{2.2}$$

where we see that that the ordinary i'th term of the log-likelihood function gets replaced by a weighted average of the possible values of the i'th distribution.

The above argument can be used to motivate a similar expression in the case when each $P_{X|i}$ is continuous. For instance, one can approximate each $P_{X|i}$ with a discrete distribution $\tilde{P}_{X|i}$ by splitting the support of $P_{X|i}$ up into intervals I_j of length δx, so that the probability of drawing an observation $x|i$ from I_j would be $f_{X|i}(\xi_j)\delta x$ for some ξ_j in I_j, and with the support of $\tilde{P}_{X|i}$ being $\{\xi_1, \xi_2, ...\}$. With these discrete distributions we would then carry out the same argument as before, resulting in a an expression similar to what we got in 2.2, but replacing $p_{X|i}(x_j)$ with $f_{X|i}(\xi_j)\delta x$, and $p_\theta(x_j)$ with $f_\theta(\xi_j)$. Note that $m_{i,j}^{(N)}$ now has the interpretation: number of times that we have drawn ξ_j (i.e. drawn a sample from the j'th interval of $P_{X|i}$) after sampling from $\tilde{P}_{X|i}$ N times. If we let δx tend to zero, we get that

$$\log\left(L(\boldsymbol{X}^{(1)},...,\boldsymbol{X}^{(N)}|\theta)^{1/N}\right) \to \sum_{i=1}^{n}\sum_{j=1}^{\infty} f_{X|i}(\xi_j)\delta x \cdot \log\left(f_\theta(\xi_j)\right)$$

$$\to \sum_{i=1}^{n}\int \log\left(f_\theta(x)\right) f_{X|\theta_i}(x)dx \quad \text{as } \delta x \text{ tends to } 0$$

assuming that we are working with sufficiently "nice" distributions. A more involved argument is necessary to establish this convergence, but here we settle for a heuristic argument. In any case, we are restricted to discrete approximations of continuous distributions in practice.

2.1.1 EVT and Mixed distributions

Now we imagine that we are in the situation where we do not know the distribution of Z but also that we do not have an expression for the distribution of X, and that we want to use the POT method to model the tail of the distribution of X based on a sample of Z. That is, we assume that the tail of X can be modeled by

$$P(X > y + u|X > u) = \left(1 + \xi\frac{y}{\sigma}\right)^{-\frac{1}{\xi}} = f_\theta(y)$$

where $\theta = (\sigma, \xi)$ is the parameter vector we wish to estimate.

The idea is to use the similar reasoning as above to form a likelihood function with which we can estimate θ. The only difference now is that when we re-sample we only include the samples that fall above u in the likelihood function. The expression analogous to 2.1 is

$$L(\boldsymbol{X}^{(1)},...,\boldsymbol{X}^{(N)}|\theta) = \prod_{i=1}^{n}\prod_{\{j|\xi_j>u\}} \left(f_\theta(\xi_j - u)dx\right)^{m_{i,j}^{(N)}}$$

where the notation used is the one explained in the paragraph about continuous distributions. Letting δx tend to zero we obtain

$$\log\left(\boldsymbol{X}^{(1)},...,\boldsymbol{X}^{(N)}|\theta)^{1/N}\right) \to \sum_{i=1}^{n}\sum_{\{j|\xi_j>u\}} f_{X|i}(\xi_j)\delta x\cdot\log\left(f_\theta(\xi_j-u)\right)$$
$$\to \sum_{i=1}^{n}\int_u^\infty \log\left(f_\theta(x-u)\right)f_{X|i}(x)dx \quad \text{as } \delta x \text{ tends to } 0 \tag{2.3}$$

which is the distribution we maximize w.r.t. θ in order to obtain our estimate. Again, we see that the i'th term $\log(f_\theta(x_i-u))$ of the log-likelihood function - which is what we would have if only one sample $\boldsymbol{X}$ was used - is substituted by a weighted sum, with the weights determined by the i'th distribution.

2.2 Numerical example

Suppose

$$Z \sim \text{Beta}(\alpha,\beta), \quad \text{where } \alpha = 3,\ \beta = 1$$
$$X|Z \sim \text{Exp}(Z)$$

so that $1/Z$ is the mean of the conditional distribution of X. The probability distribution function of X is

$$F_X(t) = \int_0^1 P(X\le t|Z=z)\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}z^{\alpha-1}(1-z)^{\beta-1}dz$$
$$= \int_0^1 (1-e^{tz})\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}z^{\alpha-1}(1-z)^{\beta-1}dz$$

which we use to compute $P(X>15) = 0.00178$, $P(X>20) = 0.000750$ and $P(X>25) = 0.000384$.

The estimated values, obtained by maximizing (2.3) w.r.t. θ with $n = 1000$, are $\hat{P}(X>15) = 0.001436$, $\hat{P}(X>20) = 0.000345$ and $\hat{P}(X>25) = 0.000083$. The estimated parameter values are $\hat{\xi} = 0$, $\hat{\sigma} = 3.503814$. These estimates are obtained using a threshold $u = 10$, selected somewhat arbitrarily after inspection of a sample of X.

The corresponding estimates obtained when estimating the probabilities based on a sample of $n = 1000$ obs. of X (using the software in2Extremes, with the same threshold $u = 10$) resulted in the estimates $\hat{P}(X>15) = 0.001531$, $\hat{P} = (X>20)0.000392$ and $\hat{P} = (X>25)0.000100$. The estimated parameter values in this case are $\hat{\xi} = 7.81932e-08$, $\hat{\sigma} = 3.66885$. Note that in both cases the true value of $P(X>u)$ was used in order to estimate these probabilities, as the only intention here was to provide an example showing that we get reasonable results when using 2.3.

2.3 Empirical distribution function for stochastic TTC

We mentioned previously that a way to evaluate goodness of fit is to compare the empirical distribution to the model distribution. If we want to model traffic safety levels using the framework of mixed distributions, how do we define the empirical distribution? We will in this section derive an expression for the empirical distribution function associated with a sample of distributions.

The empirical distribution function of $F_n(x) =$ of a sample $\{x_1, x_2, ..., x_n\}$ is defined by the equation

$$F_n(x) = \frac{\#\{x_i \leq x; \quad i = 1, 2..., n\}}{n}$$

which is equivalent to

$$F_n(x) = P(X^* \leq x) \tag{2.4}$$

where X^* is is a random variable drawn with uniform probability from the sample $x_1, x_2, ..., x_n$. To generalize this definition to the situation where we have a sample of distributions corresponding to $\mathbf{Z} = \{z_1, z_2, ..., z_n\}$, we define X^* to be the random variable obtained by first drawing Z^* with uniform probability from the sample $\mathbf{Z}$, and then drawing an observation from $P_{X|Z^*}$. We may then define the empirical distribution function of the sample by equation 2.4.

Next we consider the empirical distribution function of the excesses above some threshold u, i.e. we want to know the distribution of
$Y^* := X^* - u | X^* > u$. Note that we may draw an observation from the distribution of Y^* by drawing X^* repeatedly until we get an observation that exceeds u, after which we subtract u to get the observation of Y^*. Before deriving the distribution of Y^*, we pause to introduce some notation.

Let $\zeta_i = P(X^* > u | Z^* = z_i)$ for some threshold u arbitrary but fixed. Furthermore, define $H_i(y)$ to be the distribution function of $Y^* | Z^* = z_i$, i.e. the distribution of the excesses when they are obtained by drawing from $P_{X|i}$. We assume that the threshold is low enough so that at least one $\zeta_i > 0$.

Let D be the number of draws before before $X^* > u$ happens for the first time, and let I denote the index of the distribution from which the first exceedence is drawn. Note that the probability of failing to draw an exceedence when drawing an observation of X^* is

$$\begin{aligned} P(\text{fail to draw exceedence}) &= \sum_{i=1}^{n} P(X^* \leq u | Z^* = z_i) P(Z^* = z_i) \\ &= \sum_{i=1}^{n} (1 - \zeta_i) \frac{1}{n} = 1 - \frac{1}{n} \sum_{i=1}^{n} \zeta_i. \end{aligned}$$

Using this result, we get that the distribution of I is given by

$$\begin{aligned} P(I = k) &= \sum_{l=1}^{\infty} P(I = k, D = l) \\ &= \sum_{l=1}^{n} P(\text{fail } l-1 \text{ times to draw exceedence}) P(X^* > u | Z^* = z_k) P(Z^* = z_k) \\ &= \sum_{l=1}^{\infty} \left\{ 1 - \frac{1}{n} \sum_{i=1}^{n} \zeta_i \right\}^{l-1} \zeta_k \cdot \frac{1}{n} = \frac{\zeta_k}{n} \sum_{l=0}^{\infty} \left\{ 1 - \frac{1}{n} \sum_{i=1}^{n} \zeta_i \right\}^{l} = \\ &= \frac{\zeta_k}{n} \cdot \frac{1}{1 - \left\{ 1 - \frac{1}{n} \sum_{i=1}^{n} \zeta_i \right\}} = \frac{\zeta_k}{\sum_{i=1}^{n} \zeta_i} \end{aligned}$$

from which it follows that

$$
\begin{aligned}
P(Y^* \leq y) &= \sum_{k=1}^{n} P(Y^* \leq y | I = k) P(I = k) = \sum_{k=1}^{n} H_k(y) P(I = k) \\
&= \sum_{k=1}^{n} H_k(y) \frac{\zeta_k}{\sum_{i=1}^{n} \zeta_i}.
\end{aligned}
$$

This equation shows that the empirical distribution function of the excesses is a weighted sum of the distribution functions of the excesses for each distribution $P_{X|Z_i}$, with the i'th weight proportional to the probability of threshold exceedence for the i'th distribution. Note that if $P_{X|z_i}$ is a discrete distribution that places equal probability mass on the points $\{x_1, x_2, ..., x_m\}$, then

$$
H_k(y) = \frac{\#\{u < x_i \leq u + y; \quad i = 1, 2..., m\}}{\#\{u < x_i; \quad i = 1, 2..., m\}}.
$$

.

2.4 Data-separation method: A novel way of estimating the collision probability

We mentioned earlier that a difficulty when using stochastic TTC is the problem of predicting intended movement when the road users have started interacting. One way to get around this, is to compute stochastic TTC only at the moment of first evasive action, which will sometimes be referred to as FEA. In this section we will investigate which quantity such a random variable can be used to estimate. Furthermore, we will introduce the concept of maximum-danger-during-evasive action, which we will show to also be of potential use.

To proceed with this idea, we will need to introduce some notation. An encounter will be classified as either interactive or non-interactive encounters, IE and NIE in short. Interactive here means that either evasive action was taken or a collision occurred, and non-interactive that there was no evasive action and no collision. Furthermore, an encounter will be classified as type EA (evasive action) if evasive action was taken during the encounter, and type NEA (no evasive action) if no evasive action was taken. Defining these two sets of categories might seem a little excessive as they only differ in their classification when a collision with no evasive action happens, but they are defined in order to avoid ill defined mathematical expressions. With these definitions, we proceed to define some random variables (safety-surrogate-measures) that are defined in terms of the above categories.

Firstly, for type IE encounters, we make use of the notion of danger-at-first-interactive-action, in short DAFIA (note that in the following, "danger" refers to any severity measure, such as TTC). This random variable measures - for each interactive encounter - the danger at the first moment of interactive action. In practical application this moment is taken to be the last frame before interaction occurs, or the first frame for which there is collision. Note that the first interaction can either be an attempt at evasive action, or a collision, in other words it is the first behavioural or physical interaction.

Secondly, for encounters of type EA, we introduce maximum-danger-during-evasive-Action, in short MDEA. MDEA takes its value at the moment of maximum danger over the duration of time for which evasive action was attempted. This value is recorded by observing each time-frame of a conflict for which at least one road user is actively attempting to avoid the other, recording the severity measure at each such frame, and finally taking the minimum (if using say TTC) of those values. In other words, we are recording the maximum degree of danger reached during the attempt of (at least one of) the road users to resolve the conflict, given that there was such an attempt.

Finally, for encounters of type NEA, we define MDNEA (short for maximum-danger during encounter of type NEA) simply as the maximum level of danger over the duration of the encounter.

Now we consider how these measures can be used to estimate the probability of an encounter resulting in collision, which we will simply denote with P(C). Note that P(C) = P(C,EA) + P(C,NEA), so the task of estimating P(C) can broken down into the tasks of estimating each term.

Having broken the problem down into estimating two terms, lets see how the different types of measurements can be used to estimate them. For simplicity, assume that we are using TTC as severity measure so that TTC $\leq$ 0 is equivalent to the event of a collision. First we consider the task of estimating the probability of collision in which neither road user reacted to the other, i.e. P(C,NEA). Note that the event "C, NEA" $\Leftrightarrow$ "MDNEA$\leq$ 0, NEA", hence

$$\text{P(C, NEA)} = \text{P(MDNEA}\leq 0\text{, NEA)} = \text{P(MDNEA}\leq 0 | \text{ NEA)P(NEA)}.$$

P(MDNEA$\leq$ 0| NEA) can be estimated by applying EVT to the sample of MDDNEA coming from the encounters of type NEA, and P(NEA) is estimated by counting the number of encounters for which there was evasive action.

Another way to estimate P(C, NEA) comes from observing that "C, NEA" $\Leftrightarrow$ "IE, DAFIA$\leq$0", hence

$$\text{P(C, NEA)} = \text{P(DAFIA} \leq 0\text{, IE)} = \text{P(DAFIA} \leq 0 | \text{ IE)P(IE)},$$

where P(DAFIA $\leq$ 0| IE) can be estimated by applying EVT to the sample of DAFIA. We see that both MDNEA and DAFIA can be used to estimate P(C, NEA); the estimate based on DAFIA uses data from encounters with interaction, and the estimator based on MDNEA uses data from encounters with no interaction. Using only one of them would arguably be wasteful as we would be using data from only one type of encounter. One way to deal with this - if reasonable amounts of data exists for both types of encounter - is to take as our estimate a weighted average of both estimates, where the weights are based on how confident we are in each estimate, for instance by letting them be equal to the fraction of the data that is of type IE and NIE respectively. Another approach is to focus our efforts only on encounters of type IE. The potential downside of this approach is that we might potentially exclude interesting encounters.

To estimate P(C, EA) we can note that "C, EA" $\Leftrightarrow$ "MDDEA$\leq$0, EA", hence

$$\text{P(C, EA)} = \text{P(MDDEA}\leq 0\text{, EA)} = \text{P(MDDEA}\leq 0 | \text{ EA)P(EA)} ,$$

and as above we can estimate P(MDDEA$\leq$ 0| EA) by applying EVT to the sample of MDDEA.

An alternative way to estimate P(C, NEA) is to combine DAFIA and MD into one random variable

$$Z = \begin{cases} \text{DAFIA}, & \text{if encounter is of type EA} \\ \text{MDNEA}, & \text{if encounter is of type NEA.} \end{cases}$$

This random variable is defined for all types of encounters, and $Z \leq 0 \Leftrightarrow$ C, NEA. To see this, note that $Z \leq 0$ certainly implies collision, so to show "$\Rightarrow$" we only need to show that $Z \leq 0 \Rightarrow$ NEA. If the encounter was of type EA then Z = DAFIA, which would then have to be greater than zero in order for the encounter to be classified as type EA, which is a contradiction. "$\Leftarrow$" follows immediately since $C \Rightarrow Z \leq 0$.

A concern which may arise when applying EVT to a random variable defined this way is that the distribution function may no longer have properties that allows for EVT to be applied. However, as long as the DAFIA and MD have smooth distribution functions, the distribution function of Z will be smooth as well. This follows since $P(Z \leq x) = P(Z \leq x|EA)P(EA) + P(Z \leq x| NEA)P(NEA)$, which is a linear combination of smooth functions.

Another concern is that if MDNEA and DAFIA have distributions that are significantly different in the location of their distributions, then this could make the tail too irregular for EVT to be useful. As an example to illustrate this, consider what happens when we take a weighted average of the probability densities of the distributions exp(1) and N(10,1), so that we sample from each with say probability 1/2. The resulting probability density would have a "hump" at x=10, so to get a decent GPD fit to the tail we have to select a threshold larger than 10; thresholds lower than that would result in too much bias. If we restrict ourselves to a sample of size of say 200, and we select a threshold u=11, it is likely that all observations falling above the threshold would come from the normal distribution, and in terms of the above discussion this would be analogous to utilizing only one type of data.

Chapter 3

Traffic encounter simulation experiment

As a first test of the performance of stochastic-TTC as a safety surrogate measure, we constructed an algorithm which simulates traffic encounters between two vehicles heading in opposite directions. This is intended to give us an idea of well stochastic TTC could potentially perform on real life data relative to other methods. Furthermore, we investigate the effects of data-transformations to see if they can improve performance, and if so, how might one predict such performance improvements based on various diagnostics. finally, we want to compare the performance of the data-separation method described in the previous section in performance against the method of taking minimum-TTC for each encounter.

The simulator is not intended to be hyper-realistic, but the vehicles are programmed to act somewhat reasonably, for instance by trying to avoid collision once a dangerous situation has occurred by turning away from the predicted path of the other vehicle. The simulation also implements a rudimentary form of momentum, both in terms of speed and direction, so that late-detection events have a higher risk of collision than early-detection events. Note that an encounter in this setting is defined simply as two vehicles passing each other by, regardless of its level of danger.

3.1 Description of simulation

A simulated encounter begins with both vehicles being randomly assigned initial positions in a co-ordinate system. Specifically, the initial y-components are drawn from a normal distribution while the initial x-components are fixed. With each iteration - or time-frame, as they will be referred to - they are allowed to take one step, the length of which is sampled by adding some random noise to the length of the previous step. The step is taken in a direction given by a step-angle sampled in a similar way, thereby adding a kind of momentum to the behaviour of the vehicles. Both step-angle and step-size are sampled in a way that discourages odd behaviour, such as a vehicle moving straight upwards or downwards (corresponding to a car leaving the road) or otherwise deviate unrealistically from horizontal movement. The initial step-sizes and step-angles are sampled from gamma and normal distributions respectively.

At any point in time the encounter can be in one of two states: regular, where the the vehicles move freely without consideration of the other, and collision-avoidance-mode where they modify their behaviour in order to avoid collision, with the amount of behaviour modification being a function of factors indicating closeness to collision. For each moment in time, the algorithm rolls a random number to determine if collision-avoidance-mode should be initiated. The likelihood of this state being entered is determined by a probability which depends on both total separation distance and vertical separation distance between the vehicles in such a way that it goes to 1 as either of these approach 0. The reason for the latter variable being included is that as long as the movement of the vehicles is approximately parallel to the horizontal axis, the vertical distance can be used to discourage unnecessary and unrealistic evasive action, intended to mimic how two cars heading in opposite directions may be quite close to each other in terms of absolute distance, and yet not taking any evasive action. It is of course not very realistic that both vehicles always start engaging in evasive action at the same time, but for the sake of simplicity we decided to stick to this simplification.

The moment the vehicles has entered collision-avoidance-mode, the simulation of the encounter pauses to simulate a fixed number of random walks using each vehicle's current position, speed, and direction as initial values. The walks are simulated according to the rules of non-evasive behaviour, i.e. without attempting to avoid collision. For each simulated walk, TTC is computed. If no collision occurs during a simulation, TTC is set to ∞. If collision already has occurred, no walks are simulated (in this case the TTC-distribution is a point mass, since only one thing can happen), and TTC is computed by "reversing" time to the first moment of contact, with the collision then being indicated by a TTC value less than or equal to 0. In addition to sampling TTC, the minimum distance between the hulls of the vehicles is computed and stored as an alternative safety surrogate measure for that time-frame. This severity measure will be referred to as **minimum-distance**, or separation-distance when necessary to avoid confusion.

After collision-avoidance-mode has been engaged, the minimum distance and TTC is computed for each time-frame until the vehicles have either collided, or they have passed each other. The vehicles are defined to have *passed* when the x-coordinates of the vehicles have changed order, and this will also be used as the criteria for conflict resolution. After the vehicles have resolved the conflict or collided, the minimum value of TTC and SD of the interaction-frames are computed and stored.

Other than computing severity measures at FEA and during the attempt to avoid collision, the algorithm also computes minimum TTC and minimum separation distance for the encounter as a whole.

3.2 Design of simulation experiment

The experiments for testing the performance of stochastic-TTC and various methods takes the form of simulating 500 data sets of encounters, where each set consists of 500 encounters. For each such data set we estimate the collision probabilities and other quantities of interest. Some of the statistical properties of the estimators are investigated, and their accuracy judged by comparing with empirical estimates that are computed using data from 2 million simulated encounters.

In order to test the ability of different estimators on their ability to compare and rank-order the safety levels of different traffic locations - which we will refer to as ability to infer **relative safety levels** - we will compare estimates from two different simulated environments. One of these represents the more dangerous traffic location, as it is set to have a slightly lower probability for the vehicles to engage in evasive action. For each sample of encounters, the estimated collision-probabilities are used to predict which intersection has the lower collision-frequency. Then the fraction of correct and incorrect guesses is computed, and this fraction will be referred to as the **success rate** and **fail rate** respectively.

3.3 Indicators of performance

An important issue to discuss when talking about performance is what constitutes good performance, and how to quantify it. For reasons discussed in a previous section, a very basic requirement for an estimator to be said to perform well is for it to have a low probability of yielding a zero-estimate. This aspect of an estimator will be measured by looking at the percentage of estimates that are greater than or equal to zero, which will be denoted by P(NZE), or referred to as the **NZE-rate**. We would of course like the estimator to be accurate and reliable beyond just producing non-zero estimates. The Root Mean square error (RMSE) is an often used measure of accuracy, but it will not be used here as the method of performance comparison. This is because it does not sufficiently penalize estimators that frequently underestimate the collision probability, or estimates it to be zero. To illustrate this point, consider an estimator that always returns zero as the collision estimate. This estimator will have a RMSE equal to the probability itself, which - given that the probability is so small - will result in this estimator appearing to be quite good relative to more reasonable estimators, despite the fact that it is obviously useless for practical purposes. Also, the error of overestimating danger is a less serious error than the error of underestimating it, given that the purpose is to prevent serious accidents from occurring. For this reason, the measure of accuracy that will be used is the probability that the estimate will deviate from the true value by less than a given percentage - which will be referred to as the **cut-off value**. That way, estimators that often return zero-estimates are penalized more heavily. We will be using a cut-off value of 50%. This is somewhat arbitrary, but note that different cut-of values were tested to ensure that the conclusions are not sensitive to the particular choice of cut-off value. This measure of accuracy will be referred to as the **accuracy rating**.

Another measure of performance comes from the desire to be able to compare and order the safety levels of different traffic locations. To this end, we will look at the probability of correctly inferring which out of two simulated traffic locations is the safest based on the estimated collision probabilities, as this is one of the intended uses of EVT in traffic security. Note that the result might be indeterminate, as will be the case when both probabilities are zero. As a method of comparison, we will also use return levels as a basis for making this inference. This is motivated by a suggestion made in the article [7], where the authors refer to the lower variance as a reason why return levels might be a better tool for comparing safety levels, and for estimating safety levels in general.

In addition to scoring high on the above mentioned measures of performance, we want estimators that are robust with regards to the subjective choices that go into the estimation process. An estimator that is less sensitive to choice of threshold will be considered better, when all other things are equal. Also, it is important that a method that performs well can actually be identified without knowing the true value of the estimand, since it is possible for a method to perform well essentially by chance. For instance, an estimator of a distribution function could fit poorly over large parts of the support, but may never the less have zero bias at one particular point if the expected value of the estimator and the true value of the distribution function intersect at some point. These are all much more subjective indicators of performance than the above mentioned ones, but they are still important to consider. A strong performance result is not very meaningful unless we can extract from it some rule for how to make good modelling decisions under circumstances where we do not know the correct answers. The primary methods for evaluating goodness of fit will be visual inspection of qq-plots, plots of empirical vs model distribution functions, and shape-parameter stability plots.

3.4 Data transforms

The data transforms used in this thesis are

$$T_{exp}(x) = \exp(-p(x-2))$$
$$T_{inv}(x) = \frac{1}{(x-3.5)^p}$$

which will be referred to as the **exponential** and inverse transform respectively. The location parameters are admittedly fairly arbitrary, and were arrived at through after some brief experimentation. The parameter p will be referred to as the **power parameter**.

3.5 Selection of power-parameter and threshold

Threshold selection plays an important role in EVT, and it often relies on choices that are inevitably going to be somewhat subjective. Choices are often based on visual inspection of diagnostic plots indicating goodness of fit and validity of model assumptions. Manually selecting a threshold in each and every case would be both too time-consuming and too subjective. Therefore, in order to compare performance, all estimates are computed over a range of 10 thresholds that are selected to cover the thresholds that would be considered relevant in practice. The thresholds are selected so that the lowest (after negation of the data) threshold includes about 80% of the data, and the highest includes about 6% of the data. The remaining thresholds are spaced equidistantly between these two thresholds. For the data-transformation methods, all the thresholds are transformed as well, so that each threshold uses the same amount of the data regardless of the transformation used. When the data is of the form stochastic TTC, the minimum of each encounter is selected as the data-point to represent each distribution, after which the above rule is applied to the minima to select the range of thresholds.

Similar considerations have to be made in terms of which transformation parameter values to include in the analysis. We want to include the parameters that yield peak performance for each transformation type, and also include the values that allows us to capture the overall effects of increasing and lowering the transformation parameter. With this method it is not possible to completely eliminate all dependence on subjective choices, but it is still possible to get an overall sense of how the parameter and transform choice influences the estimates. One more thing to note on this topic is that for each transformation type, the location parameter is fixed, and only the power parameter is allowed to vary. This of course opens up the possibility for other combinations of parameter values to perform better than those covered here, but due to the large increase in complexity that this would entail, we settled for varying only the power parameter.

Chapter 4

Results and discussion

4.1 Results

The probability P(C, NEA) and P(C,EA) were empirically estimated (with 95% confidence intervals) to be 5.35e-5 ± 1.0137e-5 and 1.265e-4 ± 1.5587e-5 respectively. P(C) is estimated to be 1.8e-4 ± 1.8592e-5. The simulated environment with higher safety levels, P(C) was empirically estimated to be 3.3800e-4 ± 2.5475e-5.

Figure 4.9 and 4.10 shows accuracy plots comparing the data separation method against the benchmark method of taking minimum TTC over each encounter, with and without use of inverse data transform. For this comparison, deterministic TTC was used as severity measure. Without the use data transform, the data separation method significantly outperformed minimum-TTC, by as much as an $\approx$ 18% difference. After the transform, the benchmark method has a higher peak performance - a difference of about 13%.

In Figure 4.1 we can see from the accuracy plots that stochastic and deterministic TTC seems to have negligible difference in performance, with regards to accuracy. Separation distance has a significantly higher peak performance than the other two severity measures, with a peak accuracy rating at about 12.5%, which is roughly twice the peak performance of the other two severity measures. The difference in performance is largest at the lower thresholds, and for higher thresholds the difference between all three severity-measures becomes small.

In Figure 4.2 and 4.3 we can see that applying either data transform can very significantly improve the performance of the estimator when using stochastic TTC. The plots also suggest that the inverse transform is more suitable than the exponential transform in this case, as the peak performance for the parameters considered is approximately 20% and 37% respectively, which is an increase in peak performance by approximately 14% and 31% relative to that of the untransformed data.

In Figure 4.4 and 4.5 we can see that the transforms are not as effective at improving accuracy when using minimum distance as they were in the case of either TTC measure. Some improvement can be argued for in the case of the inverse transform, as the peak performance there was 20.8%, to be compared against the peak performance of 12.6% when not using a transform. For the exponential transform, in particular when using the higher power parameters, there is a performance improvement at the lowest thresholds, but possibly also a slight decrease at some of the higher thresholds. Note that for minimum distance, the lowest threshold corresponds to using 85% of the data rather than the 80% that was used in all other cases, to be able to better capture the peak performance.

Figure 4.6, 4.7, 4.8 and 4.11 and 4.12 shows the effects of increasing the power parameter on expected value, standard deviation and P(NZE), where the transformations are applied to stochastic TTC at first evasive action. By increasing the power parameter, we increase the expected value - and eventually bias - as well as the NZE rate. The expected value appears to grow faster as the power parameter increases, whereas the NZE rate grows more slowly as the power parameter increases. For the the exponential transform, the standard deviation increases with the power parameter. We see a similar behaviour for the inverse transform, but for threshold 8 and higher we see a reversal of this trend, with higher power parameter resulting in slight decrease in standard deviation.

In Figure 4.18, 4.19 and 4.20 shows the success and failure plots based on estimated collision probability and return levels respectively. In each case the inverse transform was used. In Figure 4.18 we see that the use of the data-transform significantly improved the rate of correct guesses, by as much as much as 60%. In Figure 4.19 and 4.20 we see that increasing the return rate results in significant reduction in prediction success, and that when using a return period of 10, we reach near 100% prediction success rate.

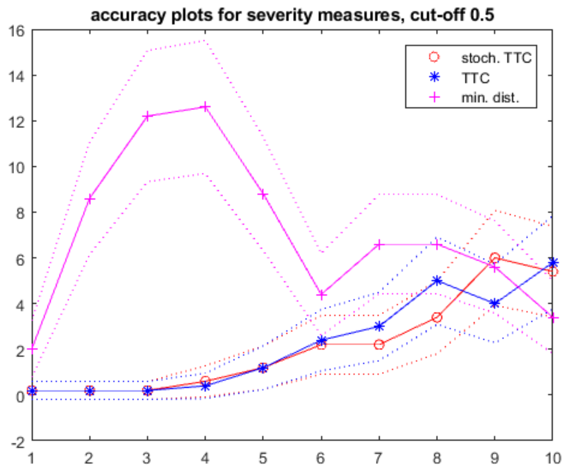

Figure 4.1: Comparison of accuracy of estimators using different severity measures at FEA, without any transforms used. The plots indicate how well each severity measure performed in terms of estimating P(C, NEA).

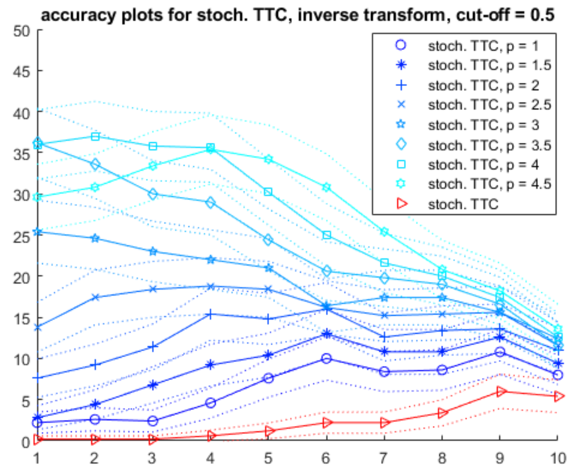

Figure 4.2: Accuracy plots of estimators that use the inverse transform applied to stochastic TTC at FEA for various choices of power parameter. The plots indicate performance in terms of estimating P(C, NEA).

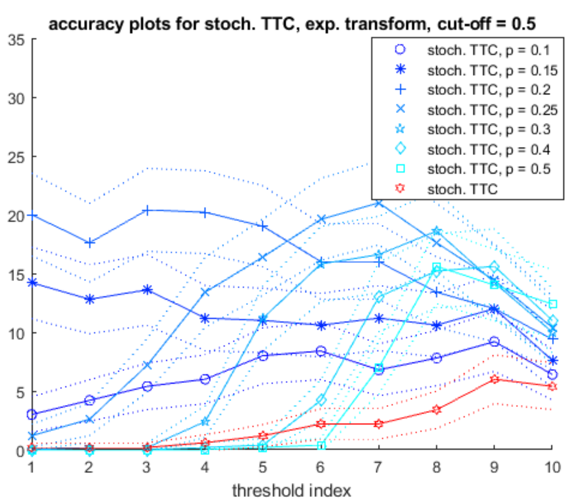

Figure 4.3: Accuracy plots of estimators that use the exponential transform applied to stochastic TTC at FEA for various choices of power parameter. The plots indicate performance in terms of estimating P(C, NEA).

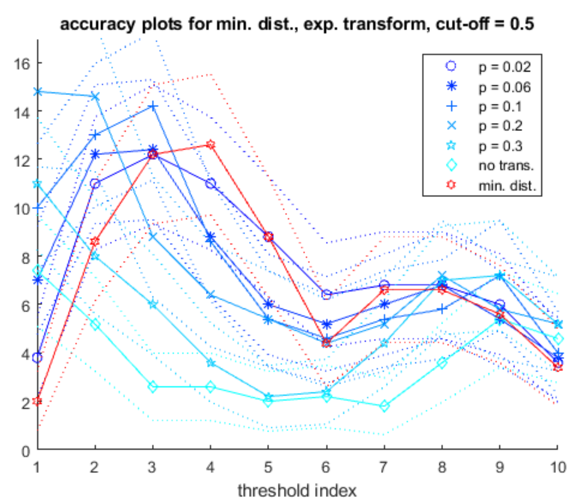

Figure 4.4: Accuracy plots of estimators that use the exponential transform applied to minimum distance at FEA for various choices of power parameter. The plots indicate how well each severity measure performed in terms of estimating P(C, NEA).

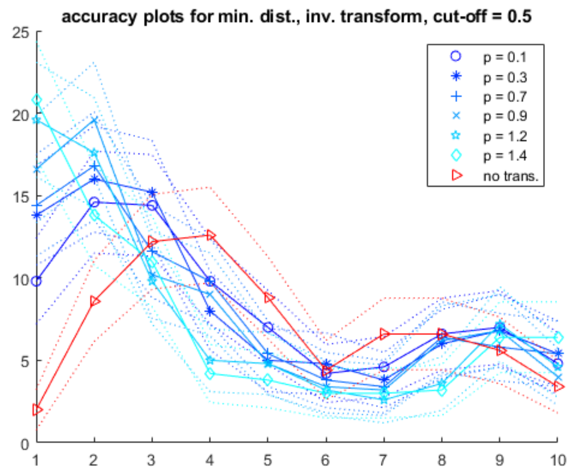

Figure 4.5: Accuracy plots of estimators that use the inverse transform applied to minimum distance at FEA for various choices of power parameter. The plots indicate how well each severity measure performed in terms of estimating P(C, NEA).

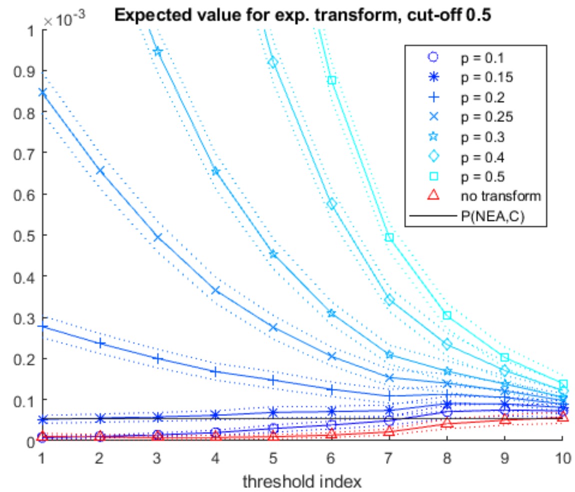

Figure 4.6: Mean value plots of estimators that use the exponential transform applied to stochastic TTC at FEA for various choices of power parameter.

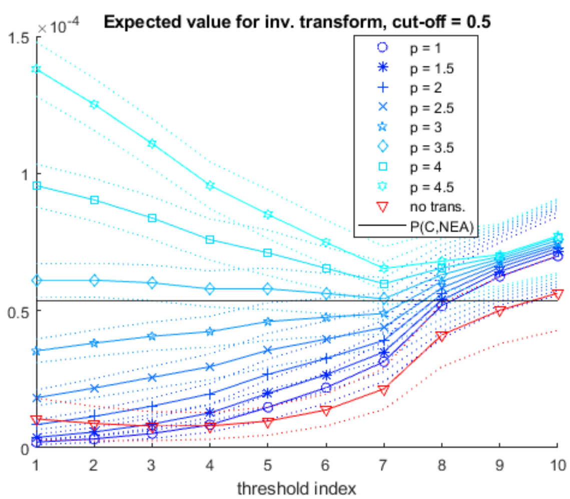

Figure 4.7: Mean value plots of estimators that use the inverse transform applied to stochastic TTC at FEA for various choices of power parameter.

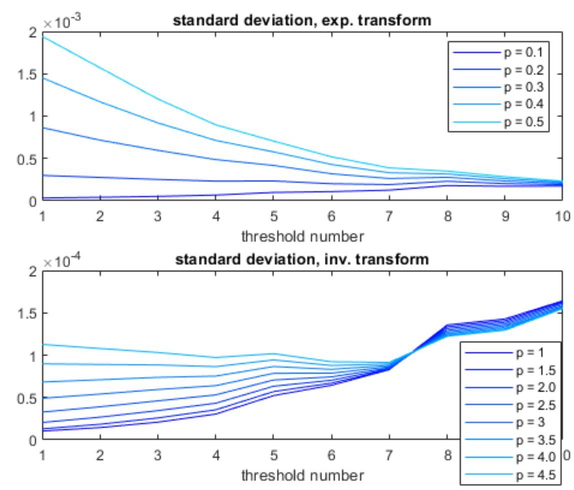

Figure 4.8: Plots shows estimated standard deviation of estimators that use the exponential transform (upper) or inverse transform (lower) applied to stochastic TTC at FEA for various choices of power parameter.

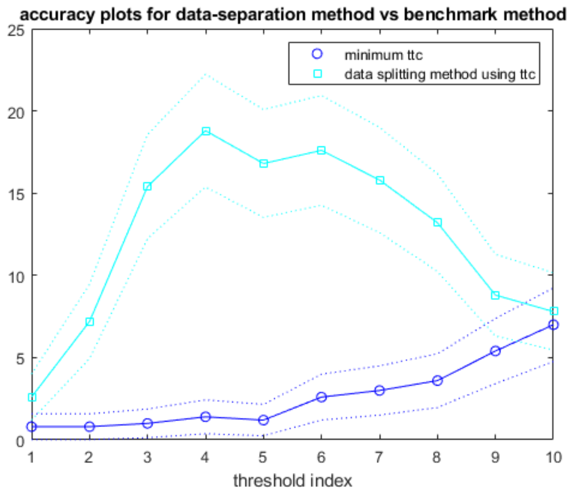

Figure 4.9: Plots shows accuracy plots for data-separation method vs minimum-TTC, where the minimum is taken over all frames of an encounter. Both methods use regular TTC to estimate P(C). Cut-off value is 50%.

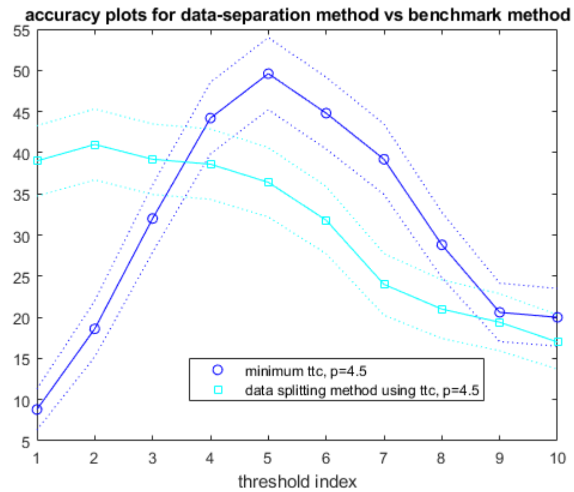

Figure 4.10: Plots shows accuracy plots for data-separation method vs minimum-TTC - both using inverse data transform - where the minimum is taken over all frames of an encounter. Both methods use regular TTC to estimate P(C). Cut-off value is 50%.

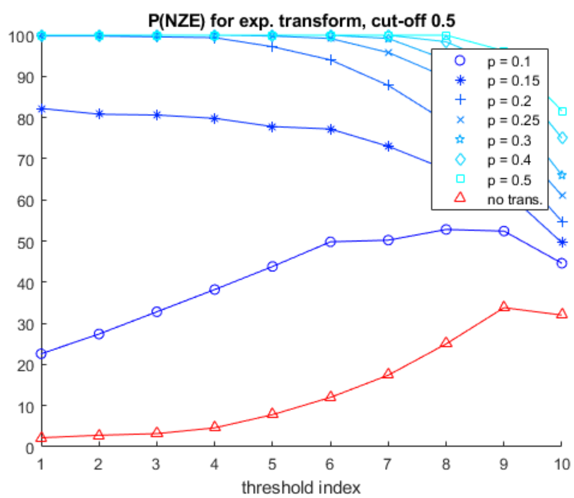

Figure 4.11: Plots shows estimated P(NZE) of estimators that use the exponential transform applied to stochastic TTC at FEA for various choices of power parameter.

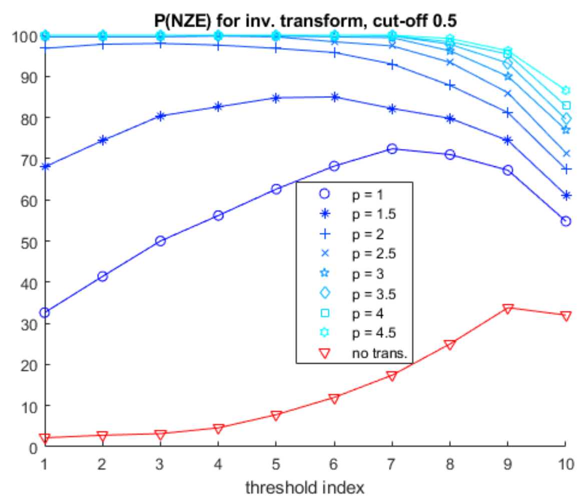

Figure 4.12: Plots shows estimated P(NZE) of estimators that use the inverse transform applied to stochastic TTC at FEA for various choices of power parameter.

Model vs empirical dist. function, stoch. TTC and min. distance

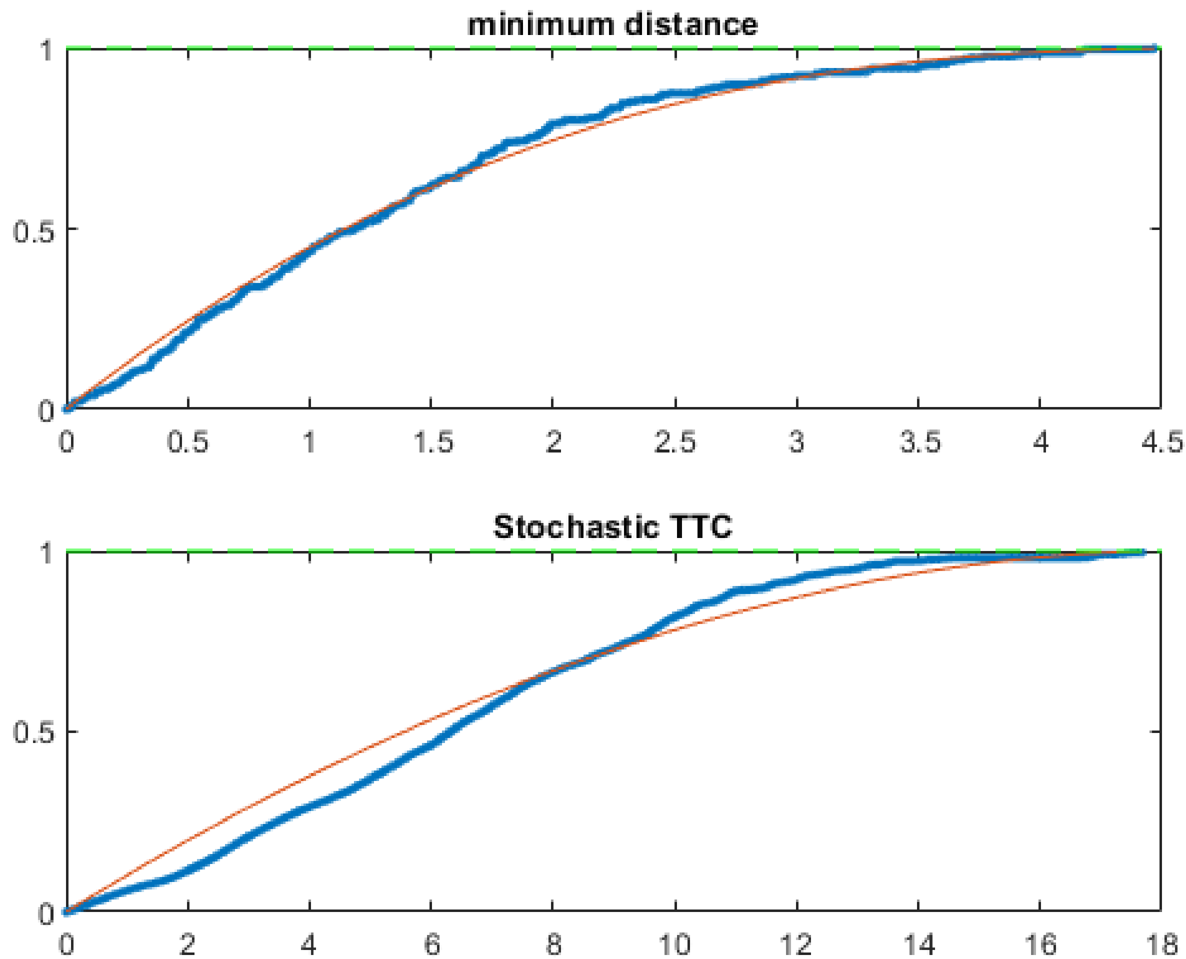

Figure 4.13: Example of model (red) and empirical (blue) distribution function for minimum distance and stochastic TTC. Threshold 3 is used in each case.

Example of diagnostic plots, with and without data transform

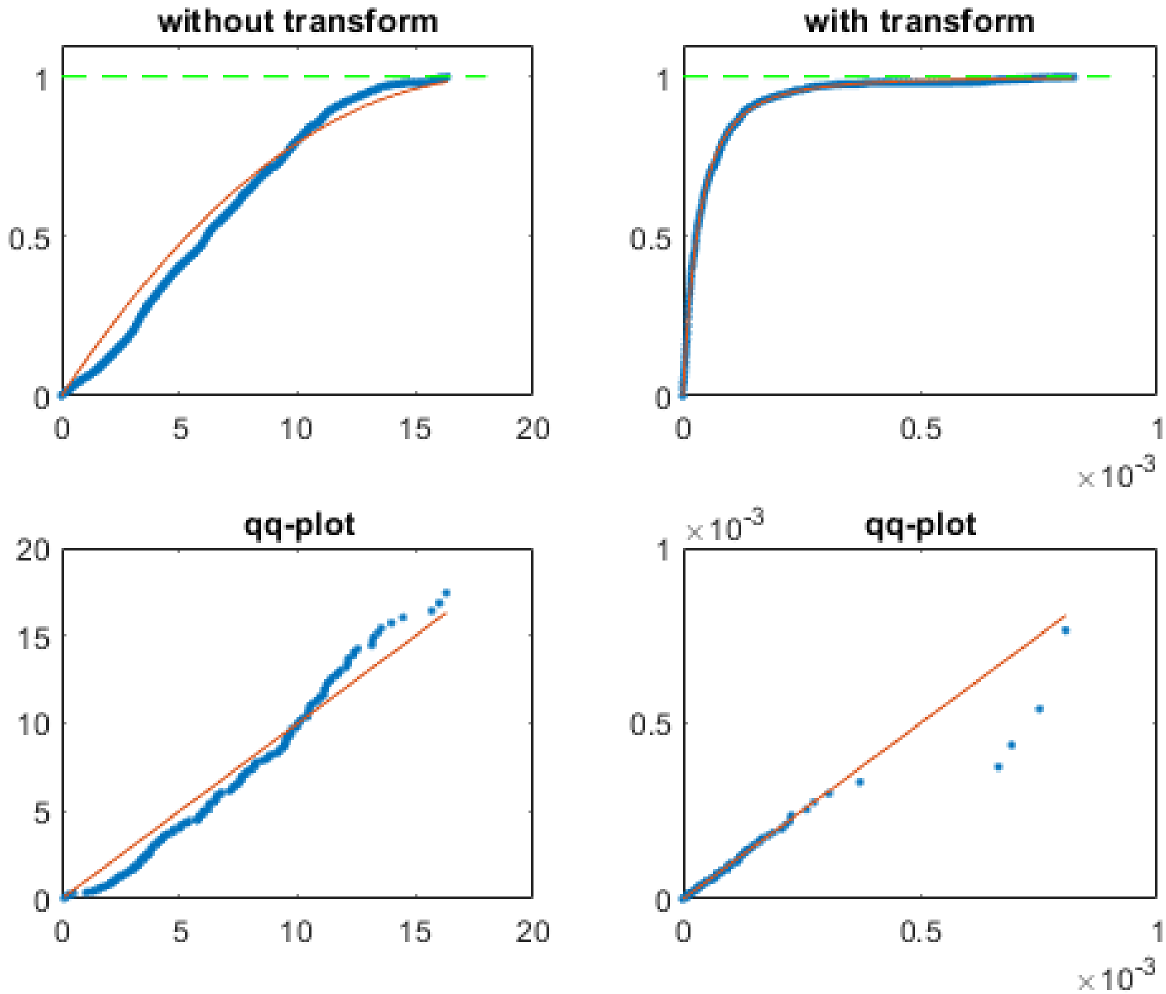

Figure 4.14: Example of model and empirical distribution function (top) and qq-plots (bottom) with and without transformation. Threshold 3 is used in each case, and the transformation used is the inverse transform with power parameter 3.5.

Example of diagnostic plots with different power parameter.

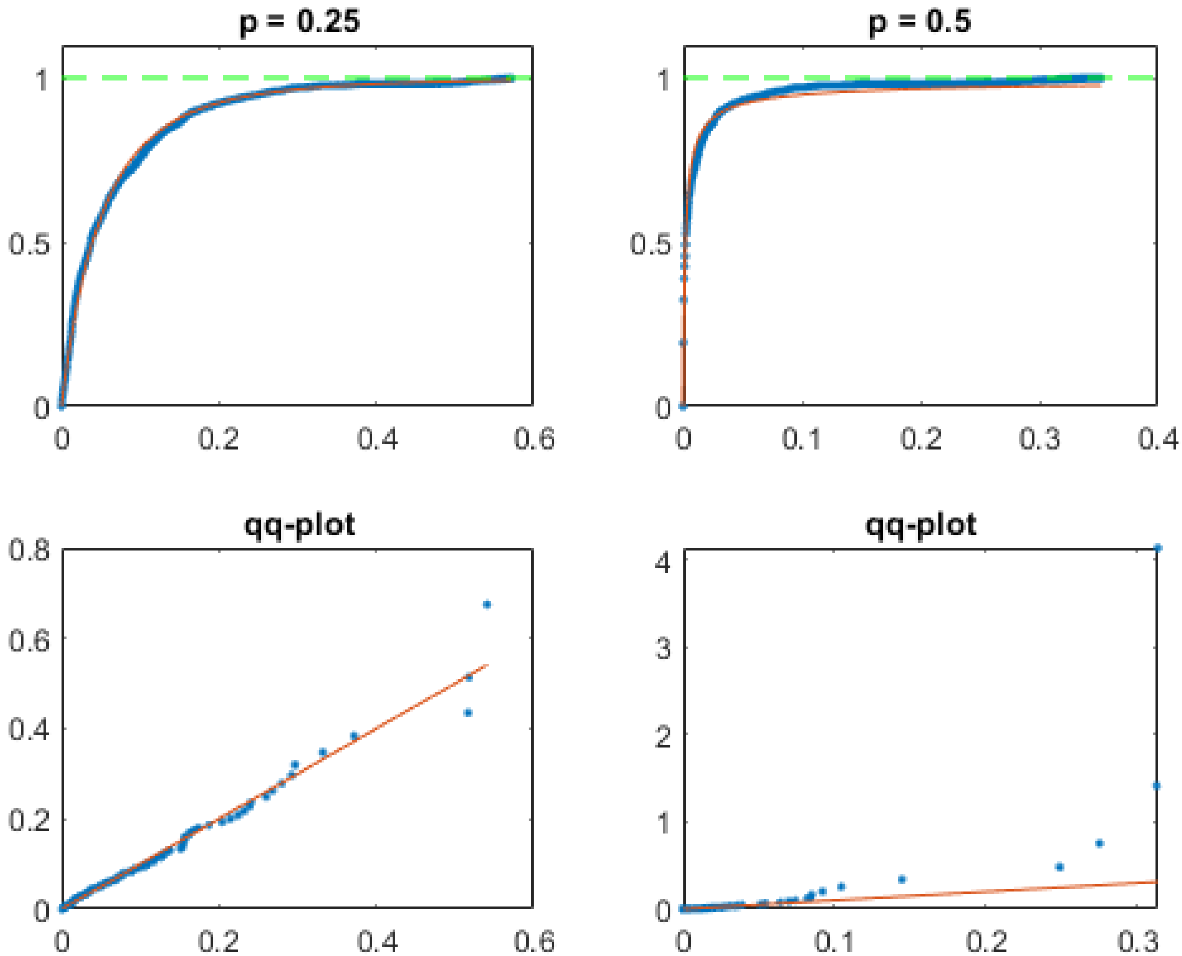

Figure 4.15: Example of model and empirical distribution function (top) and qq-plots (bottom) with different power parameters used. Threshold 3 is used in each case, and in both cases the exponential transform is used on stochastic TTC at FEA.

Example of diagnostic plots with different power parameter.

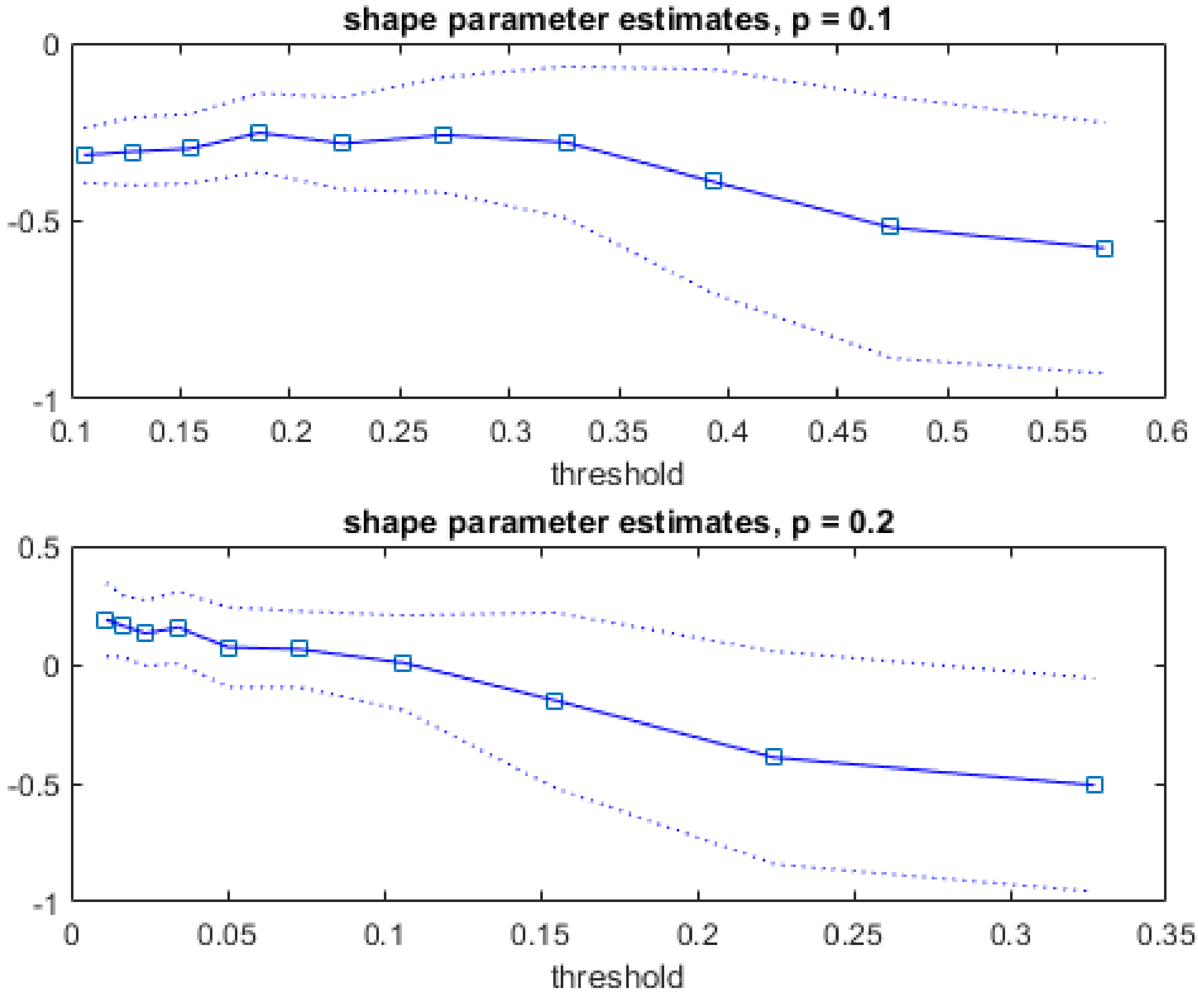

Figure 4.16: Example of shape parameter stability plot for power parameter 0.1 (top) and 0.2 (bottom). Threshold 3 is used in each case, and in both cases the exponential transform is used on stochastic TTC at FEA.

Example of diagnostic plots with different power parameter.

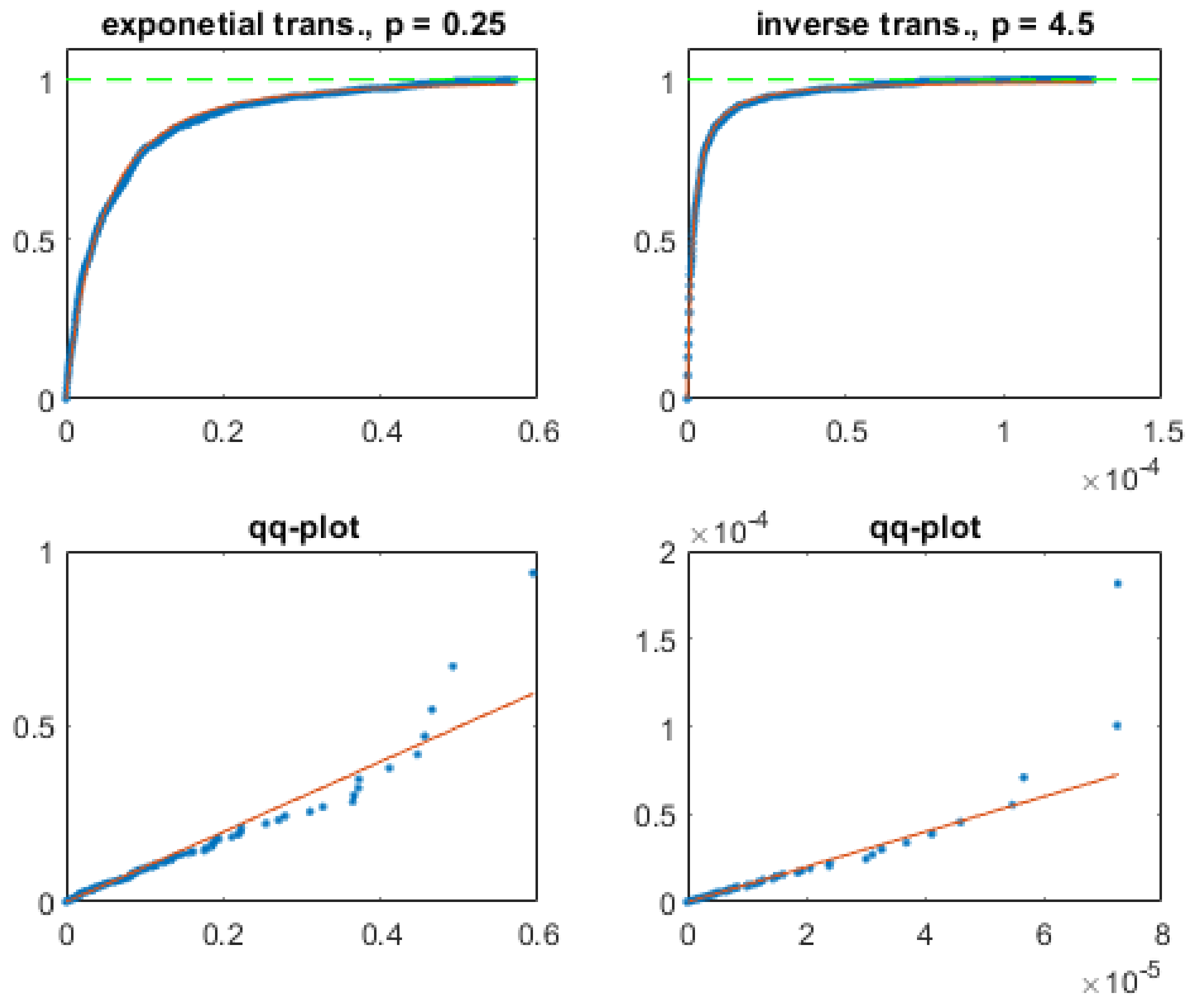

Figure 4.17: Example of model and empirical distribution function (top) and qq-plots (bottom) with different transformations used. In the left-side plots, inverse transform is used, and in the right-side plots the exponential transform is used. Threshold 3 is used in each case

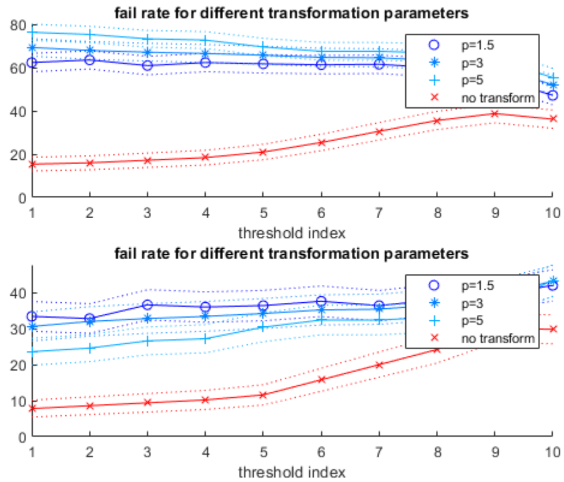

Figure 4.18: The plots shows the estimated probability of correctly (top) and incorrectly (bottom) predicting which simulated traffic location was safest when the prediction is based on estimated collision probability. The data used is minimum TTC of each encounter, with the inverse transform applied.

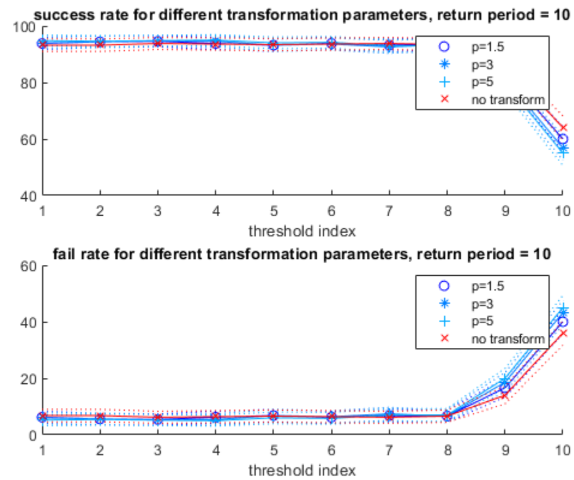

Figure 4.19: The plots shows the estimated probability of correctly (top) and incorrectly (bottom) predicting which simulated traffic location was safest when the prediction is based on return levels with return period 10. The data used is minimum TTC for each encounter.

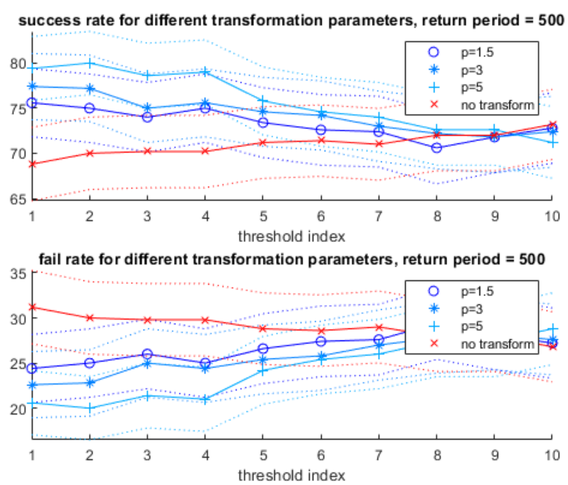

Figure 4.20: The plots shows the estimated probability of correctly (top) and incorrectly (bottom) predicting which simulated traffic location was safest when the prediction is based on return levels with return period 500. The data used is minimum TTC.

4.2 Discussion

4.2.1 Advantages of stochastic TTC

We have assumed the existence of an intended path, which we use to define collision course and TTC. However, this information is not necessarily known to us, and this makes the danger of the situation uncertain. Consider, for instance an encounter between two road users - denote them driver A and driver B - crossing paths in an intersection. Suppose we are concerned with the task of evaluating the level of danger at the moment when driver A or B first reacts to the presence of the other driver, after which that driver engages in corrective behaviour. By the assumptions outlined above, the degree to which this is a dangerous situation depends on how each road user **would** have driven had the other not been present. Consider for instance, if driver A habitually slows down significantly in the middle of the intersection regardless of the traffic situation, i.e. safe driving is driver A's default behaviour. then the situation might be quite safe, because even if driver A does not notice driver B, driver A will still slow down out of habit, thus allowing driver B more time to get out of the collision course. However, if driver A habitually maintains constant speed when driving through the intersection, then there is greater potential for danger. The consequence of driver A not noticing driver B is now more severe, as in this case there is less time for driver B to get out of the collision course, and the margin for error in the decision making is smaller.

The above example illustrates that there is inevitably going to be uncertainty in the degree to which a situation is dangerous, and this uncertainty should be accounted for by the severity measure. This is a strong argument in favour of using stochastic TTC over deterministic TTC. By using stochastic TTC we are incorporating naturalistic behaviour to compute the possible future motions of the drivers, and we are accounting for the frequency with each type of behaviour occurs.

The more realistic definition of collision course that is implicit when using stochastic TTC together with the above discussion on inherent uncertainty with regards to safety levels suggests that it is a more appropriate severity measure to use than deterministic TTC. It allows us to incorporate information about how road users are likely to behave in the traffic environment in question, and restricts attention to only the movements that are realistic. It is of little interest or relevance that two cars are defined to be on collision course if the collision course in question has zero probability of being realized. There are infinitely many ways that two cars could conceivably move in terms of what is physically possible, but only a small subset of all such trajectories are of interest when trying to determine the potential for collision in a given situation. For instance, we could have a situation where a car is in the process of turning left and crossing over the lane of an approaching vehicle, and is on collision course with the approaching car if defined in terms of current speed and velocity. But we know that the probability of this path being realized is for all intents and purposes equal to zero, even if there were no other road user to react to. For one thing, the driver already has the wheel rotated into a position that forces the car to continue turning. Furthermore, we know that the act of maneuvering a vehicle in traffic to some extent becomes automated and habitual over time, so that the driver would not have realized the constant-motion trajectory regardless of the presence of other vehicles. In other words, deterministic TTC is in this case based on a trajectory that is irrelevant.

4.2.2 Severity measure performance comparison

The comparison between the data-separation method and the benchmark method yielded mixed results. It is however ensuring to see that the method has performance comparable to the benchmark method. One possible explanation as to why minimum-TTC performed better in terms of peak performance after an optimal transform was used, is that minimum TTC always uses information from all 500 encounters, wheres the data separation method only uses data from encounters where interaction occurred, so that on average it effectively uses information from only ≈ 420 encounters.

The fact that minimum distance performed better at the lower thresholds can be better understood by looking at Figure 4.13, showing representative examples of plots comparing the empirical and model distribution function for minimum distance and stochastic TTC at the third threshold. These plots suggests that the modelling assumption - namely that the exceedences are GPD - is more accurate for minimum distance than for stochastic TTC at the lower thresholds, which is consistent with the observation that minimum distance performs better at these thresholds.

Stochastic TTC was nearly indistinguishable in performance from regular TTC. On one hand, this demonstrates that it works about as well as the benchmark comparison under ideal circumstances; that is, when we do not have to deal with the real life complications such as those associated with assigning a TTC-distribution to each moment of an encounter. On the other hand, an improvement - or at least a difference - in performance was expected, which raises the question of why this did not happen. The reason for this result seems to be that there was very little difference in TTC between the simulated trajectories at first moment of evasive action, so that the distributions were close to being point masses, i.e. they were highly concentrated around single points. Since assigning deterministic TTC is equivalent to assigning a point mass distribution at a particular value, the difference between the severity measures was small. Put differently, the behaviour of the vehicles was too homogeneous in order for stochastic TTC to make much difference. Also, the vehicles were not following curved trajectories, which is the situation were we expect stochastic TTC to significantly improve upon deterministic TTC.

Initially, it was somewhat surprising that minimum distance ended up performing better than any of the TTC measures when no transforms were used, given that TTC seems like a much better measure of the danger of a traffic situation. However, this reaction comes from the assumption that the key feature defining a good severity measure is how well it captures the danger of a situation. This might be a fairly reasonable statement when taking real world concerns into consideration, but within the setting of the simulation, the only purpose of the severity measure is to allow for extrapolation beyond the range of observed data. A key feature of a good severity measure from this point of view is that observations in the non-extreme range contains information about the behaviour of the process at the tail of the distribution, that is to say the frequency of rare events. Based on this criteria, it becomes less surprising that minimum distance performed well, since more close encounters - in the minimum distance sense - does imply more collisions within this fairly simple simulated environment. It is not clear that the same can be said about real life traffic however, as it is not obvious that there is such a strong connection between rate of spatially close encounters and rate of collisions; extreme spatial closeness during an encounter can happen for reasons that have nothing to do with the factors that causes accidents. It may have simply been the case that the road users were driving extremely slowly, in which case the spatial proximity may be deceptive, and the situation may still be well within the comfort zone of each road user.

Even if there was such a connection between rate of spatially close encounters and rate of accidents, There are still other drawbacks to consider that relates to the application of the severity measure. In traffic security, the primary interest is in events that involve potential for harm to road-users, and not necessarily collisions or close encounters in general. For instance, if an intersection has a high rate of low impact collisions resulting only in light vehicle damage due to low speeds and high traffic volume, it might still not be as urgent of a problem as an intersection where collisions happen less often, but are very serious when they do occur. If we were to use comparison of return levels - with minimum distance as the severity measure - as a method for inferring which out of the two traffic locations is more dangerous, we would likely conclude incorrectly that it was the former. In other words, it would falsely signal that "dangerous" encounters occur at a high rate. TTC does suffer from the same disadvantage to some extent, as it too can incorrectly label non-dangerous situations as dangerous(encounters with low speed, but close proximity), but it is arguably less likely to miss-label actually dangerous situations as non dangerous, and as such still seems like a more reasonable severity measure to use. In any case, this drawback should be eliminated by using stochastic TTC implemented using empirically determined trajectories.

4.2.3 Effect of data transforms

An important effect of increasing the power parameter is to increase the expected value of the estimators. The results of the simulation suggests that estimators that perform well tend to be overestimators. In addition to increased performance in terms of increased accuracy, we might prefer estimators that over rather than under estimate in any case since the consequence of under estimation of danger-levels is arguably worse than the consequence of overestimation when considering the application.

Figure 4.14 shows a representative example of diagnostic plots for the third threshold when using stochastic TTC, with and without use of an inverse transform respectively. These figures indicate that the modelling assumption - i.e. that the exceedences are GPD - is more accurate at this threshold after applying the transform. This is consistent with the result that the inverse transform significantly improved performance, in particular at the lower thresholds. Similar results were observed in the case of the exponential transform. Note that good fits being obtained at lower threshold effectively allows us to use more of the data in the estimates, potentially reducing variance.

The fact that the transformations had more of an effect on performance for the TTC measures than for minimum distance is possibly due to the fact that, in the case of minimum distance, the fit is quite good already at lower thresholds without a transformation, see Figure 4.13, and so there is perhaps less to gain by applying a transformation in order to improve the model fit.

In Figure 4.11 and 4.6 we see a much higher NZE rate and expected value after applying the transformation, and the hit-rate increases from around 30% to 100% at the lower thresholds when changing the power parameter from 0.1 to 0.25. Since $P(\text{Accurate estimate}) = P(\text{Accurate estimate}|\text{NZE})P(\text{NZE})$, this will result in an increase in accuracy if $P_{0.1}(\text{Accurate estimate}|\text{NZE}) \approx P_{0.25}(\text{Accurate estimate}|\text{NZE})$. So, one way to explain the performance increase is to note that the large increase in $P(\text{NZE})$ compensates for the small decrease in $P(\text{Accurate estimate}|\text{NZE})$ that results from the increased bias.

The above discussion suggests decomposing the effect of increasing the power parameter into a negative and a positive effect on performance. On one hand, increasing the power parameter seems to increase $P(\text{NZE})$. On the other hand, $P(\text{Accurate estimate}|\text{NZE})$ will eventually start decreasing due to increase in bias and variance. A way to think of this then, is as a trade-off between these two effects, with the optimal parameter choice (for a fixed threshold) being reached once the performance gain produced by the increase in P(NZE) can no longer compensate for the performance loss resulting from the increase in bias and variance. Figures 4.6 and 4.11 strongly suggest that there is such an optimal point, since the bias grows more rapidly with the power parameter as it gets larger, whereas the $P(\text{NZE})$ on the other hand quickly reaches strongly diminishing returns, especially for $0.1 < p < 0.2$.

In the case of the exponential transform applied to the TTC data, we see that there is very poor performance at the low thresholds when using a high power parameter. This poor performance can be predicted by the diagnostic plots, as is illustrated by Figure 4.15 where we see an example of diagnostic plots of p=0.25 and p = 0.5 respectively, using threshold 3. The strong bias we see in Figure 4.6 at threshold 2 is indicated by the poor fits observed in both of these plots. In particular, we can predict that p=0.5 is likely to result in overestimation as the model distribution is significantly below the empirical distribution towards the right end tail of the distribution.

Smaller differences in performance is harder to predict and explain. In Figure 4.16 there is arguably indication of slightly more bias for p=0.2 at the lower thresholds. In hindsight, we know that it is worthwhile to accept this slight increase in bias in return for the large increase in P(NZE). Without this knowledge however, it is hard to make such a decision. It is not clear how to reliably infer from the visual indication of bias in the diagnostic plots the actual bias of the estimator itself, or which amount of bias results in the optimal trade off. A suggestion for how to approach the problem is to plot the probability estimates over a range of thresholds for different values of the power parameter, and then select - if possible - the power parameter that produces both non-zero estimates as well as reaching reasonable stability in the diagnostic plots, reflecting our goal to strike a good balance between bias and higher NZE rate. To simultaneously satisfy the condition of stability as well as the condition of getting non-zero estimates is not guaranteed however, as we might sometimes have to tune the power parameter up a lot before non-zero estimates are obtained, resulting in large bias, and stability plots from which it is hard to select an non-arbitrary threshold. More research is needed to make more definitive statements about how such choices should be made, and under what circumstances this type of procedure could result in performance improvements.

4.2.4 Performance difference between transforms

A more difficult result to explain is why the inverse transform performed better than the exponential transform, as there was not an obvious visual difference in goodness of fit when considering the diagnostic plots of the transformation parameters that performed best. In Figure 4.17 we see an example comparison of diagnostic plots for the exponential transform with parameter 0.25 and the inverse transform with parameter 4.5, both applied to stochastic TTC at first evasive action. Both fits looks quite good, and it is not obvious from these plots that the inverse 4.5 performs best. Furthermore, figures 4.6 and 4.7 show that the exp 0.25 transform has a slightly lower bias than the inv 4.5 transform, and the P(NZE) in both cases are close to 100% percent over the first few thresholds, so these factors do not seem like they can explain the performance difference. It appears that it is the much lower standard deviation of the inverse transform - about 5 times smaller than that of the exponential transform at threshold 3 - that is the main reason for the performance difference, see Figure 4.8. If this difference in variance can be reliably inferred from a single sample of observations, then it can be used as a basis for deciding which transform to use in the cases when such a choice it not made obvious from the diagnostic plots. By using standard bootstrapping for estimating the standard deviation, with 200 bootstrap samples for each estimate, it turns out that this estimator will correctly indicate which estimator has the lower variance about 95.5, 97.5, 94.0, 89.5 and 88.5 percent of the time for the 5 lowest thresholds. For higher thresholds, the variance estimate becomes less reliable. Perhaps then the estimated variance could act as a tie breaker when there are otherwise no clear reasons to prefer one transform over the other.

4.2.5 Suggestions for choice of transformation type and parameter

The discussion so far suggests - at least in this particular simulated setting - selecting transform type and parameter according to the following tentative rules of thumb. Pick a transform if it improves the fits according to the diagnostic plots, in particular when it improves the fits at lower thresholds so that more data can be used efficiently. Also, consider the effect on the variance of the estimand, keeping in mind that the variance estimates at the lower thresholds are more reliable than those at the higher thresholds. In other words, try to find a combination of transformation type and parameter that offers improved goodness of fit and low variance. Also, as previously mentioned, increase the power parameter until both estimation stability and non-zero estimates are obtained, if possible. Note that these are suggestions that are made in hindsight after seeing which methods performed well, and as such they should be viewed as speculations that need to be tested further to see if they generalize, and if so - to which situations.

4.2.6 Viability of data-transformation when applied to traffic data

There are a couple of points that needs to be made with regards to the above suggestions. First of all, it is likely that these rules fail to offer benefits under circumstances when the data is not as "smooth" and suitable for extrapolation as the data dealt with in this simulation experiment. As mentioned, the results overall seems to indicate that good performance was associated with estimators for which there were good fits achieved already at low thresholds, allowing for more of the data to be used. However, using lower thresholds essentially means we are extrapolating from data that is less extreme, and in doing so, we are assuming that this data contains information about the extreme part of the distribution. Otherwise, it is hard to see why using this data for extrapolation should reliably improve results. To translate this into terms of a practical situation, it would be similar to assuming that fairly regular traffic, i.e. what we would consider as non dangerous situations, contains information about the rate at which collisions occur. Although there is research establishing a relationship between the rate of traffic conflicts with the rate of collisions [ref], it is not clear at what point along the traffic severity spectrum that the encounters start providing information about the rate of collisions. Therefore, it is not obvious that the suggested methods of this thesis would offer benefits to traffic data, or to what extent one should attempt to include more "regular" traffic data in the estimation of safety levels.

4.2.7 Performance in ability to infer relative safety levels

The fact that the data transforms improve the inference success rate can be largely attributed to the fact that they increase P(NZE), as this reduces the risk of getting an inconclusive outcome. Also, as mentioned above, some of the transforms allows us to use more of the data in the inference, thus reducing the variance of the estimates. We saw that this may come at the cost of increased bias, but this need not be a problem as long as the bias is the same for each traffic location. To illustrate this point, if we were to instead use mean value of minimum TTC, $\bar{X}$, as an indicator to base our inference on, we could just as well use $\bar{X} + 1$, and it would make no difference to the inference, despite the fact that we would be using a biased statistic. This reasoning is also likely to explain why performance was highest at the lowest threshold, as overestimation does not hurt performance in this case.

Based on how reliably the return levels correctly indicated which traffic location was safer, especially relative to the estimator of the collision probability, it is tempting to conclude that one should always use return levels as a basis for such inference. It may be the case that they would be better to use in some cases due to the higher robustness, but it is important to keep in mind the conditions required for them to work. We are relying on the assumption that higher return levels imply higher (assuming inverted data) probability of collision. This need not be true however. For instance, consider two GPD distributions with scale and shape parameters $(1.8, 0.08)$ and $(1.5, 0.15)$ respectively. the 100th return levels are 10.0224 and 9.9526 respectively. But if the threshold corresponding to collision is say 30, then the collision probabilities would be 2.5136e-05 and 9.6887e-05 respectively. In this scenario, the 100th return levels would then on average indicate (assuming unbiased estimator) the wrong conclusion with regard to collision frequency. To use return levels for inference then, ideally requires investigation into whether or not it is true that a higher return level implies higher collision probability. So, to some extent, the increase in performance w.r.t. this kind of inference is due to the fact that we got lucky, since a precondition for the method to work happened to hold. It may also be the case that for instance the mean value of minimum TTC could indicate with high reliability the safest traffic location, but the same conclusions would hold for such a statistic.

The fact that performance increased as we used return levels with lower return periods can be explained by the fact that we are essentially basing inference on a more robust statistic. The return levels associated with low return periods provide us with information about the less extreme behaviour of the process in study, and so it is not too surprising that they might result in more reliable predictions, as they will naturally have lower variance. However, using lower return levels for this inference involves problems similar to those discussed above, as it means assuming that the frequency of non-extreme events is predictive of the frequency of extreme events, and this assumption may of course not hold, especially not in a system as complex as road traffic. The topic of which statistics to use for inferring relative safety levels is a topic that warrants further research.

4.2.8 Suggestions for future work

In order to make conclusions that can be confidently generalized, more research is needed. A first step could be to use the discussed methods on data generated by different distributions, such as the gamma, beta, chi square and normal distributions, under blind conditions where the true value of the "collision probability" is unknown. This was done to some extent on the aforementioned distributions with what seemed to be similar and promising results, but it would need to be tested more systematically to make any conclusions. In addition, it is desirable to study these questions in a more mathematically rigorous manner, beyond simulation testing. For instance, the results seems to strongly indicate that the effect of increasing the power parameter for either transform has the effect of increasing the expected value of the estimator. It would be interesting to see under what circumstances this happens. That is, which properties of the transform and data being transformed are essential to get such an effect. Perhaps this occurs when using any monotonically decreasing and strictly convex transformation on a random variable that has a smooth and convex left tail, such as the above mentioned distributions.

A suggestion for a follow up to this simulation experiment - in order to hopefully make better use of the idea of stochastic TTC - is to introduce the concept of driver types. At the beginning of each encounter, randomly draw the type of each driver, where the type determines the rules dictating the statistical behaviour of the road user. Then at the moment of first evasive action, simulate trajectories as usual in order to estimate the TTC distribution, with the exception that another layer of randomization is added by drawing the type of each road user before each walk is initiated. This is similar to asking the question "How dangerous would this situation have been if road user A and B was of type X and Y?" and including the answer in the estimate of the danger of the situation. As a very simple example, we could have two types of road users, one which has a lot of angular momentum and little variability in terms of the step-angle, and one which has much lower angular momentum and can change direction more quickly.

4.3 Conclusions

Stochastic TTC performed about as well as regular TTC in the simulation experiment. More features need to be added to the simulation to make the vehicle behavior more variable and realistic before the two methods can be expected to deliver significantly different results. Despite the results not being significantly different from the main benchmark method, the simulation offers a first proof of concept, and overall encourages continued research.

Similarly, the data-separation method performed comparably to the benchmark method. This encourages application to actual traffic data to see how well it performs under such circumstances. It offers a way to use stochastic TTC in practice, but it remains to investigate how to most optimally measure the severity levels during the attempt to resolve the conflict. Even if we can measure P(C,NEA) (the probability of collision where no attempt at evasive action was made) well by virtue of using stochastic TTC, this gain might be negligible if we still have large error in the estimate P(C,EA) (probability of collision where evasive action was attempted).

Data transformations proved to be highly useful in improving performance results, especially when applied to the TTC data. A common theme that emerged from the results was that the estimators that performed well were the estimators for which there seems to be a good GPD fit achieved already at lower thresholds, allowing for more efficient use of the data and a reduction of variance. The improved fit appears to be one of the main reason as to why the transformation improved performance. Another reason for improved performance seems to be the increase in expected value, resulting in an increased probability for the tail of the fitted distribution to cover the collision-threshold, at the expense of increased bias.

Although suggestions were made for how performance improvement might be predicted from the diagnostic plots, it still remains to be properly demonstrated that suitable choice of transformation type and parameter can be reliably predicted without knowledge of the true collision probability. Also, it remains to investigate under what conditions the suggested methods can offer improvements, and in particular if the benefits extends to traffic data.

Bibliography

[1] Paul St-Aubin, Nicolas Saunier, and Luis F. Miranda-Moreno. "Road User Collision Prediction Using Motion Patterns Applied to Surrogate Safety Analysis." In: *Transportation Research Record: Journal of the Transportation Research Board* (2013).

[2] Stuart Coles. *An Introduction to Statistical Modeling of Extreme Values*. London, UK: Springer-Verlag, 2001.

[3] Chr. Hyden. *A traffic-conflicts technique for examining for examining urban intersection problem.* Lund Institute of Technology.

[4] Nicolas Saunier and Tarek Sayed. "Probabilistic Framework for Automated Analysis of Exposure to Road Collisions." In: *Transportation Research Record: Journal of the Transportation Research Board.* (2008).

[5] Andrew P. Tarko. "Estimating the expected number of crashes with traffic conflicts and the Lomax Distribution – A theoretical and numerical exploration." In: *Accident Analysis Prevention* 113 (2018), pp. 63–73.

[6] Andrew P. Tarko. "Use of crash surrogates and exceedance statistics to estimate road safety." In: *Accident Analysis Prevention* 45 (2012), pp. 230–240.

[7] Lai Zheng, Karim Ismail, and Xianghai Meng. "Freeway safety estimation using extreme value theory approaches: A comparative study." In: *Accident Analysis Prevention* 62 (2014), pp. 32–41.

[8] Lai Zheng, Karim Ismail, and Xianghai Meng. "Shifted Gamma-Generalized Pareto Distribution model to map the safety continuum and estimate crashes." In: *Safety Science* 64 (2014), pp. 155–162.