Modeling Recurrent Safety-critical Events among Commercial Truck Drivers: A Bayesian Hierarchical Jump Power Law Process

Miao Cai
Saint Louis University, Saint Louis, MO 63104

Qiong Hu
Auburn University, Auburn, AL 36849

Amir Mehdizadeh
Auburn University, Auburn, AL 36849

Mohammad Ali Alamdar Yazdi
Johns Hopkins University, Baltimore, MD 21202

Alexander Vinel
Auburn University, Auburn, AL 36849

Fadel M. Megahed
Miami University, Oxford, OH 45056

Karen C. Davis
Miami University, Oxford, OH 45056

Hong Xian
Saint Louis University, Saint Louis, MO 63104

Steven E. Rigdon
Saint Louis University, Saint Louis, MO 63104

May 25, 2020

Abstract

JUST COMMENT ON TARGET JOURNALS, NOT REAL ABSTRACT. The target of this manuscript will be statistical journal, so the writing will involve quite a bit of math and simulation. The target journals I have in mind include: a) Journal of the American Statistical Association (JASA, impact factor: 3.412, rank: 5/123), b) Journal of Computational and Graphical Statistics (impact factor: 1.882, rank: 28/123), c) Statistics in Medicine (impact factor: 1.847, rank: 29/123), d) Journal of Applied Statistics (impact factor: 0.767, rank: 84/123). This will be written once the paper is complete.

1. INTRODUCTION

The methods used in trucking safety is usually predictive model, while reliability models that account for multiple events are less used.

Traditional data use retrospective crash reports collected at certain road segments.In studies using these data, there are at most one crash in one shift and recurrent data models do not make sense in these scenarios. However, there can be multiple unsafe driving events in shift generated in NDS, which motivate the application of innovative recurrent event models in transportation safety modeling.

Literature review of recurrent event analyses application in transportation safety science. The most common recurrent event analysis model is probably Poisson regression, which assumes the events in a time interval is generated by a homogeneous Poisson process (Kim et al., 2013). (Chen and Guo, 2016; Li et al., 2017; Liu and Guo, 2019; Liu et al., 2019; Guo et al., 2019; Li et al., 2018).

2. DESCRIPTION OF DATA

[Figure 1 about here.]

Here are the notations for the data:

- Driver $d: 1, 2, \dots, D$,
- Shift $s, 1, 2, \dots, S_d$,
- Trip $r: 1, 2, \dots, R_{d,s}$,
- SCE $i: 1, 2, \dots, I_{d,s}$.
- $t_{d,s,i}$: time to the i-th SCE for driver d measured from the beginning of the s-shift,
- $n_{d,s,r}$: the number of SCEs for trip r within shift s for driver d,
- $a_{d,s,r}$: the end time of trip r within shift s for driver d.

3. MODELS

3.1 Non-homogeneous Poisson Process (NHPP) and Power Law Process

We assume the time to SCEs t follows a non-homogeneous Poisson process, whose intensity function $\lambda(t)$ is non-constant. The intensity function is assumed to have the following function form

$$\lambda_{\text{PLP}}(t) = \beta\theta^{-\beta}t^{\beta-1}, \tag{1}$$

where the shape parameter β indicates reliability improvement ($\beta < 1$), constant ($\beta = 1$), or deterioration ($\beta > 1$), and the scale parameter θ determines the rate of events. Here we assume the intensity function of a power law process because it has a flexible functional form, relatively simple statistical inference, and is a well-established model (Rigdon and Basu, 1989, 2000).

3.2 Bayesian Hierarchical Power Law Process (PLP)

The Bayesian hierarchical power law process is parameterized as:

$$
\begin{aligned}
t_{d,s,1}, t_{d,s,2}, \cdots, t_{d,s,n_{d,s}}, \tau_{d,s} &\sim \text{PLP}(\beta, \theta_{d,s}) \\
\beta &\sim \text{Gamma}(1, 1) \\
\log \theta_{d,s} &= \gamma_{0d} + \gamma_1 x_{d,s,1} + \gamma_2 x_{d,s,2} + \cdots + \gamma_k x_{d,s,k} \\
\gamma_{01}, \gamma_{02}, \cdots, \gamma_{0D} &\sim \text{i.i.d. } N(\mu_0, \sigma_0^2) \\
\gamma_1, \gamma_2, \cdots, \gamma_k &\sim \text{i.i.d. } N(0, 10^2) \\
\mu_0 &\sim N(0, 5^2) \\
\sigma_0 &\sim \text{Gamma}(1, 1),
\end{aligned}
\tag{2}
$$

where $t_{d,s,i}$ is the time to the i-th event for driver d in shift s, $\tau_{d,s} = a_{d,s,R_{d,s}}$ is the length of time of shift s (truncation time) for driver d, and $n_{d,s} = \sum_{r=1}^{n_{d,s}}$ is the number of SCEs in shift s for driver d. The likelihood function of event times generated from a PLP for driver d in shift s is given in Rigdon and Basu (2000, Section 2.3.2, Page 60):

$$
\begin{aligned}
L_{d,s}(\beta, \gamma_{0d}, \gamma | \mathbf{X}_d, \mathbf{W}_s) &= \Big(\prod_{i=1}^{n_{d,s}} \lambda_{\text{PLP}}(t_{d,s,i}) \Big) \exp(- \int_0^{\tau_{d,s}} \lambda(u) du) \\
&= \begin{cases} \exp\Big(- (\tau_{d,s}/\theta_{d,s})^\beta \Big), & \text{if } n_{d,s} = 0, \\ \Big(\prod_{i=1}^{n_{d,s}} \beta \theta_{d,s}^{-\beta} t_{d,s,i}^{\beta-1} \Big) \exp\Big(- (\tau_{d,s}/\theta_{d,s})^\beta \Big), & \text{if } n_{d,s} > 0, \end{cases}
\end{aligned}
\tag{3}
$$

where $\mathbf{X}_d$ indicates driver specific variables (e.g. driver age and gender), $\mathbf{W}_s$ represents shift specific variables (e.g. precipitation and traffic), and $\theta_{d,s}$ is the function of parameters $\gamma_{0d}, \gamma_1, \gamma_2, \ldots, \gamma_k$ and variables $x_{d,s,1}, x_{d,s,2}, \ldots, x_{d,s,k}$ given in the third line of Equation 2. The full likelihood function for all drivers are:

$$
L = \prod_{d=1}^{D} \prod_{s=1}^{S_d} L_{d,s}(\beta, \gamma_{0d}, \gamma | \mathbf{X}_d, \mathbf{W}_s) \tag{4}
$$

where $L_{d,s}(\beta, \gamma_{0d}, \gamma | \mathbf{X}_d, \mathbf{W}_s)$ is given in Equation 3.

3.3 Bayesian Hierarchical Jump Power Law Process (JPLP)

Since the Bayesian hierarchical PLP in Subsection 3.2 does not account for the rests $(r : 1, 2, \ldots, R_{d,s})$ within shifts and associated potential reliability improvement. In this subsection, we proposes a Bayesian hierarchical JPLP, with the following piecewise intensity function:

$$
\begin{aligned}
\lambda_{\text{JPLP}}(t | d, s, r, \beta, \gamma_{0,d}, \gamma, \mathbf{X}_d, \mathbf{W}_s) &= \begin{cases} \kappa^0 \lambda(t | \beta, \gamma_{0,d}, \gamma, \mathbf{X}_d, \mathbf{W}_s), & 0 < t \le a_{d,s,1}, \\ \kappa^1 \lambda(t | \beta, \gamma_{0,d}, \gamma, \mathbf{X}_d, \mathbf{W}_s), & a_{d,s,1} < t \le a_{d,s,2}, \\ \cdots & \cdots \\ \kappa^{R-1} \lambda(t | \beta, \gamma_{0,d}, \gamma, \mathbf{X}_d, \mathbf{W}_s), & a_{d,s,R-1} < t \le a_{d,s,R}, \end{cases} \\
&= \kappa^{r-1} \lambda(t | d, s, r, \kappa, \beta, \gamma_{0,d}, \gamma, \mathbf{X}_d, \mathbf{W}_s),\ a_{d,s,r-1} < t \le a_{d,s,r},
\end{aligned}
\tag{5}
$$

where the introduced parameter κ is the percent of intensity function recovery once the driver takes a break, and $a_{d,s,r}$ is the end time of trip r within shift s for driver d. By definition, the end time of the 0-th trip $a_{d,s,0} = 0$, and the end time of the last trip for the d-driver within the s-th shift $a_{d,s,R_{d,s}}$ equals the shift end time $\tau_{d,s}$. We assume that this κ is constant across drivers and shifts.

The Bayesian hierarchical JPLP model is parameterized as

$$
\begin{aligned}
t_{d,s,1}, t_{d,s,2}, \cdots, t_{d,s,n_{d,s}}, \tau_{d,s} &\sim \text{JPLP}(\beta, \theta_{d,s}, \kappa) \\
\beta &\sim \text{Gamma}(1, 1) \\
\log \theta_{d,s} &= \gamma_{0d} + \gamma_1 x_{d,s,1} + \gamma_2 x_{d,s,2} + \cdots + \gamma_k x_{d,s,k} \\
\kappa &\sim \text{Uniform}(0, 1) \\
\gamma_{01}, \gamma_{02}, \cdots, \gamma_{0D} &\sim \text{i.i.d. } N(\mu_0, \sigma_0^2) \\
\gamma_1, \gamma_2, \cdots, \gamma_k &\sim \text{i.i.d. } N(0, 10^2) \\
\mu_0 &\sim N(0, 5^2) \\
\sigma_0 &\sim \text{Gamma}(1, 1),
\end{aligned}
\tag{6}
$$

The notations are identical with those in Equation 2 except for the extra κ parameter. Similarly, the likelihood function of event times generated from a JPLP for driver d on shift s is

$$
L^*_{d,s}(\kappa, \beta, \gamma_{0d}, \gamma | \mathbf{X}_d, \mathbf{W}_s) =
\begin{cases}
\exp\Big(-\int_0^{\tau_{d,s}} \lambda_{\text{JPLP}}(u)du\Big), & \text{if } n_{d,s} = 0, \\
\Big(\prod_{i=1}^{n_{d,s}} \lambda_{\text{JPLP}}(t_{d,s,i})\Big) \exp\Big(-\int_0^{\tau_{d,s}} \lambda_{\text{JPLP}}(u)du\Big), & \text{if } n_{d,s} > 0,
\end{cases}
\tag{7}
$$

where the piecewise intensity function $\lambda_{\text{JPLP}}(t_{d,s,i})$ is given in Equation 5.

However, since the intensity function depends on the trip r for the same driver d and shift s, it is hard to write out specific form of Equation 7. Instead, we can rewrite the likelihood function at trip level, where the intensity function λ_{JPLP} is fixed for driver d on shift s and trip r:

$$
L^*_{d,s,r}(\kappa, \beta, \gamma_{0d}, \gamma | \mathbf{X}_d, \mathbf{W}_r) =
\begin{cases}
\exp\Big(-\int_{a_{d,s,r-1}}^{a_{d,s,r}} \lambda_{\text{JPLP}}(u)du\Big), & \text{if } n_{d,s,r} = 0, \\
\Big(\prod_{i=1}^{n_{d,s,r}} \lambda_{\text{JPLP}}(t_{d,s,r,i})\Big) \exp\Big(-\int_{a_{d,s,r-1}}^{a_{d,s,r}} \lambda_{\text{JPLP}}(u)du\Big), & \text{if } n_{d,s,r} > 0,
\end{cases}
\tag{8}
$$

where $t_{d,s,r,i}$ is the time to the i-th SCE for driver d on shift s and trip r measured from the beginning of the shift, $n_{d,s,r}$ is the number of SCEs for driver d on shift s and trip r. Compared to the PLP likelihood function given in Equation 4 where $\mathbf{W}_s$ are assumed to be a constant during an entire shift, the rewritten likelihood function for JPLP in Equation 8 assumes external covariates $\mathbf{W}_r$ vary between different trips in a shift. In this way, JPLP can account for the variability between different trips within a shift.

Therefore, the overall likelihood function for drivers $d = 1, 2, \ldots, D$, their corresponding shifts $s = 1, 2, \ldots, S_d$, and trips $r = 1, 2, \ldots, R_{d,s}$ is:

$$
L^* = \prod_{d=1}^{D} \prod_{s=1}^{S_d} \prod_{r=1}^{R_{d,s}} L^*_{d,s,r},
\tag{9}
$$

where $L^*_{d,s,r}$ is a likelihood function given in Equation 8, in which the intensity function λ_{JPLP} has a fixed functional form provided in the last line of Equation 5 for a certain driver d in a given shift s and trip r.

[Figure 2 about here.]

4. SIMULATION STUDY

4.1 Simulation setting

We conducted a simulation study to evaluate the performance of our proposed JPLP. 1000 simulations were performed to each scenario with various number of drivers ($D = 25, 50, 75, 100$).

4.2 Simulation results

5. DATA ANALYSES

The hierarchical Bayesian PLP and JPLP were performance using the probabilistic programming language (Carpenter et al., 2017; Stan Development Team, 2018).

6. DISCUSSION

In this article we have proposed a Bayesian hierarchical jump power law process, which accounts for the characteristics of multiple rests within a shift among commercial truck drivers. The simulation results shows XXXX. In the application to a 496 truck driver NDS dataset, the results suggest that XXX.

REFERENCES

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. "Stan: A Probabilistic Programming Language." *Journal of Statistical Software*, 76(1) (2017).

Chen, C. and Guo, F. "Evaluating the influence of crashes on driving risk using recurrent event models and Naturalistic Driving Study data." *Journal of Applied Statistics*, 43(12):2225–2238 (2016).

Guo, F., Kim, I., and Klauer, S. G. "Semiparametric Bayesian models for evaluating time-variant driving risk factors using naturalistic driving data and case-crossover approach." *Statistics in Medicine*, 38(2):160–174 (2019).

Kim, S., Chen, Z., Zhang, Z., Simons-Morton, B. G., and Albert, P. S. "Bayesian hierarchical Poisson regression models: an application to a driving study with kinematic events." *Journal of the American Statistical Association*, 108(502):494–503 (2013).

Li, Q., Guo, F., Kim, I., Klauer, S. G., and Simons-Morton, B. G. "A Bayesian finite mixture change-point model for assessing the risk of novice teenage drivers." *Journal of Applied Statistics*, 45(4):604–625 (2018).

Li, Q., Guo, F., Klauer, S. G., and Simons-Morton, B. G. "Evaluation of risk change-point for novice teenage drivers." *Accident Analysis & Prevention*, 108:139–146 (2017).

Liu, Y. and Guo, F. "A Bayesian Time-Varying Coefficient Model for Multitype Recurrent Events." *Journal of Computational and Graphical Statistics*, 1–12 (2019).

Liu, Y., Guo, F., and Hanowski, R. J. "Assessing the Impact of Sleep Time on Truck Driver Performance using a Recurrent Event Model." *Statistics in Medicine*, 38(21):4096–4111 (2019).

Rigdon, S. E. and Basu, A. P. "The Power Law Process: A model for the Reliability of Repairable Systems." *Journal of Quality Technology*, 21(4):251–260 (1989).

—. *Statistical Methods for the Reliability of Repairable Systems*. Wiley New York (2000).

Stan Development Team. "RStan: the R interface to Stan." (2018). R package version 2.18.2. URL `http://mc-stan.org/`

List of Figures

1 NDS real-time ping data (colored points) and the aggregation process from pings to shifts and trips (colored arrows). 8
2 Simulated intensity function of power law process (left) and jump power law process. The x-axis shows time in hours since start and y-axis shows the intensity of SCEs. The red crosses mark the time to SCEs and the green vertical lines indicates the time of the rests. Parameter values for simulation: shape parameters $\beta = 1.2$, rate parameter $\theta = 2$, jump parameter $\kappa = 0.8$. 9

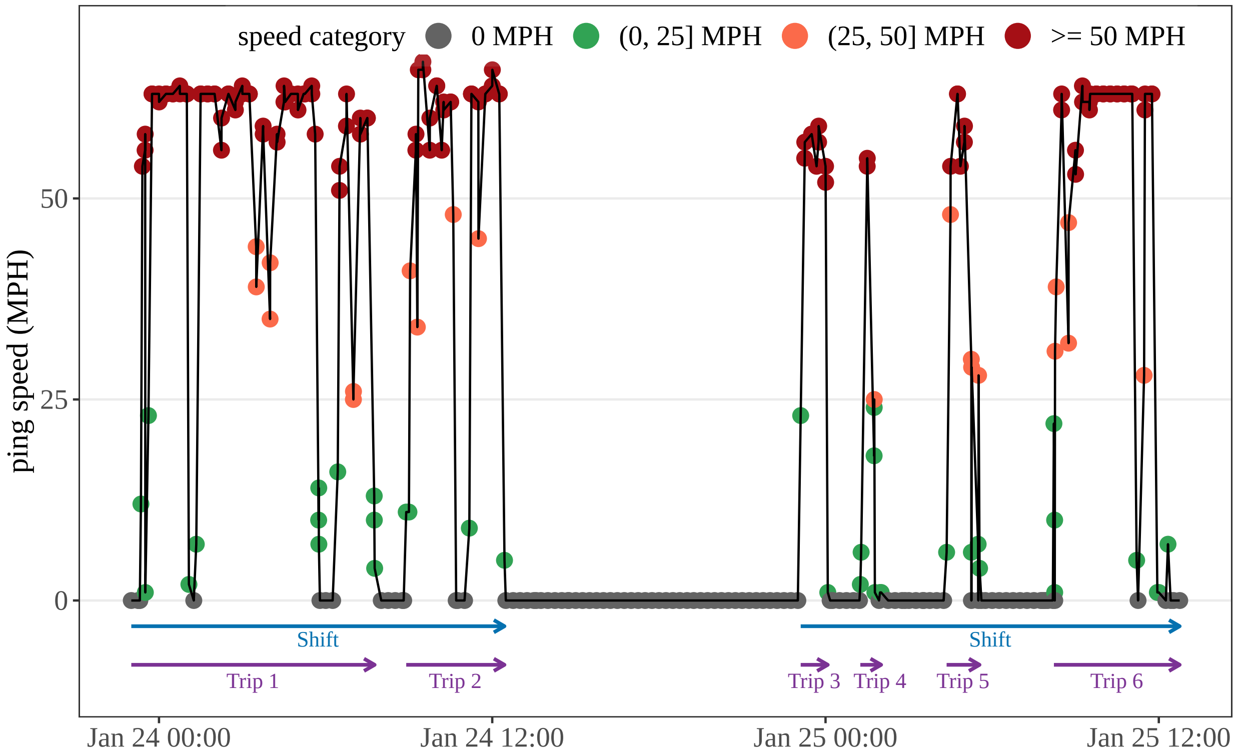

Figure 1: NDS real-time ping data (colored points) and the aggregation process from pings to shifts and trips (colored arrows).

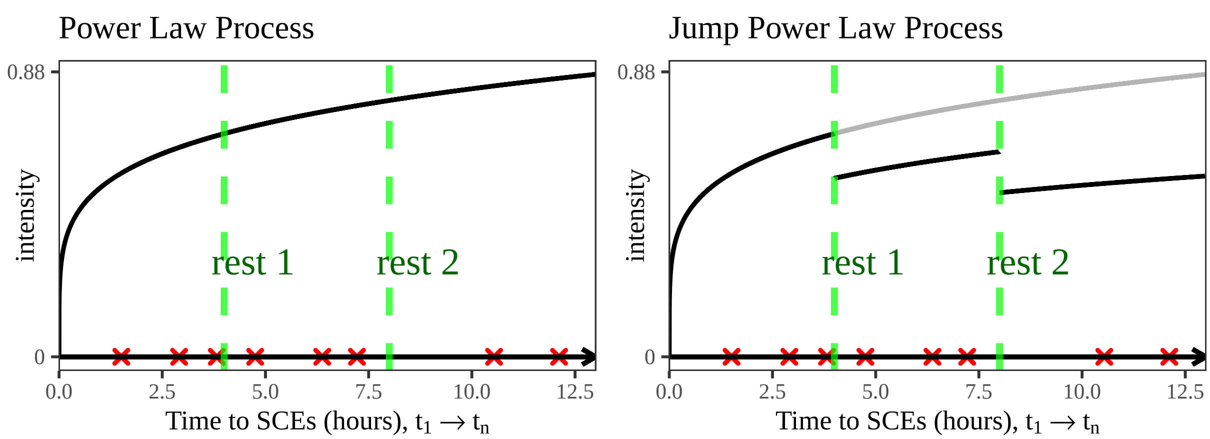

Figure 2: Simulated intensity function of power law process (left) and jump power law process. The x-axis shows time in hours since start and y-axis shows the intensity of SCEs. The red crosses mark the time to SCEs and the green vertical lines indicates the time of the rests. Parameter values for simulation: shape parameters $\beta = 1.2$, rate parameter $\theta = 2$, jump parameter $\kappa = 0.8$.