Accident Analysis and Prevention 108 (2017) 139–146

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Evaluation of risk change-point for novice teenage drivers

Qing Li[a], Feng Guo[b,c,*], Sheila G. Klauer[c], Bruce G. Simons-Morton[d]

[a] Department of Statistics, University of Wisconsin-Madison, United States
[b] Department of Statistics, Virginia Tech, 250 Drillfield Drive, Blacksburg, VA 24061, United States
[c] Virginia Tech Transportation Institute, United States
[d] Eunice Kennedy Shriver National Institute of Child Health and Human Development, United States

ARTICLE INFO

Keywords:
Change-point
Driving risk
Naturalistic Teenage Driving Study
Non-homogeneous Poisson process
Recurrent events

ABSTRACT

The driving risk of novice teenagers is the highest during the initial period after licensure but decreases rapidly. This paper applies two recurrent-event change-point models to detect the time of change in driving risks. The models are based on a non-homogeneous Poisson process with piecewise constant intensity functions. We show that the maximum likelihood estimators of the change-points can only occur at the event times and they are consistent. A simulation study is conducted to demonstrate the model performance under different scenarios. The proposed models are applied to the Naturalistic Teenage Driving Study, which continuously recorded *in situ* driving behaviour of 42 novice teenage drivers for the first 18 months after licensure using sophisticated in-vehicle instrumentation. The results indicate that approximately half of the drivers have lower risk after 73.0 h of independent driving after licensure while the risk for others increases. On the average the driving risk deceases after the change-point. The results provide critical information for safety education, safety countermeasure development, and Graduated Driver Licensing policy making.

1. Introduction

Traffic crashes are the leading cause of death for teenagers in the United States, and the fatality rate for teenage drivers is substantially higher than that for experienced drivers (National Highway Traffic Safety Administration, 2000). Driving risk is typically measured by the rate of safety critical events. Studies like Williams (2003) have shown that the driving risk among novice teenagers is the highest early in licensure, declines rapidly for a period of time right after independent driving, and changes slowly later. The rapid decline in crash risk after licensure has been documented in a number of recent studies (Guo et al., 2013; Lee et al., 2011; Mayhew et al., 2003; Simons-Morton et al., 2011). Reducing the high risk in the initial driving period is the foundation for the Graduated Driver Licensing (GDL) law and also the target of teenage driver safety education and countermeasure programs.

Various factors contribute to the high driving risk of teenage drivers. Jackson et al. (2013) used high G-force events, i.e., hard braking, to predict the occurrence of crash and near-crash (CNC) events and the findings implied that driving style could contribute to the driving risk. Kim et al. (2013) stated that having friends with poor driving habits and risky behaviour increased the driving risk. Klauer et al. (2014) analyzed the Naturalistic Teenage Driving Study (NTDS) data and concluded that performing secondary tasks while driving would significantly increase the driving risk and the impacts on teenagers were considerably higher than experienced drivers. Ouimet et al. (2014) showed that teenage drivers with higher cortisol level tend to have lower CNC risk and the driving risk decrease faster over time. Simons-Morton et al. (2014) concluded that the driving risk for teenage drivers was directly associated with total time eyes off the forward roadway. The high initial risk and rapid change after licensure are likely to be the results of multiple factors including drivers' experience increase and driving skills.

The duration of the initial high risk period is a key parameter of interest. Many GDL laws impose certain constraints on driving privilege, e.g., teen passengers and night time driving, to decrease the likelihood of CNC events. Using a naturalistic driving study approach, Simons-Morton et al. (2011) showed that the CNC rate declined significantly at six months after licensure. Mayhew et al. (2003) showed similar results with the Nova Scotia Motor Vehicle Accident Database. Guo et al. (2013) examined the variation of the CNC rate change across time and revealed substantial variations among teenage drivers: only about thirty percent of drivers at moderate risk showed a significant decrease in CNC rate after six months, while no significant change was found for high and low risk groups.

A common approach for evaluating the time trend of event rate is to aggregate data into pre-defined time intervals, e.g., three-month

* Corresponding author.
E-mail address: feng.guo@vt.edu (F. Guo).

http://dx.doi.org/10.1016/j.aap.2017.08.007
Received 11 May 2017; Received in revised form 10 July 2017; Accepted 4 August 2017

intervals, for analysis (Guo et al., 2013; Simons-Morton et al., 2011). The low safety-critical event rate observed in real life requires a long time period as base analysis unit for stable estimation. The long time interval, however, reduces the time resolution and can only detect the time of change in terms of months or even quarters. In addition, the calendar time interval used in many studies does not necessarily correspond well with the actual driving time, which is arguable to be a more direct measure of driving experience. It is useful to identify the change-points of crash rates in driving time which may quantify the amount of experience required for relatively safe driving (Simons-Morton et al., 2011). For example, according to the GDL of Virginia (Governor's Highway Safety Association, 2015), the teenagers have to drive at least 45 h under supervision to obtain a license. There is a need for more suitable analytical approaches and new data collection method for estimating the risk change in driving time.

The Naturalistic Driving Study (NDS) method is a novel driving data collection method characterized by continuous data collection from advanced in-vehicle instrumentation (Dingus et al., 2006). There is typically no specific instruction for drivers to ensure that the collected data truly reflect their normal driving behaviour. The NDS provides unrepresented *in situ* driving data under natural, non-experimental driving conditions that could not be obtained from crash database and driving simulator/test track studies. It is especially suitable for evaluating the rapid change in teenage driving risks. The Naturalistic Teenage Driving Study (NTDS) conducted between 2006 and 2009 examined the novice teenage driving. The study included 42 teenage drivers who had just obtained their Virginia driver's licenses; their vehicles were equipped with cameras, accelerometers, Global Positioning Systems, and other vehicle-monitoring devices. The crashes and near-crashes were identified from the collected driving data (Lee et al., 2011). This NTDS provides an opportunity to evaluate the change-points of driving risk with higher precision for novice teenage drivers.

Change-point models are a natural fit for detecting the risk change-point. Differing from terminal diseases, majority of crashes or safety critical events are not fatal and most likely will occur again in the future. In this paper, we focus on developing change-point models in the context of recurrent event. There have been a number of change-point detection models based on constant hazard functions but in a non-recurrent-event context (Loader, 1991; Matthews and Farewell, 1982; Nguyen et al., 1984; Yao, 1986). Ghosh et al. (1993) and Karasoy and Kadilar (2007) presented Bayesian estimators of the change-point assuming a piecewise constant hazard function. Mü and Wang (1994) reviewed different change-point estimators in hazard functions. Pons (2002) estimated the change-point in the time-varying covariates in a Cox model by maximizing the likelihood. Wu et al. (2003) proposed a non-parametric change-point estimator of the hazard function in the counting process context allowing for random censoring. Dupuy (2006) provided consistent estimators of the change-point for both hazard function and regression coefficients in a parametric survival regression model.

Piecewise constant intensity function is robust (Lawless and Zhan, 1998), and is the assumption of a large part of literatures on change-point detection in the recurrent-event context. Lawless and Zhan (1998) analyzed the interval-grouped recurrent-event data under such assumption, where only the event counts in given intervals were recorded. Achcar et al. (2007) and West and Odgen (1997) implemented likelihood approaches to estimate the change-point for only one individual with recurrent events. Frobish and Ebrahimi (2009) proposed a maximum likelihood estimator (MLE) and a Nelson-Aalen estimator for estimating the unknown change-point of multiple individuals with recurrent events. Li et al. (2017) proposed a Bayesian finite mixture model to detect the change-points in the driving risk of teenage drivers and cluster the drivers by risk patterns.

Recurrent-event models are commonly based on counting processes (Cook and Lawless, 2007, Chap. 1). A counting process is a non-decreasing stochastic process $\{N(t);\, t \geqslant 0\}$ with positive integer values, in which $N(t)$ is the cumulative number of events before or at time t. According to the Doob–Meyer decomposition theorem, $N(t) = \Lambda(t) + M(t)$, where $\Lambda(t)$ is the cumulative intensity process and $M(t)$ is a martingale (Andersen et al., 1993). If $\Lambda(t)$ exists, the intensity is $\lambda(t|\pounds_t) = \lim_{\triangledown \to 0} P(N[t,\, t+\triangledown) \geqslant 1|\pounds_t)/\triangledown$, where $\pounds_t$ is the process history before time t, $\triangledown$ is a small time interval, and $N[t,\, t+\triangledown)$ is the number of events between time t and $t+\triangledown$.

The Poisson process is canonical for event count analysis, of which the intensity function $\lambda(t|\pounds_t)$ has the form $\lambda(t)$. Given the marginal probability of an incident at time t, $\lambda(t)$ is the intensity function of a Poisson process. The non-homogeneous Poisson process (NHPP) is a Poisson process when the intensity parameter $\lambda(t)$ is not a constant (Ross, 2006, p. 32). Considering the NHPP $\{N(t);\, t \geqslant 0\}$ with cumulative intensity $\Lambda(t)$, the number of events that occurred between time zero and t is a random variable that is Poisson distributed with parameter $\Lambda(t)$. The intensity function $\lambda(t)$ of a NHPP can increase or decrease abruptly and the shift point in time is the change-point.

This paper advances the likelihood-based approach of Frobish and Ebrahimi (2009) in several ways. First, we relax the constraint that the censoring times be larger than the change-points for all subjects. Second, the intensity functions for individuals could vary, which fits the driving risk change better because of high variability among subjects. Third, we infer the standard errors (SEs) of the intensity estimators by the information matrix in identical-intensity models. The confidence intervals (CIs) and SEs of change-point estimators are obtained by block bootstrapping, where only the drivers are re-sampled with the event times left unchanged (Field and Welsh, 2007). The change-points of the bootstrapped samples are detected by the method in this paper and are used to calculate the CIs and SEs. Lastly, multiple change-points are allowed in our models and we propose a method to determine the number of change-points. All the change-points are assumed unknown, which is different from the approach in Frobish and Ebrahimi (2009) which considered two change-points and assumed the first change-point to be known.

The rest of the paper is as follows. Section 2 introduces the NTDS and presents an exploratory data analysis. Section 3 describes the change-points detection in the context of recurrent-event models. Section 4 reports the simulation findings, and Section 5 shows the NTDS data application results. Section 6 presents concluding remarks.

2. Naturalistic teenage driving methods

The NTDS is a naturalistic driving study focusing on novice teenage drivers. The participants consist of 22 females and 20 males with an average age of 16.4 years. Recruitment of newly-licensed teenagers was conducted via newspaper advertisements, flyers, and driving schools from the New River Valley and Roanoke Valley areas of the Commonwealth of Virginia. The participants were still subject to the GDL when they were in this study. To be eligible for the study, the drivers have to be younger than 17 years old and hold a provisional driver's license with the maximum independent driving time to be three weeks. (Lee et al., 2011). The data collection started within two weeks after licensure. The average driving time over the 18-month study period per participant was 250.6 h and the stand deviation (SD) was 107.5 h.

The driving risk is quantitatively measured by crashes and near-crashes (CNC) rate. A crash involves energy transfer between the participant vehicle and another vehicle or object; a near-crash is similar to a crash except that the driver makes an evasive maneuver to avoid a crash (Lee et al., 2011). It is a state-of-practice to combine crashes and near-crashes in driving risk evaluation mainly because of the relatively small sample sizes in naturalistic driving studies (Guo et al., 2010; Guo and Fang, 2013; Klauer et al., 2010, 2014; Ouimet et al., 2014; Simons-Morton et al., 2011). The validity of using crash and near-crash has been examined in previous studies (Guo et al., 2010).