Multidimensional risk assessment for vehicle trajectories by using copulas
Abdourahmane Koita, Dimitri Daucher, Michel Fogli

To cite this version:

HAL Id: hal-00865839
https://hal.archives-ouvertes.fr/hal-00865839
Submitted on 25 Sep 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Multidimensional risk assessment for vehicle trajectories by using copulas

A. Koita & D. Daucher  
IFSTTAR, IM, LEPSIS  
University of Paris-Est, F-75732, Paris, France  

M. Fogli  
Department of LaMI  
University of Blaise Pascal, 63175, Clermont Ferrand.

ABSTRACT: Fight against road unsafety is a French government priority. The policy conducted since 2002 allowed to obtain undeniable success. Despite this improvement, road accidents have very serious consequences on human level for road users. The proposed methods to reduce these accidents involve independence between criteria of accident risk. The objective of this study is to estimate the multidimensional risk of failure trajectory. It consists to investigate a new method of risk assessment in order to better characterize the dependence structure between the vehicle criteria for safety acceptance. This requires the use of simulation techniques such as copulas methods. This function connects the joint probability distribution to the marginal distribution. Thus, it contains all information on the dependence structure of models. However, the difficulty of multidimensional risk is to choose the copula which capture the better dependence between criteria. To select an adequate copula must be based on a statistical test. The $\chi^2$ test used in the framework of the adjustment of a parametric distribution to an empirical distribution is in this regard an interesting tool for the choice of copula. The experiments have shown the relevance and effectiveness of this method. The results will help to better assess the risk of failure trajectory for vehicles.

1 INTRODUCTION

Despite the improvement of road safety, vehicle accidents have very serious consequences on human level for road users, (ONISR 2011). We note that road accidents depend on several factors, including driver and vehicle environment, acting on trajectory (e.g. effects of random external actions like wind, or unexpected reactions of drivers), (IRTAD 2010). Therefore, taking into account random phenomena in the civil engineering problems has become increasingly common and necessary in order to ensure the safety of structures and systems. It must be taken into account in modeling of systems as well as in the formulation of reliability problems.

In the past, many strategies have been proposed to investigate this complex problem. We remember that our laboratory (IFSTTAR) worked in the road safety problems since several years. In this context, a reliability analysis has been applied on experimental site (Nantes/France). It consists to predict failure trajectories, starting from a given bend configuration and representative observations of real trajectories. One calls failure trajectory, any trajectory not respecting the safety conditions defined by the legislator. The risk levels of failure trajectory were evaluated according to several criteria. This methodology has been a contribution to the development of warning procedures designed to significantly reduce the number of accidents in bend, (Koita 2011).

However, the failure criteria of vehicle trajectory were considered independently in risk assessment. The reproach to this result is to not include the dependence structure between criteria. One of the most important questions in accident analysis is what role correlation between failure criteria plays in risk assessment. Specifically does correlated or dependent activity convey more information than criteria used independently. Then, how are one-dimensional correlated risks? Note that the correlations between variables of risk are often difficult or impossible to quantify. In particular, in the case of two variables will require $n^2$ measures where $2n$ are sufficient to estimate two marginal densities.
In a previous study (Koita, Daucher, & Fogli 2012a), we have measured the impact of the variation of criteria as function of the failure probability \(P_f\). It represents the probability to exceed a safety threshold. To deepen up this approach, we propose in this study to investigate a new method of risk assessment in order to better characterize the dependence structure between the relevant failure criteria. This requires the use of simulation techniques such as copulas methods. This function connects the joint probability distribution to the marginal distribution. Thus, it contains all information on the dependence structure of model. In practice, at the time of the copula model implementation, it is advisable to be able to estimate the dependence by using available data. The failure criteria will be assimilated to components of a \(\mathbb{R}^n\)-valued stochastic vectorial process defined on \((\Omega, \mathcal{F}, \mathbb{P})\). Copula involving modelling two or more random variables, (Frees & Valdez 1998). This technique will allow to estimate the multidimensional risk of failure trajectory.

2 METHODOLOGY

After this introduction, we first provide some definitions about relevant criteria of failure trajectory. The vehicle’s trajectory is defined as realizations of a \(\mathbb{R}^6\)-valued stochastic process \(U = (U(t), t \in \mathbb{R})\) defined on \((\Omega, \mathcal{F}, \mathbb{P})\). We obtain a trajectory \(u(t) = U(t, \omega), \forall t \in T\) where \(\omega \in \Omega\).

Then, we introduce copulas method in order to estimate multidimensional risk of failure trajectory. That will be the key focus in this study. The difficulty of multidimensional aspects of risk is to choose the copula which capture the better dependence structures between failure criteria. To select an adequate copula must be based on a statistical test. The \(\chi^2\) test used in the framework of the adjustment of a parametric distribution to an empirical distribution is in this regard an interesting tool for the choice of copula, (Fermanian & Scaillet 2003).

Finally, we propose to estimate the risk level of failure trajectory by taking into account of the dependence between criteria. We will also show the effect of the dependence structure in the vehicle risk assessment.

3 CHOICE OF FAILURE CRITERIA

In this section, we describe criteria used to estimate the risk of failure trajectory. In research work(Koita, Daucher, & Fogli 2012b), we considered three failure criteria:

- The relative distance between the center of traffic lane and the vehicle’s trajectory \(u\). This geometrical distance must be limited because one cannot deviate indefinitely from the center of traffic lane while remaining on road space rationing. This criterion noted \(K_1\) is defined by:

\[
\sup_{t \in T} |D(t)| > \delta^s
\]

where \(\delta^s\) is a safety threshold value. This criterion relates on the \(\mathbb{R}, \mathbb{P}\)-valued process of control \(D = (D(t), t \in \mathbb{R})\) defined on \((\Omega, \mathcal{F}, \mathbb{P})\).

- Lateral acceleration and its variation (jerk) are criteria often used in the literature to determine the \textbf{vehicle trajectory dangerosity}. It is used as failure criterion because the going beyond a threshold \(\delta^s\) can generate loss of control, see (Revue 2003). This criterion noted \(K_2\) is defined:

\[
\sup_{t \in T} |\Gamma_N(t)| > \delta^s
\]

It relates on the \(\mathbb{R}\)-valued process of control \(\Gamma_N = (\Gamma_N(t), t \in \mathbb{R})\) defined on \((\Omega, \mathcal{F}, \mathbb{P})\). This process is easy to determine because it is a coordinate of trajectory \(u\) in the Galilean coordinate system \(\mathcal{R}^4_0\).

- The vehicle’s orientation compared to the axis of traffic lane plays a significant role in the search for strategy to control vehicle in corner driving. This criterion noted \(K_3\) is defined by:

\[
\sup_{t \in T} |\psi(t)| > \bar{\psi}
\]

where \(\bar{\psi} = \delta^s\) is a safety threshold of vehicle orientation to be chosen in an interval \([\psi_{\min}, \psi_{\max}]\). This criterion relates on the \(\mathbb{R}\)-valued process of control \(\psi = (\psi(t), t \in \mathbb{R})\) defined on \((\Omega, \mathcal{F}, \mathbb{P})\).

Note that, these 3 criteria of failure seem to be relevant. They relate on the processes of control: \(D(t), \Gamma_N(t)\) and \(\psi(t)\). The max of these processes will be limited by thresholds \(\delta^s\) to guarantee the vehicle stability on road. The processes \(D(t), \Gamma_N(t)\) and \(\psi(t)\) are obtained by transformation of the process \(U\) via a functional \(F\). Where \(F\) is a functional of \(\mathbb{R}^6\) with value in \(\mathbb{R}\), relative to criteria chosen to define failure trajectory, (Koita 2011).

Starting from these criteria, we calculated the failure probability \(P_f\) by using reliability of components in series (integration of Sup criterion). That means as soon as criterion is not respected the system becomes failing. For that, the probability of a union noted \(P_f-\text{sys}\), has been calculated:
\[ P_{f_{sys}} = \text{Prob}(\cup E_i), i = 1, 2, 3 \] (4)

The reproach made with this result is to not calculate probability of interaction between failure criteria. The object of this study is to estimate multidimensional risk of failure trajectory. It consists to calculate the failure probability \( P_f \) by taking into account of the risk related to the structure dependence between criteria.

For that, we propose to choose the relevant criteria of failure. In research work (Koita, Daucher, & Fogli 2012a), a sensitivity analysis was carried out to show the effect of criteria on the failure probability estimation. Let us recall that a classical analysis of sensitivity seeks to calculate the statistical parameters of output variables according to the statistical parameters of input data. For this study, we sought to characterize the sensitivity \( \frac{\partial P_f}{\partial K_i} \) of the probability \( P_f \) by report to failure criteria \( K_i \). To compare criteria on the same scale, the min-max normalization is used. The standardized safety threshold \( \delta^* \) for each criterion varies in \([0; 1] \). The results showed that the functions \( \pi(\delta^*) \) are different according to criteria \((K_1, K_2 \text{ and } K_3)\). These functions represent the variation of \( P_f \) as function of \( \delta^* \).

The criteria do not have same variability. The most important variation is between \( K_2 \) and \( K_3 \). This result corresponds to a driving situation where drivers generate a strong lateral acceleration \( \Gamma_N \) by minimizing the vehicle’s orientation \( \psi \) compared to the axis of traffic lane. It is also noted that failure criteria are treated on a hierarchical basis with respect to \( P_f \). From this result, we obtain the following order relation:

\[ K_3 \prec K_1 \prec K_2 \] (5)

We conclude that the criterion of orientation \( K_3 \) has less effect on the failure probability \( P_f \) than two other criteria. What implies the relevance of \( K_1 \) and \( K_2 \). Also, let us note that lateral acceleration criterion \( K_2 \) is most relevant to evaluate failure trajectory. Thus, we have just carried out a study of sensibility of \( P_f \) compared to the effect of criteria.

To estimate the multidimensional risk, we propose to simulate the processes of control \( D \) and \( \Gamma_N \) because being most relevant to characterize the failure trajectory. That requires the simultaneous simulation of \( \mathbb{R}^2 \)-valued vectorial stochastic process \((D, \Gamma_N)\) defined on \((\Omega, \mathcal{F}, \mathbb{P})\). For this purpose, the copula model is proposed in order to include the dependence structure between criteria.

4 DEPENDENCE MODELLING VIA COPULA

Copula is a multivariate distribution function \( C \) defined on the unit cube \([0; 1]^n\), with uniformly distributed marginals \([0; 1]\). Let \( F \) be an \( n \)-dimensional cumulative distribution function with marginals \( F_1, \ldots, F_n \), there exists \( n \)-copula \( C \) such as for each \( x \) of \( \mathbb{R}^n \), \( F(x_1, \ldots, x_n) = C(F_1(x_1), \ldots, F_n(x_n)) \). The density \( c \) of a copula \( C \), if it exists, is defined as follows:

\[ c(u_1, \ldots, u_n) = \frac{\partial^n C}{\partial u_1 \cdots \partial u_n}(u_1, \ldots, u_n) \] (6)

The copulas constitute a statistical tool making it possible to model the dependence between random variables.

4.1 Dependence structure

The research of dependence relates at the same time to the various strategies. There are many ways of measuring the dependence (concordance) between random variables. In statistic, the Kendall’s Tau \( \tau \) and the Spearman’s \( \rho_S \) are two classical measurements of correlation. They provide the perhaps best alternatives to the linear correlation coefficient as a measure of dependence for nonelliptical distributions, for which the linear correlation coefficient is inappropriate and often misleading. In addition, they offer the advantage of being expressed simply according to the copula associated with the couple of random variables. For more details about Kendall’s tau and Spearman’s rho and their estimators, we refer to (Kendall & Stuart. 1979), (Kruskal 1958), (Lehmann 1975) and (Caperaa & Genest. 1993). For other interesting scalar measures of dependence see Schweizer and Wolff (1981).

4.1.1 Kendall’s rank coefficient of correlation

Kendall’s tau \( \tau \) is a measure of rank dependence commonly computed nonparametrically, but does have an analytical relationship with respect to most parametric copulas. Let \((X, Y)^T\) be a vector of continuous random variables with copula \( C \). Then Kendall’s tau for \((X, Y)^T\) is given by:

\[ \tau(X, Y) \triangleq 4 \int \int_{[0,1]^2} C(u, v)dC(u, v) - 1 \] (7)

For instance, Kendall’s tau equals the difference between the probabilities of concordance and discordance, for a randomly selected pair of observations.

Written in another form by:

\[ \tau(X, Y) = \begin{cases} \Pr \{ (X - X')(Y - Y') \geq 0 \} & \\ -\Pr \{ (X - X')(Y - Y') < 0 \} & \end{cases} \] (8)
Where \((X', Y')\) a vector of same law as \((X, Y)\) and independent.

4.1.2 Spearman coefficient of correlation

The Spearman’s \(\rho\) is defined by:

\[
\rho_S(X, Y) = 12 \int \int_{[0,1]^2} C(u, v) \, du \, dv - 3 \quad (9)
\]

Different copula functions exhibit different dependence patterns. Therefore, if a researcher wants to explore the structure of dependence, he may estimate several copulas and choose one on the basis of best fit to the data. The choice of copula represents the first difficulty in the practical application of the dependence modeling.

4.2 Copula choice

It is necessary to find the copula family that describes best the correlation between variables. Once a researcher has specified the marginal distributions, an appropriate copula is selected. Because copulas separate marginal distributions from dependence structures, the appropriate copula for a particular application is the one which best captures dependence features of data. A large number of copulas have been proposed in the literature, and each of these imposes a different dependence structure on the data.

Regarding classes of copulas, the AMH, Clayton and Frank copulas are Archimedean, however the Gaussian is not. The Gaussian copula is a multivariate density that captures the elliptical dependence shape, but as is the case with all copulas, is capable of binding arbitrary marginal distributions. It has parameters that would need to be estimated in the same fashion as any parametric copula, typically through maximum likelihood estimation. The standard multivariate Gaussian density of course assumes that the marginals are Gaussian. The Gaussian copula is mathematically less tractable than the Archimedean family of copulas. Although the AMH copula is somewhat more tractable to work with analytically, it is not as flexible as the Clayton copula in terms of the possible range of dependence. Both of these copulas are asymmetric, with the strongest dependence in the lower tails. The Frank copula (Genest 1987), also an Archimedean copula, is radially symmetric, a characteristic it shares with the Gaussian copula.

We developed procedures to simulate each copula and estimate the parameter underlying each copula (dependence structure) for a given data set using an approach similar to that of (Genest, Ghoudi, & Rivest 1995) and Dias et.al (2004), and apply several goodness-of-fit tests to test the appropriateness of these copulas, see figure 1.

4.2.1 Fit test description

The choice of the best copula making it possible to model the dependence between random variables is of major importance. However, the tests of adequacy for the copulas are relatively recent. It is to be announced that one finds few articles on the subject, but the research field is under development constant. To our knowledge, this is the first multivariate goodness-of-fit test for Archimedean copulas that is actually implementable and, indeed, implemented.

Even though the test statistic is based on the classical \(\chi^2\)-statistic, its asymptotic distribution is non-standard. The null hypothesis \(H_0\) states that the dependence structure of data at hand can be captured by a particular parametric family of Archimedean copulas. The \(\chi^2\) test relies upon an arbitrary binning of data to calculate the test statistic. In this paper, we have used equal probability binning in all procedures to minimise error. Furthermore, all data is transformed prior to binning into uniform variates that are uncorrelated under \(H_0\). The classical Pearson’s \(\chi^2\)-statistic is then calculated according to the following equation:

\[
\chi^2 = \sum_{i=1}^{n} \frac{(N_i^{obs} - N_i^{exp})^2}{N_i^{exp}} \quad (10)
\]

Figure 1: Copulas.

Gaussian copulas are the most used in practice partly because their parameters can be estimated very easily by empirical means. Moreover, they extend the simple multidimensional gaussian framework naturally. The Student’s copula is similar to the Gaussian copula, but it has an extra parameter to control the tail dependence. Student copulas are of interest because they include the gaussian copulas family as a limit case. But contrary to the latter, they exhibit dependence in the tails. As noticed in (Ling 2003), the three other families represent three degree of dependence. Gumbel (resp. Clayton) copulas allow to modelize dependence in the right (resp. left) tails. After this description, the main question is: which is the best structure of dependence being able to be adapted to the studied phenomenon?
where \( N_i^{obs} \), \( N_i^{exp} \) are the number of observed and expected observations in the \( i^{th} \) bin respectively. To calculate the \( p \)-value, we note that the degrees of freedom are simply the number of bins that are not empty, and thus obtain the \( p \)-value from statistical tables.

It’s possible to get optimal copula in terms of log-likelihood by using one with lowest likelihood. While the \( \chi^2 \) test is the most commonly-used statistical test for goodness-of-fit, it should be noted that the binning is entirely arbitrary and can produce different \( p \)-values depending upon which binning procedure is used. With this in mind, we make use of five different binning procedures for \( \chi^2 \) testing in this paper. Then, the optimal copula will permit to better simulate the studied phenomenon. From these simulations, we describe: how to estimate the risk by using the failure criteria \((D, \Gamma_N)\)?

5 BIVARATE RISK ANALYSIS

This study considers only the bivariate (2-dimensional) case. Let be considered 2 random variables \((X, Y)\) representing respectively the vehicle’s criteria of failure \((D, \Gamma_N)\). A law of marginal distribution could be adjusted by using experimental data. Let \( F_1 \) and \( F_2 \) be the distributions of \( X \) and \( Y \). The values in risk at the safety threshold \( \delta^* \) are:

\[
VaR(X; \delta^*) = F_1^{-1}(\delta^*) \tag{11}
\]

\[
VaR(Y; \delta^*) = F_2^{-1}(\delta^*) \tag{12}
\]

We seek to aggregate these two values at risk. Let \( F_{1+2} \) be the distributions of \( X \) and \( Y \):

\[
F_{1+2}(z) = \int \int_{x+y \leq z} dC(F_1(x), F_2(y)) \tag{13}
\]

We consider continuous distributions where \( F \) is the joint cumulative distribution function (Cdf) of the random vector \((X, Y)\) and \( F_X \) and \( F_Y \) are the marginal Cdfs of \( X \) and \( Y \) respectively. Note particularly that \( X \) and \( Y \) do not necessarily have the same distribution, and that the joint distribution may differ again: for example, it is quite possible to link a normally-distributed variable and an exponentially distributed variable together through a bivariate gamma function. Bivariate copulas further satisfy three necessary and sufficient properties (Joe 1997).

By fixing \( \delta^* \in \mathbb{R}_+^* \), we can estimate the vehicle multidimensional risk assessment in this form:

\[
VaR(X + Y; \delta^*) = F_{1+2}^{-1}(\delta^*) \tag{14}
\]

The risk value depends on the marginal distributions, but also of the copula \( C \). The experiment results will show the relevance and effectiveness of this method for vehicle risk assessment.

6 METHODOLOGY APPLICATION

The aim of this section is to apply the proposed method. It consists to estimate the bivariate risk of failure trajectory by using copula method. Copulas are useful for examining the dependence of multivariate random vectors (criteria). Let us remember that the relative distance \((D)\) and lateral acceleration \((\Gamma_N)\) are criteria used to estimate the 2-dimensional risk. From real trajectories, we determine the dependence between criteria. Then, we choose a optimal copula and estimate the risk of failure trajectory. Finally, we make a comparison between risks in order to show the effect of the criteria dependence in risk assessment.

6.1 Criteria dependence analysis

The figure 2 makes it possible to apprehend the form of dependence existing between the criteria \((D, \Gamma_N)\). One gives below a graphic illustrating the dependence between empirical uniforms. In the event of perfect independence, we should find in each square of graph 232/20, that is to say approximately 12 points. The

![Structure of dependence](image)

Figure 2: Structure of dependence between criteria \((D, \Gamma_N)\).

Figure 2 shows that the cloud points is very close to the second bisector. It corresponding to a negative dependence between \( D \) and \( \Gamma_N \). Then, the cloud points is very concentrated on the two tally of the first bisector. That, means strong dependence between criteria. In addition to the direction and the intensity of the dependences, this graph gives first information on the dependence of tail. We notice central values more “pricked” and thicker tails of distribution. In fact, the empirical distribution presents rare observations with a slower decrease adjusted in power whereas the normal law has an exponential decay. We are going to fit copulas in order to choose the optimal copula.

6.2 Choice of copula

We suggest a \( \chi^2 \) test of fit for parametric families of bivariate copulas. The \( \chi^2 \) relies upon an arbitrary of data to calculate the statistic test. Most of usual copulas are rejected by the test at the \( p \)-value equal to 1%
in every case. Nonetheless, Frank’s copula seems to be the best one, especially for the couple \((D, \Gamma_N)\). At the opposite, the Clayton, Student and Gumbel copulas are strongly rejected. When Frank’s and Plackett’s copulas performances are closed, Gumbel’s copulas provide a worse fit. The Frank copula was retained to apprehend the nature of dependence between criteria. Note that the parameter of dependence increased for the Frank’s copula. However, more this parameter is high, larger is the dependence. Now, we simulate Frank’s copula to estimate the 2-dimensional risk.

### 6.3 Simulation result

The figure 3 represents 10000 observations resulting from the simulation of Frank’s copula for \((D, \Gamma_N)\).

![Figure 3: Simulated Dependent for the couple \((D, \Gamma_N)\).](image)

As \(D\) increases, \(\Gamma_N\) decreases and vice versa (negative correlation). It makes it possible to apprehend the form of the generated dependences and to appreciate their adequacy with the profile of copula. The simulation will permit to estimate the bivariate risk.

### 6.4 Risk analysis

The use of copulas enables to plot contours lines of the 2-dimensional risk \(VaR\) and to examine for a given \(\delta^*\), the marginal rate of substitution between the \(VaR\) of 2-univariate risks. The figure 4 shows \(P_f\) as function of \(\delta^*\) according to criteria and dependence structure of criteria. The risk \(VaR(\Gamma_N; \delta^*)\) is much more important for lateral acceleration \(\Gamma_N\) when we consider the dependence. At the opposite, it is less more important for the relative distance \(D\).

The figure 5 shows the evolution of \(P_f\) as function of \(\delta^*\). We calculated \(P_f\) by using reliability of components in series. That means as soon as criterion is not respected the system becomes failing. The function with discontinuous line represents the risk without the dependence between criteria. Afterwards, the function with a continuous line represents the risk with the dependence between failure criteria. The two values at risk \(VaR(D; \delta^*)\) and \(VaR(\Gamma_N; \delta^*)\) were aggregated to give \(VaR(D + \Gamma_N; \delta^*)\).

![Figure 5: \(P_f\) as function of \(\delta^*\) with/without dependence.](image)

The risk level is similar for the 2 functions in the interval \(\delta^* = [0.5; 0.55]\). We note that the risk may also lead to severe underestimation when the dependence between the risks does not take into account.

### 7 CONCLUSION

The aim of this study was to estimate the multidimensional risk of failure trajectory by using vehicle’s real observations data. For that, we proposed to use the copulas method in order to take into account the dependences between failure criteria of vehicle trajectory. The choice of the most adapted copula was made by using the test of \(\chi^2\). The Frank’s copula is adapted to our real observations data. From the simulations of this copula, the 2-dimensional risk is estimated by aggregating the two values at risk. Then, we have discussed how various estimation procedures impact the vehicle risk assessment. The results showed that risk is much more important with the dependence between failure criteria. In particular we have demonstrated how copulas can contribute to the computation of multivariate information.

To conclude, experiments have shown the relevance and effectiveness of this method. The method analysed above might help to better understand the risk of accident. So the present results of this new approach are promising for road safety. This method is relatively robust and applicable to all kinds of mobile objects (motorcycles, pedestrians,...) with different data acquisition systems and different characteristics.
REFERENCES


