Summer Program on Transportation Statistics
August 14-18, 2017

SPEAKER TITLES/ABSTRACTS

Roya Amjadi
Federal Highway Administration

"Why Highway Crashes Have Recurring Patterns?"

Highway crash data with average of 39 thousand fatalities and 2.4 million nonfatal injuries per year have repetitive and predictable patterns, and may benefit from statistical predictive models to enhance highway safety and operation efforts to reduce crash fatalities/injuries. Highway crashes have patterns that repeat over fixed periods of time within the data set for crashes such as motorcycle, bicycles, pedestrians, nighttime, fixed object, weekend, and winter crashes. In some States, these crashes are weekly, monthly, or seasonally. Contributing factors such as: age category, light condition, weather, weekday, underlying state of the economy, and others impact these variations.

Anjishnu Banerjee
Medical College of Wisconsin

"Injury Risk Curves from Impact Biomechanical Experiments for Risk Analysis in Motor Vehicle Crashes"

Injury probability curves form the primary basis to mitigate injuries and fatalities in environments such as motor vehicle crashes. They are used in US Federal Motor Vehicle Safety Standards (FMVSS) for ensuring crashworthiness and safety of vehicles sold in the USA, consumer information appearing as star ratings in the MSRP stickers displayed on the automobiles for public awareness, and designs of safety systems such as seatbelts and airbags, and their evaluations using crash test devices, also termed anthropomorphic test devices or dummies.

The presence or absence of injures identified from testing human cadavers (termed post mortem human subjects, PMHS, in impact biomechanics literatures) are used in conjunction with biomechanical outcomes such as force, deflection, and acceleration to derive injury probability curves. They are then converted to dummy risk curves for purposes stated above. For example, the US frontal and side impact FMVSS have used results from human cadaver tests and applied to Hybrid III and ES-2re dummies, respectively. Experimental design followed in PMHS impact biomechanics research yields: (1) right censored data wherein the subject responds to the dose (impact) with no injury, (2) left censored data wherein the subject responds to the dose with injury, or, (3) interval censored data wherein the same subject responds to the lower dose with no injury and a higher dose with injury. The exact dose for injury (uncensored data point) is difficult to discern from these tests and have not been used in any Standards/regulations. This study focuses on the statistical methods, in particular the use of Bayesian hierarchical modeling techniques for the development of these injury probability curves and a discussion of experimental design in these contexts.

Chou-Lin Chen
NHTSA

"Statistical and Data Issues for Automatic Driving Safety Analysis and Research"

The downward trend of motor vehicle traffic crashes has been a remarkable success story of saving lives and preventing injuries on nation's highway. The underpinning of this success is a data-driven approach to tackle many traffic issues under NHTSA's leadership and the safety community's continuous efforts. Despite these efforts, 35,092 people died on United States' roadways in 2015 and 94% of traffic crashes can be attributed to human choices or errors. Traditional traffic safety data are from a wide variety of sources and cover areas of crash, roadway, vehicle, driver, citation, and EMS. As both automobile technology and communication innovation advance rapidly, transportation in the future will be very different than it is today. These changes require the transportation community to embrace the transformation and plan ahead to maximize these advancements to improve safety and mobility. This presentation will discuss the role of traditional traffic data in improving highway safety and the data needs for the future development of automated vehicles, roadway infrastructure, and other regulatory tools to ensure technologies are safely introduced and achieve their full potential.

Carol Flannagan
University of Michigan

"Evaluating Safety in Self-driving Cars: data and analytical challenges"

As the analysis by Kalra & Paddock (2016) demonstrated, traditional crash data and analysis approaches may require hundreds of millions or billions of self-driving miles to achieve sufficient power to demonstrate that automated vehicles (AVs) have lower injury/fatality risk than human-driven vehicles. Moreover, crash risk for AVs is a moving target as algorithms and systems change, and the mistakes AVs will make are not necessarily the same mistakes humans make. Thus, we need to rethink both the data that will make up transportation safety datasets in the near future as well as the analytical approaches used. I will present some newer data-collection approaches along with some specific challenges that might call for different analytical approaches than are being used for crash data today.

Feng Guo
Virginia Tech University

"Statistical Issues of Evaluating Traffic Safety Using Naturalistic Driving Study"

This talk will focus on challenges in analyzing naturalistic driving study (NDS) from the perspectives of epidemiological design, causal inference, recurrent event models, and other related issues such as crash surrogates. The application of NDS to autonomous driving safety evaluation will also be discussed.

John Ivan
University of Connecticut

“Application of Spatio-Temporal Methods to Identify Crash Prone Speed Conditions on Limited Access Roadways”

Crashes on limited access roadways typically occur due to drivers being unable to react in time to avoid collisions with vehicles ahead of them either moving slower or merging unexpectedly. Prevailing traffic stream conditions with high volume and low or variable speed downstream of low volume and high speed conditions can increase the possibilities for such collisions to occur. Real time trajectories of vehicles collected through crowdsourcing methods can give information about the distribution of speeds in the traffic stream by space and time. Spatio-temporal models relating these observed speed distributions to the occurrence of crashes or near crashes can help to identify crash prone traffic conditions as they arise, offering the opportunity to warn drivers before crashes occur.

Vishesh Karwa
Harvard University

“Causal Inference in Transportation Safety Studies: Comparison of Potential Outcomes and Causal Diagrams”

The research questions that motivate transportation safety studies are causal in nature. Safety researchers typically use observational data to answer such questions. In this talk, I will focus on exploring the applicability of two frameworks “Causal Diagrams and Potential Outcomes” to answer causal questions for a specific transportation safety problem. The causal effects of pavement marking retroreflectivity on safety of a road segment were estimated. More specifically, the results based on three different implementations of these frameworks on a real data set were compared: Inverse Propensity Score Weighting with regression adjustment and Propensity Score Matching with regression adjustment versus Causal Bayesian Network. The effect of increased pavement marking retroreflectivity was generally found to reduce the probability of target nighttime crashes. However, we found that the magnitude of the causal effects estimated are sensitive to the method used and to the assumptions being violated.

Karthik Konduri
University of Connecticut

“Trajectory-based Regression Approach to Predict Real-Time Traveler Information Using Crowdsourced Location Traces”

Advanced traveler information systems (ATIS) are increasingly being deployed as a way to address the congestion issues faced by transportation systems. Traditionally, information in ATIS solutions is collected using physical sensor networks which are prohibitively expensive to install, costly to maintain and operate, limited in their coverage and suffer from unreliability. In an effort to overcome these challenges, crowdsourcing based solutions are increasingly being deployed that combine technological advances with new paradigms of information sharing to collect data. In crowdsourcing-based ATIS, users equipped with location-aware and data-enabled portable devices collect and share information. However, despite growth and promise of crowdsourcing-based ATIS, number of issues about its feasibility and applicability remain.

A critical challenge with crowdsourcing is that the data collected is incomplete as it only provides information about subset of links from the entire network. This is primarily because at any given moment, users in a region are not traversing all links. Further, even if users traverse all links, only a small portion of those users participate in crowdsourcing solutions. In this presentation, a trajectory based regression technique that combines incomplete data (about travel conditions on a subset of links afforded by crowdsourced solutions) to predict traveler information for the entire transportation network will be presented. The methodology is demonstrated by utilizing location traces of shuttles serving the University of Connecticut to predict estimated arrival time for any given origin-destination pair. Further, robustness of the methodology is presented by varying the coverage and sampling frequency of the location traces.

Eric Laber
North Carolina State University

"Efficient Planning in Large Spatio-temporal Decision Problems"

Motor vehicle accidents are responsible for tens of thousands of fatalities and millions of injuries each year. Thus, improving the safety of our roadways is a top priority for state and federal agencies. An adaptive resource allocation strategy for roadway safety is a sequence of decision rules, one per time point, that maps up-to-date information on roadway and traffic information to a recommended investment of current resources into roadway safety projects. These investments could include structural changes, e.g., installing a stoplight at an intersection or adding reflective markers along a road segment, or data-gathering actions, e.g., installing embedded sensors to measure traffic flow and driving behaviors at selected locations. An optimal allocation strategy maximizes the mean of some cumulative measure of roadway safety. Estimating an optimal allocation strategy for roadway safety is challenging as: (i) data are noisy, incomplete, and not generally collected in a designed experiment; (ii) interventions can induce spillover effects complicating causal inference; and (iii) system dynamics are generally non-stationary. We discuss a general framework for estimating optimal allocation strategies for roadway safety that can be applied with noisy, heterogeneous, and non-stationary data streams. We discuss some of the key technical challenges and lines for future research.

Fan Li
Duke University

"Causal Inference in Traffic Safety Studies"

A central goal in traffic safety research is to evaluate the effectiveness of traffic safety programs and countermeasures from observational studies, which is inherently a causal inference problem. In this talk we review several causal inference methods highly relevant to traffic safety research, including propensity scores, difference-in-difference, and simulation based sensitivity analysis. We make the connection to the established methods in the traffic safety literature such as the empirical Bayes method. These methods will be illustrated with a real data set on rumble strips in Pennsylvania.

Sharon-Lise Normand
Harvard University

"Improving Motor Carrier Safety Measurement"

This talk will describe the findings from a congressionally mandated study to assess the methodology used by the U.S. Federal Motor Carrier Safety Administration (FMCSA) in its evaluation of motor carrier safety. A 12-person National Academies of Sciences, Engineering, and Medicine committee developed recommendations to improve the collection and analyses of enforcement data used to monitor roughly 550,000 commercial motor carriers.

Sharon-Lise Normand on behalf of the National Academies Panel on the Review of the Compliance, Safety, and Accountability Program of the Federal Motor Carrier Safety Administration

Victoria Nneji
Duke University

"What about the Driver in Driverless Cars?"

As we prepare for a future of driverless cars, what new risks must we work to understand? Despite the connotation of driverless, we can expect that humans will remain in the loop at each iteration of increasingly autonomous technology integration. While our technology is advancing, our population and economics are also transitioning to present challenging paradigm shifts that we should account for in assessing the risks of driverless cars. Let us take this holistic systems engineering approach to exploring transportation at the Statistical and Applied Mathematical Sciences Institute.

Eun Sug Park
Texas A&M University

"Bayesian Analysis of Multivariate Crash Counts Using Copulas"

There has been growing interest in jointly modeling correlated multivariate crash counts in road safety research over the past decade. To assess the effects of roadway characteristics or environmental factors on crash counts by severity level or by collision type, various models including multivariate Poisson regression models, multivariate negative binomial regression models, and multivariate Poisson-Lognormal regression models have been suggested. We introduce more general copula-based multivariate count regression models with correlated random effects within a Bayesian framework. Copulas provide a flexible way to construct valid multivariate distributions by decomposing any joint distribution into a copula and the marginal distributions. Our models incorporate the dependence among the multivariate crash counts by modeling multivariate random effects using copulas. Overdispersion as well as general correlation structures including both positive and negative correlations in multivariate crash counts can easily be accounted for by this approach. Our copular-based models can also encompass previously suggested multivariate negative binomial regression models and multivariate Poisson-Lognormal regression models.

The proposed method is illustrated with the crash count data of five different severity levels collected from 451 three-leg unsignalized intersections in California.

Co-authors: Rosy Oh, Man-Suk Oh, and Jae Youn Ahn

Michael Porter
University of Alabama

"Understanding and Evaluating Predictive Crash Models"

This talk will discuss how predictive crash models should be understood and evaluated. The focus will be on evaluating point process models that estimate the intensity and density of crash events over time. By viewing the crash events as realizations from a marked point process (in space or on a grid), a new understanding of *hotspots* emerges. The hotspot detection problem can then be solved in a decision theoretic manner that addresses the trade-offs between interventions and cost. Methods to compare competing models are also addressed. If time permits, I will also discuss how machine learning models can be used to construct better SPFs by removing the rigid functional form in the standard linear models and incorporating multi-way interactions.

Nalini Ravishanker
University of Connecticut

"Dynamic Modeling and Prediction Approaches for Sparse Spatial and Temporal Data"

This talk will review dynamic modeling and prediction for temporal and spatio-temporal data and describe algorithms for suitable state space models. Use of dynamic models for modeling crash types by severity will be briefly illustrated. Extension of these approaches for handling irregular temporal spacing and spatial sparseness will be discussed, and a potential application to travel time prediction will be explored.

Galen Reeves
Duke University

"Modeling the Impact of Driverless Cars on Traffic and Parking"

We explore some potential tradeoffs for parking and traffic in urban settings with driverless cars. Suppose that a driverless car could drop off passengers at a desired destination and then park in a remote location that is less crowded. While such a strategy could reduce the parking load, it would also increase congestion on the roads. To understand the overall impact, we introduce an asymmetric simple exclusion process (ASEP) to model the interaction between through traffic and parking traffic. By focusing on the steady state behavior, we are able to characterize achievable rate regions based on the number of the vehicles and their behaviors.

James Rosenberger
Pennsylvania State University

“Measurement Effects of Aging on Physical Properties of Pressurized Aging Vessel Residue”

Standard Practice for Accelerated Aging of Asphalt Binder Using a Pressurized Aging Vessel (PAV) and AASHTO T-240, Standard Method of Test for Effect of Heat and Air on a Moving Film of Asphalt Binder (Rolling Thin-Film Oven Test - RTFO), make no provision for the effects of ambient barometric pressure (elevation) on the properties of the test residue. An experiment to test for these elevation effects was developed based on reports from agencies located at elevations above 3,000 - 4,000 feet that shows variability exists between their test results and those from laboratories located at lower elevations. Statistical analyses performed as part of this project on data supplied by AASHTO-resources and the WCTG verified that the effect is of engineering consequence and that it is binder-specific, confirming an earlier analysis of the WCTG data. Preliminary results indicate that the preferred method for accounting for elevation (ambient barometric pressure) is to increase the RTFO test time as a linear function of elevation. A minimal experimental design will be presented to confirm the effect of elevation on these accelerated aging test results.

Co-author: David A Anderson

Erik Schlicht
Computational Cognition Group, C2-g

"Using Multifidelity Methods to Estimate the Risk Associated with Transportation Systems"

Multifidelity methods combine inexpensive low-fidelity simulations with costly but high-fidelity simulations to produce an accurate model of a system of interest at minimal cost. They have proven useful in modeling physical systems and have been applied to engineering problems such as wing-design optimization and robotics. This talk will overview recent efforts to extend multifidelity methods to estimate the risk associated with human-in-the-loop situations where humans are interfacing with technological systems, (e.g., aerospace and transportation). First, theoretical results will be presented that quantitatively evaluate different approaches to multifidelity modeling based on their predictive performance under different data conditions (Schlicht, et al, 2012). Then, the talk will provide a quantitative method for estimating the risk associated with candidate transportation technology, before it is developed and deployed (Schlicht & Morris, 2017). These empirical results utilize transportation data from low-fidelity simulation environments and high-fidelity sources. The multifidelity methods allow for candidate technologies to be evaluated at the stage of conception, and enables a mechanism for only the safest and most effective technology to be developed and released. Finally, the potential for using multifidelity models for other human-technology interactions will be discussed.

Schlicht, E.J., Lee, R., Kochenderfer, M.J., Wolpert, D., & Tracy, B. (2012). Predicting the behavior of interacting humans by fusing data from multiple sources. In the Proceedings of the 28th Conference on Uncertainty in Artificial Intelligence, (UAI-2012). https://arxiv.org/abs/1408.2053

Schlicht, E.J., & Morris, N.L. (2017). Estimating the risk associated with transportation technology using multifidelity simulation. Submitted to the International Conference on Machine Learning (ICML-2017). https://arxiv.org/abs/1701.08588

Jiguang Zhao
CH2M Hill

"Application of Bootstrap for Transportation Safety Study"

The highway safety professionals always face the challenge of analyzing the safety performance of roadway facilities with limited samples, especially when evaluating the safety performance of innovative roadway design and traffic management techniques. In the past several decades, researchers have utilized different statistical methods to draw meaningful conclusions from limited samples, including bootstrap. This presentation is to review the current status of bootstrap method and its application on highway safety study, the advantages and disadvantages of the method, and possible improvements to widen its application on highway safety studies.