Statistical Error Components Frameworks for Modeling Autonomous Systems Safety

Lan Ventura, PhD Student (lanventu@ttu.edu);
Rohan Shrestha, PhD Student (rohan.shrestha@ttu.edu);
Narayan Venkataraman, Senior Research Associate (narayan.venkataraman@ttu.edu);
Venky Shankar, Professor and Barnhart Chair (venky.shankar@ttu.edu)

Department of Civil, Environmental, and Construction Engineering
Texas Tech University, Lubbock, TX

Introduction

Autonomous systems safety includes several considerations with respect to methodology, modeling, and interpretation of findings:

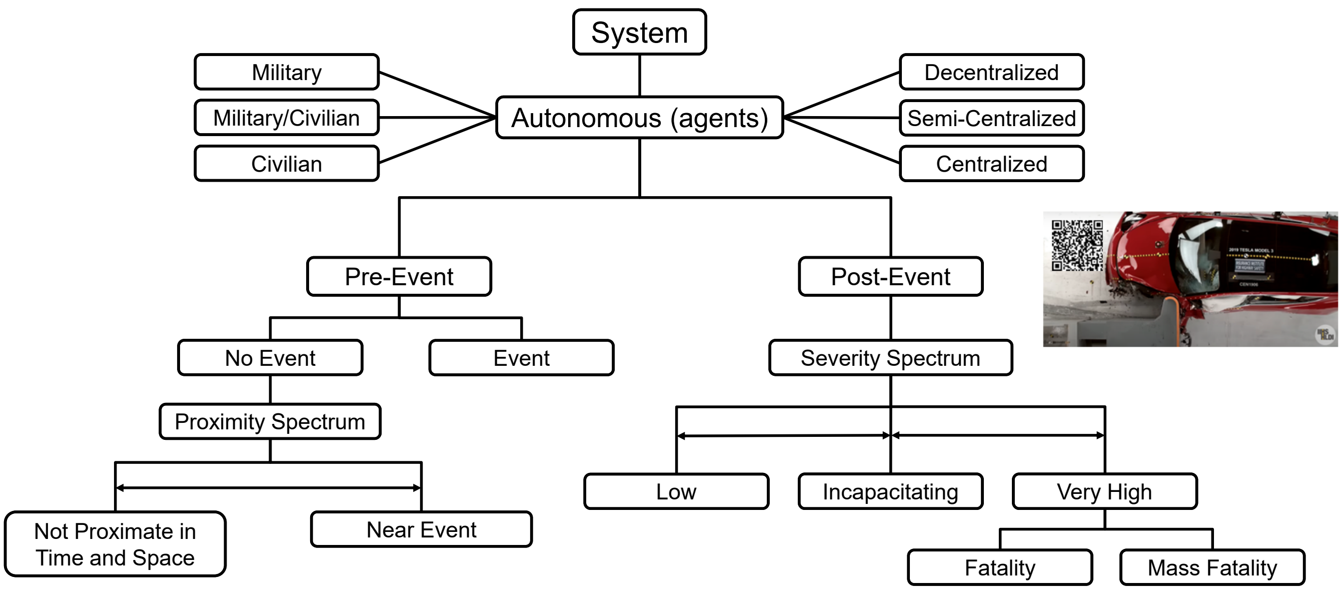
- Diversity and complexity in autonomous systems architecture
    - Diversity gives rise to unobserved heterogeneity
    - Complexities arise from nonlinearities in system interactions
- Autonomous systems safety is a complex, probabilistic challenge
- System parameters would be statistically biased, can compromise predictive accuracy, and would limit model portability

Modeling Results

Significant variables in autonomous involved crashes included:

- Traffic control type (signalized versus stop controlled)
- Driver contributing factors (inattention/disregarding)
- Road function (city street/interstate/farm-to-market)
- Fixed object collision involvement
- Nonlinear random parameter effects
    - Proportion of autonomous capable vehicles
    - Number of occupants involved
    - Number of vehicles involved

Follow the link in the QR code for our published application of this modeling methodology in the context of fixed object collisions.



Statistical Methodology

The outcome probabilities of a mixed logit model of crash injury severity, which accounts for unobserved heterogeneity in the data and error components, can be derived as:

$$P_n(k) = \frac{EXP[\beta'_{kn}X_{kn} + \sum_{m=1}^{M} d_{km}\theta_m \exp(\gamma'_m he_n)E_{nm}]}{\sum_{q=1}^{K} EXP[\beta'_q X_{qn} + \sum_{m=1}^{M} d_{qm}\theta_m \exp(\gamma'_m he_n)E_{nm}]}$$

where $P_n(k)$ is the probability of occupant n having the crash injury severity k, β_{kn} is a vector of estimable parameters, X_{kn} is a vector of explanatory variables that affect occupant's injury severity level k, E_{nm} is the individual specific underlying random error component, d_{km} is one if E_{nm} appears in the severity class k and zero otherwise, θ_m is the scale factor for the error component, and $\exp(\gamma'_m he_n)$ is the heterogeneity in variance of the error components, which can be broken down as γ'_m (parameters in the heterogeneity in variance of the error components) and he_n(individual choice invariant characteristics that produce heterogeneity in the variances of the error components).[1]

Interpretation and Systems Analysis Challenges

The nonlinear mixed logit uncovers the ***nonlinear effects of vehicle technology exposure (example: proportion of autonomous vehicles in a crash, number of occupants, and number of vehicles)***. The nonlinear mixed logit approach is applicable in domains where there are unobserved correlations across multiple severity outcomes.

Challenges associated with autonomous systems safety analysis:

- Big data (data sparsity and statistical adaptations)
- Optimization (improving efficiency and performance)
- Variable selection (nuanced in the context of automation)

The future of autonomous systems safety necessitates a methodological bridge between AI and statistics at the nexus of:

- Sociotechnical environment
- Micro and macro policy spectrum
- Geopolitical constraints

[1] Greene, W., 2007. NLOGIT version 4.0: reference guide. Econometric Software, Inc.