FROM FEW TO MANY: USING COPULAS AND MONTE CARLO SIMULATION TO ESTIMATE SAFETY CONSEQUENCES

Vindhya Venkatraman[1], John D. Lee[1], Chris W. Schwarz[2]
University of Wisconsin-Madison[1]; University of Iowa-Iowa City[2]
Madison, Wisconsin[1]; Iowa City, Iowa[2]
venkatraman@wisc.edu, jdlee@engr.wisc.edu, chris-schwarz@uiowa.edu

Summary: With the introduction of more advanced vehicle technology, it is paramount to assess its safety benefit. Advanced driver assistance systems (ADAS) can reduce crashes and mitigate crash severity, if designed appropriately. Driver behavior models are integral to the ADAS design process, complementing time and resource intensive human participant experiments. We introduce a method to model driver responses to forward collision events by quantifying multivariate behavior with copulas and Monte Carlo simulation. This approach capitalizes on the data from small samples of crash events observed in naturalistic or simulator studies. Copulas summarize data by capturing the underlying joint distribution of variables, and Monte Carlo methods can be used to repeatedly sample from these distributions. A driver model can be parameterized with these samples, and run on a desktop driving simulation environment.

INTRODUCTION

Driving is undergoing a transformation. Vehicles with Advanced Driver Assistance Systems (ADAS) such as collision warning and imminent braking systems are already in the market, with the promise of more intelligent vehicles being available in the near future. Understanding driver responses to crash situations can guide design and evaluation of these ADAS interventions. Assumptions about drivers' responses and the environment inform algorithm design. These factors can include demographics (e.g. age, and gender), event kinematics (e.g. lane position, headway to forward vehicle, speed), traffic rules (e.g. traffic rules, cultural differences between drivers in the U.S and Europe), and vehicle dynamics (e.g. weight and deceleration profile). Computational models of the driver make assumptions about how these factors affect drivers' response explicit. The driver model is coupled with a model of the ADAS and their joint performance is studied in a simulated roadway environment. The outcomes, such as crash probability and severity, can be used to estimate ADAS safety benefits. These computational models support a more rapid iterative design of ADAS, than is possible with time and resource intensive human participant experiments.

Data from driving simulator and naturalistic studies provide a detailed record of driver response behavior from which models could be developed. However, the development of realistic and robust driver models poses a substantial challenge. Often, elements of driver responses are assumed to be independent, but this may not hold and neglecting the interdependence could result in unreliable predictions. In addition, human participant data on behavior to crash-imminent situations is sparse. Multiple exposures to collision situations can change the expectations and responses of drivers. As a consequence, even with a driving simulator, collecting hundreds of responses for each participant is infeasible and so researchers are left with only one or two imminent collision responses from each driver. Developing accurate models of

driver response to imminent collision situations requires techniques that address the drawbacks of sparse and interdependent behavioral data.

MODELING TOOLSET: COPULAS AND MONTE CARLO SIMULATION

Drivers' responses to collisions reflect a dynamic process comprised of interdependent elements that play out over time: a driver's reaction time influences subsequent braking behavior. Copulas model the relationships between multiple variables of a complex system (Frees & Valdez, 1997) by capturing their dependence structure. The dependence structure of a composite response can be fully described using the marginal distributions of each variable and the association between them. Copulas have historically been popular in the fields of financial analysis (Melchiori, 2003), biological system modeling (Kim et al., 2008), and more recently applied to driver injury modeling in vehicle crashes (Yasmin, Eluru, Pinjari, & Tay, 2014). They offer a promising approach for describing the interdependence between elements of drivers' responses.

In comparing copulas to classical regression analysis, Frees and Valdez (1997) point out that, unlike studying a single dependent measure of interest, copulas allow for combining several dependent measures to describe behavior. As an example of multivariate outcomes in driving, the brake reaction time of a driver to a lead-vehicle braking event is one dependent measure that can be collected. In addition, there are other measures such as the deceleration profile, the peak brake force during the maneuver, and the steering profile. These measures are different elements of a composite response. Multivariate analysis is needed to model the interdependence of these elements. The advantage of copulas over other multivariate techniques is their capacity to model the outcomes generated by any distribution, rather than only the normal distribution (Frees & Valdez, 1997), as well as their ability to account for different associations between the variables.

Many families of copulas have been developed, and among these, the Archimedean approach is preferred because it compresses the multiple dimensions of the system outcome into one function which is called the generator of the copula (Frees & Valdez, 1997; Melchiori, 2003). The marginal distributions modeled by copula equations are uniformly distributed, hence the variables can be converted to the unit uniform distribution [0, 1] by a probability integral transformation. Their correlation coefficient is then estimated. A generator function of the joint distribution captures the inter-dependence of these marginal distributions using a measure of association. This measure of association is related to the correlation coefficient between the variables and Table 1 shows how the correlation coefficients can be used to estimate the parameters for three of the most popular Archimedean copulas. Once estimated, a copula can generate a joint distribution of the multiple elements of drivers' collision avoidance response.

The joint distribution of drivers' response can extrapolate behavior from tens of participant's data points to hundreds and thousands of similar responses. Monte Carlo simulation can sample from this joint distribution for parameters of a driver model. Different parameterized instances of the model can then be run within a simulation environment. The simulation environment consists of a scenario, a vehicle, an ADAS, and a driver. The scenario is represented by the lane and road networks, pavement types, and traffic such as a lead vehicle. The vehicle dynamics consist of the vehicle kinematic equations. ADAS models are usually linked to the vehicle dynamics. The

driver model and the ADAS contribute to vehicle control functions. This study used a simulation environment that replicated the scenario and vehicle dynamics used in the National Advanced Driving Simulator (NADS).

Table 1. Copulas that follow Archimedean approach (adapted from Frees & Valdez, 1997)

Name	Bivariate Representation u and v are the bivariates on unit uniform distribution	Measure of Association (α) as a function of correlation measure Kendall's τ	Description
Clayton	$(u^{-\alpha} + v^{-\alpha} - 1)^{-1/\alpha}$	$\dfrac{2\tau}{(1-\tau)}$ $\in [-1, \infty)$	Asymmetric copula with greater dependence in the negative tail over the positive tail.
Gumbel	$exp\left\{-[(-\ln u)^{\alpha} + (-\ln v)^{\alpha}]^{1/\alpha}\right\}$	$\dfrac{1}{(1-\tau)}$ $\alpha \in [1, \infty)$	Asymmetric copula with greater dependence in the positive tail over the negative tail.
Frank	$\dfrac{1}{\alpha} \ln\left(1 + \dfrac{(e^{\alpha u} - 1)(e^{\alpha v} - 1)}{e^{\alpha} - 1}\right)$	*No closed form* $\alpha \in (-\infty, \infty)$	Symmetric copula

METHODS

The dataset used in this study is from the Crash Warning Interface Metrics Phase 3 study commissioned by the National Highway Traffic Safety Administration. The study was conducted in 2012 at NADS, and tested the effects of ADAS on responses to crash-imminent forward collision and lane departure events (Brown, Marshall, Chrysler, & Schmitt, 2014). Only the forward collision event data were analyzed for this study. The assistance systems used were passive – they provided warnings without initiating responses. Different warning modes were studied – auditory beeps, head-up display, seat vibration, brake pedal pulse, and seatbelt tug – with a baseline no-warning condition. In addition to warning mode, the timing of the warning was varied as either an early warning at 3.5 seconds time to collision (TTC) or a late warning at 2.1 seconds TTC. The warning mode and timing were between-subjects and the collision events were within-subjects conditions. Each participant was exposed to two forward collision events and two lane departure events.

Data Reduction and Model Variables

Ninety-six participants, aged 35-55 years old, were part of the study. Eight participants, evenly divided by gender, experienced one of the twelve conditions (6 warning mode x 2 warning timing). The participants were asked to follow a lead vehicle at a posted speed limit of 40 mph, and engaged in a number recall task during the initiation of FCW events. The recall task required the participants to look at a screen to their right, behind the passenger head-rest (Figure 1a). A series of five random numbers were displayed, one at a time, on the screen. This drew the drivers' eyes away from the forward roadway, during which time the lead car swerved to reveal a

stopped car. The warning was triggered when the TTC with the stopped vehicle reached 3.5 s, for the late warning group, or 2.1 s, for the early warning group.

Only the seatbelt tug and the baseline no-warning conditions, at the late 2.1 seconds TTC were selected for analysis in this study. The two conditions were selected due to the difference in diverting the driver's attention back to the roadway from the task screen. The baseline condition resulted in the greatest visual distraction – participants looked away from the forward roadway for an average of 1.2 seconds. On the other hand, participants in the seatbelt warning condition looked away for the shortest duration – an average 0.8 seconds. The brake reaction time after the onset of alert (BRT) and the peak brake force (PBF) characterized drivers' response.

The experiment design stipulated 16 data points for each condition – 8 participants with two collision imminent FCW events each. However, analysis of the data revealed that the BRT and PBF were only mildly correlated for the second events. This indicates possible behavior changes due to exposure of the experiment manipulation. In addition to low correlation between the BRT and PBF in the second events, participants also tended to steer into the empty neighboring lane rather than brake. Such behavior is inconsistent driver response to actual collision situations, where steering and a combination of steering and braking is rare (Najm, Toma, & Brewer, 2013). Consequently, this analysis focuses on the response to the first event. After accounting for missing data, the baseline and seatbelt warning had 7 and 6 responses respectively.
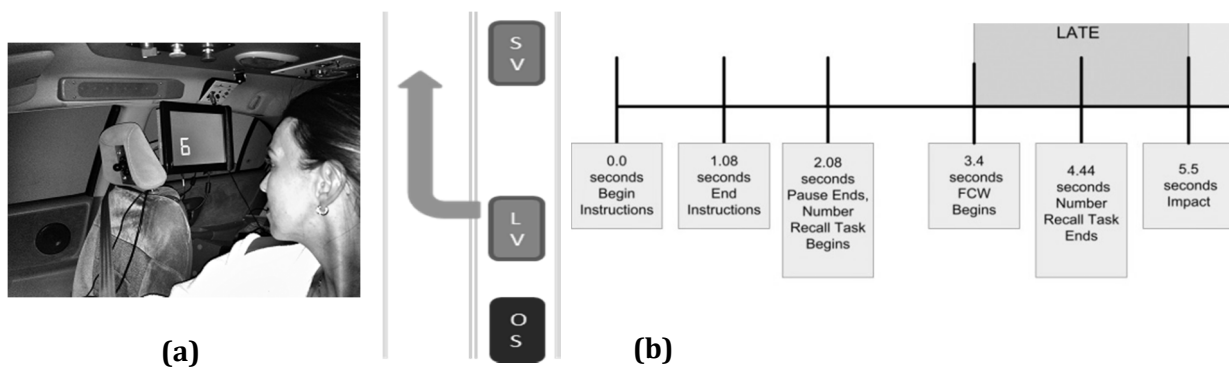

Figure 1. (a) Number presented to participant during FCW Event; (b) Diagram of FCW event scenario and timing- OS: OwnShip, LV: Lead Vehicle, SV: Stopped Vehicle

Copula Modeling

The two variables, BRT and PBF were modeled as a joint distribution using copulas. All statistical analyses used the R statistical software (R Core Team, 2014). The copulas were constructed using the Copula package (Yan, 2007), and distributions were fit using the fitdistplus (Delignette-Muller & Dutang, 2015) and MASS (Venables & Ripley, 2002) packages.

Histograms were plotted to visualize the distributions and the variables in both conditions were fitted to different distributions. A comparison of distributions using the Akaike Information Criterion (AIC) revealed that both BRT and PBF, under both the warning conditions, were best described by the normal distribution. The Table 2 summarizes BRT and PBF data. The BRT and PBF were negatively correlated: Kendall's tau for baseline = -0.43 and seatbelt tug = -0.33. This

indicates that the later a driver responded, the lesser the brake force. Visual analysis of the driver trajectories showed that many late responding drivers swerved around the vehicle and so relied less on braking to avoid collision. The Clayton copula from the Archimedean family was selected (Table 1) to model the joint distribution of BRT and PBF and to generate data from the joint distribution. From the values of tau, the respective copula parameters were calculated using the equations in Table 1 as baseline = -0.6 and haptic= -0.50. The generated data and the experimental points are shown in Figure 2. In addition to the copula data, another dataset was generated by sampling normal distributions, assuming independence between BRT and PBF. These data were used to assess the consequences of assuming independence of BRT and PBF.

Table 2. Summary statistics for the experimental data (CWIM Phase 3) as normal distribution

Variable Name	Baseline	Seatbelt Tug
Brake Reaction Time (BRT) in seconds	Mean=1.49, SD=0.51	Mean=1.16, SD=0.60
Peak Brake Force (PBF) in pound force	Mean=111.4, SD=40.1	Mean= 88.5, SD= 29.94

Simulation Architecture

The modeling environment was the NADS desktop simulation software, the NADS-MiniSim™. The driver and ADAS model was built using Matlab Simulink™. The ADAS model replicated the warning timing used for the data collection, issuing a warning when the TTC of the driver's vehicle dropped below 2.1 seconds. A simple driver response model was based on BRT and PBF, in which the driver is assumed to apply a constant brake force after a delay defined by the reaction time. The delayed response begins after the TTC to the lead vehicle dropped below 2.1 s, which is also when the warning is issued.
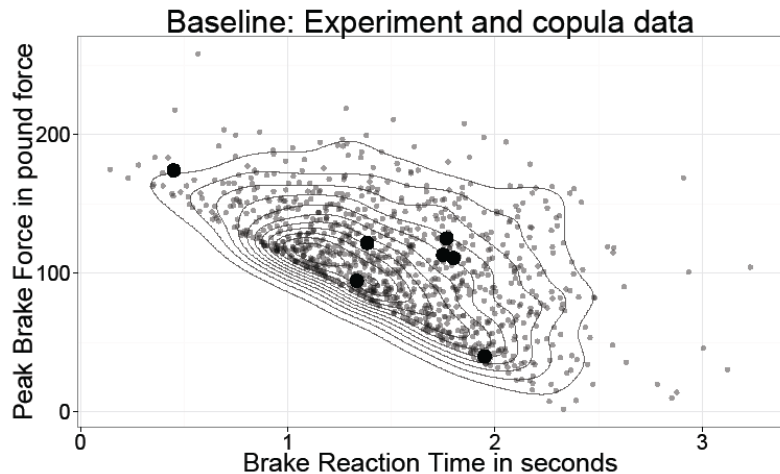

Figure 2. Data generated by the Clayton copula and the experimental data (large dots) for the baseline condition

RESULTS

The Adjusted Minimum Time to Collision (AMTTC) measured crash severity. AMTTC is defined as the time that a driver has to avoid collision (Brown, 2005). Positive values of AMTTC indicate the minimum time to collision and negative values indicate that the driver collided. The magnitude of the negative value indicates how much sooner the driver needed to react to avoid collision. AMTTC is calculated using the deceleration and velocity at impact.

The driver model used either BRT and PBF generated by the copula, or BRT and PBF generated by independent distributions. The comparison of AMTTC for 150 runs of each of the datasets (Figure 3) shows that the independence assumption predicts slightly more, but less severe crashes (less negative AMTTC). Similar small differences in results were observed for the

baseline condition. The results show no obvious benefit of the copula approach given the modest correlation between BRT and PBF (-0.33 for the seatbelt tug).
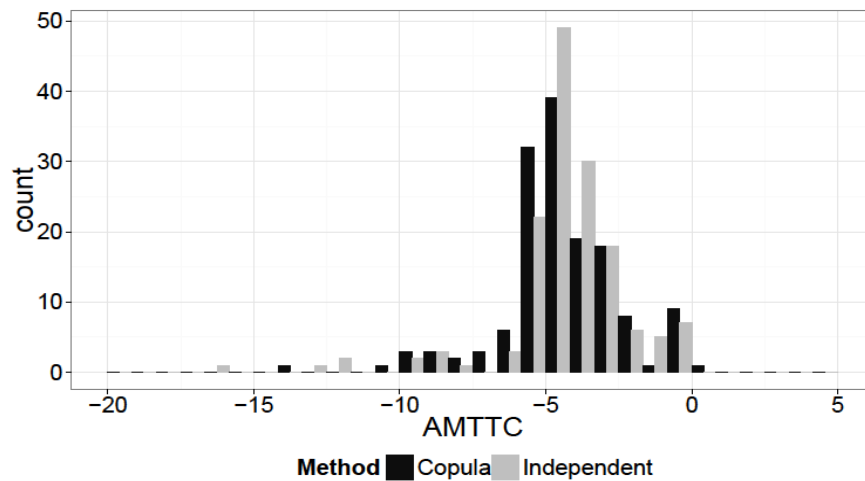

Figure 3. AMTTC comparison for the seatbelt tug condition between driver models using a copula to model the joint distribution of BRT and PBF and a model that assumes independent of BRT and PBF

DISCUSSION

The methods described in this study help address central challenges to developing accurate models of driver response to imminent collision situations: sparse and interdependent data. Statistical methods, such as copulas, that describe relationships among data are useful tools for understanding and modeling driver responses. Combined with Monte Carlo simulations, copulas provide an efficient method to extrapolate sparse experimental data.

However, the research described in this paper has multiple limitations. The dataset contains very few data points, with just 13 total observations across both conditions. This is clearly too sparse to make robust estimates of the underlying distributions. Indeed, copula modeling requires thoughtful application of the summary distributions—a single point can have a substantial influence on the copula. Collecting sufficient data for the copula plagues both controlled simulator experiments and naturalistic data collection. In both situations, filtering data to a set of qualitatively similar responses reduces the number of data points. As the event description gets more specific, the number of matching cases shrinks (Rasmussen, 1997). The benefit of copula modeling is that with many sparse datasets it could be possible to create a combined dataset that considers the dependencies between response elements and avoids potentially dangerous assumptions of independence.

Another limit of this work is that the driver model is overly simplified. The brake force is assumed constant, and the reaction time is a composite of the perception, motor, and vehicle response times. Other models have been developed, such as ACT-R and queueing network model that account for the dynamic process of response, as well as models specific to rear-end collisions (Brown, Lee, & McGehee, 2001). Perceptual cues of imminent collisions and the associated neurological processes have been identified and can be integrated into a driver model (Flach, Smith, Stanard, & Dittman, 2004). Future work in this effort will focus on the integration of perceptual cues into a dynamic process model of driver response.

ACKNOWLEDGEMENTS

REFERENCES

Brown, T. L., Lee, J. D., & McGehee, D. V. (2001). Human performance models and rear-end collision avoidance algorithms. *Human Factors*, *43*(3), 462–482.

Brown, T. L. (2005). Adjusted Minimum Time-To-Collision (TTC): A Robust Approach to Evaluating Crash Scenarios. In *DSC North America*, 40–48.

Brown, T.L., Marshall, D., Chrysler, S., & Schmitt, R. (August, 2014). Crash Warning Interface Metrics: Protocol Completion, *Internal Report,* National Advanced Driving Simulator, The University of Iowa.

Flach, J. M., Smith, M. R. H., Stanard, T., & Dittman, S. M. (2004). Chapter 5 Collisions: Getting them under control. In H. H. and G. J. P. Savelsburgh (Ed.), *Advances in Psychology* 135, 67–91. North-Holland.

Frees, E. W., & Valdez, E. A. (1997). Understanding Relationships Using Copulas. In *32nd Actuarial Research Conference*,pp. 1–25.

Kim, J.-M., Jung, Y.-S., Sungur, E. A., Han, K.-H., Park, C., & Sohn, I. (2008). A copula method for modeling directional dependence of genes. *BMC Bioinformatics*, *9*, 225. doi:10.1186/1471-2105-9-225

Marie Laure Delignette-Muller, Christophe Dutang (2015). fitdistrplus: An R Package for Fitting Distributions. *Journal of Statistical Software*, 64(4), 1-34.

Melchiori, M. R. (2003). Which Archimedean Copula is the Right One? *SSRN Electronic Journal*.

Najm, W. G., Toma, S., & Brewer, J. (2013). *Depiction of Priority Light-Vehicle Pre-Crash Scenarios for Safety Applications Based on Vehicle-to-Vehicle Communications* (No. DOT HS 811 732).

Rasmussen, J. (1997). Risk management in a dynamic society: a modelling problem. *Safety Science*, *27*(2), 183-213.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

Venables, W. N. & Ripley, B. D. (2002) Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0

Yan, J. (2007). Enjoy the Joy of Copulas : With a Package copula. *Journal Of Statistical Software*, *21*, 1–21.

Yasmin, S., Eluru, N., Pinjari, A. R., & Tay, R. (2014). Examining driver injury severity in two vehicle crashes - A copula based approach. *Accident Analysis and Prevention*, *66*, 120–135.