A Poisson-Multinomial Mixture Approach to Grouped and Right-Censored Counts

Qiang Fu, Assistant Professor
Department of Sociology, The University of British Columbia, Vancouver, Canada，
qiang.fu@ubc.ca

Xin Guo, Assistant Professor
Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong,
x.guo@polyu.edu.hk

Kenneth C. Land, John Franklin Crowell Professor Emeritus

Department of Sociology and Center for Population Health and Aging, Duke University, Durham, NC, United States, kland@soc.duke.edu

The authors gratefully acknowledge financial support from a start-up grant from The Hong Kong Polytechnic University [Grant No. 1-ZVEB], Hampton New Faculty Award at The University of British Columbia, A Junior Scholar Grant at The Chiang Ching-kuo Foundation for International Scholarly Exchange and a 2015 Major Project of the National Social Sciences Foundation in China [Grant No. 15ZDB172]. Please direct all correspondence to Qiang Fu, Department of Sociology, The University of British Columbia, Vancouver, BC, Canada, V6T 1Z1 (Email: qiang.fu@ubc.ca).

Abstract: Although count data is often collected in social, psychological and epidemiological surveys in grouped and right-censored categories, there is a lack of statistical methods simultaneously taking both grouping and right-censoring into account. In this research, we propose a new generalized Poisson-multinomial mixture approach to model grouped and right-censored (GRC) count data. Based on a mixed Poisson-multinomial process for conceptualizing grouped and right-censored count data, we prove that the new maximum-likelihood estimator (MLE-GRC) is consistent and asymptotically normally distributed for both Poisson and zero-inflated Poisson models. The use of the MLE-GRC, implemented in an R function, is illustrated by both statistical simulation and empirical examples. This research provides a tool for epidemiologists to estimate incidence from grouped and right-censored count data and lays a foundation for regression analyses of such data structure.

Keywords: Grouped and right-censored count data, Mixed Poisson models, Zero-inflated Poisson distribution, Multinomial distribution, MLE-GRC

INTRODUCTION

Statistical distributions for analyzing count data are often built upon Poisson distributions (Allison and Waterman 2002; Cameron and Trivedi 1998; Greene 1997; Long 1997). One important goal of these Poisson family models is to estimate λ, or the average number of events occurring in a fixed interval of time, which is closely related to incidence in epidemiological research. Although statistical models are well developed for estimating λ from typical count data (e.g., once, twice, 3 times, 4 times, 5 times, 6 times,…), the estimation of λ becomes very challenging when response categories to survey questions consist of grouped counts (e.g., the response category "3-4 times" rather than the separate "3 times" and "4 times" categories) and are right censored (e.g., the upper end response category "5 or more times"). Recently, Poisson regression models based on maximum likelihood estimation (MLE) have been proposed for analyzing right-censored count data (Saffari and Adnan 2011; Saffari, Adnan and Greene 2012). In survival analysis of right-censored data, an interval-censored approach has been developed (Chen, Sun and Peace 2012; Fay and Shaw 2010; Lindsey and Ryan 1998). Statistical models for analyzing grouped Poisson data and their asymptotic properties have also been carefully investigated (Dickman et al. 2004; McGinley, Curran and Hedeker 2015; Rao and Scott 1999). Nevertheless, scholars remain unclear about the estimation of λ if the count data is both grouped and censored (GRC), because virtually no Poisson-based likelihood functions are readily developed for grouped and right-censored count data.

This paper presents a new multinomial-Poisson approach to model GRC count data, derives its asymptotic properties and applies this method to adolescent drinking data. Two members of the Poisson family of frequency distributions, the Poisson and Zero-Inflated Poisson (ZIP), are discussed in detail. The first step of the method is to introduce a conceptual multinomial model consisting of mixed Poisson random variables. Next, we estimate parameters of the underlying Poisson distributions using a maximum likelihood method and derive the statistical properties (consistency, asymptotical efficiency and asymptotic unbiasedness) of this maximum likelihood estimator for grouped and right-censored count data (the MLE-GRC), followed by a discussion on goodness of fit. Our analyses are illustrated by empirical applications to data on adolescent drinking. The next section commences with a brief description of count-data response categories in sample surveys.

WHY GROUPED AND RIGHT-TRUNCATED COUNT DATA?

The GRC data structure is often adopted by social scientists exploring behaviors and attitudes in areas such as public health, crime and delinquency, mental health and urban studies. As examples, respondents in surveys have been asked to list their frequencies of being sick, suicide attempts, criminal victimization, substance use, neighborly interactions and residential moves. There are two reasons why the GRC data structure is employed. First and foremost, for data collection on sensitive topics, such as personal income, number of sex partners, incidents of delinquent behaviors and history of drug use, respondents may feel much more comfortable in reporting grouped categories instead of exact numbers (Groves et al. 2009). Based on this reasoning, two of the major surveys on adolescent behaviors and attitudes in America, the *Monitoring the Future Study* (MTF) and the *National Longitudinal Study of Adolescent to Adult Health* (Add Health) adopted GRC response categories for multiple questions related to juvenile delinquency and drug use. Second, a precise enumeration of number of events imposes a cognitive burden on interviewees and leads to excessive missing data. For example, although medical sociologists and psychiatrists would like to know exactly how many days in the past week individuals experienced a variety of depressive symptoms, respondents, especially those with depressive symptoms, often get frustrated when required to distinguish between, for example, 2 days and 3 days. Thus, the CES-D scale, a standard self-report depression measure, offers four grouped response categories: less than 1 day, 1-2 days, 3-4 days and 5-7 days (Radloff 1977).

As discussed above, an alternative approach is to collect right-censored and grouped count data. For instance, a following question is adopted by the *Monitoring the Future* (MTF) to elicit frequency of binge drinking:

> "Think back over the LAST TWO WEEKS. How many times have you had five or more drinks in a row? (A "drink" is a glass of wine, a bottle of beer, a wine cooler, a shot glass of liquor, a mixed drink, etc.)"

Response categories for this question are none, once, twice, 3 to 5 times, 6 to 9 times, 10 or more times. Such GRC response categories demand much less efforts from both interviewers and interviewees. Yet, no existing statistical models can be readily applied to analyze this distinct data structure unless count data are treated as ordinal measures and then analyzed accordingly using, for example, proportional odds models. The following sections address this by developing a maximum likelihood method for analyzing GRC data and investigating the statistical properties of this new maximum-likelihood estimator for both the Poisson and zero-inflated Poisson cases.

MAXIMUM LIKELIHOOD ESTIMATION FOR GROUPED AND RIGHT-TRUNCATED DATA

The Poisson Case

Statistical models for analyzing count data are often based on the *Poisson distribution*, whose probability mass function for a random event count variable *y* is given as follows (Long and Freese 2006):

$$f(y \mid \lambda) = \mathrm{e}^{-\lambda} \frac{\lambda^{y}}{y!}, \quad y = 0,1,2,\cdots, \tag{1}$$

where λ is both the expected value/mean and the variance of Poisson distribution.

To define a Poisson-based likelihood function for GRC count data, we first denote by $\mathrm{G} = \{I_j\}_{j=1}^{G}$ the division of all nonnegative integers into a grouping and right censored scheme of counts. For identically and independently distributed (iid) observations x_i from a $\mathrm{Poisson}(\lambda)$ distribution, let

$$\alpha_j(x_i) = \begin{cases} 1, & when\ x_i \in I_j, \\ 0, & otherwise \end{cases} \tag{2}$$

so that we have a G-dimensional random vector $(\alpha_1, \cdots, \alpha_G)$ for denoting the GRC responses. For example, $(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6)$ summarizes the response of the aforementioned MTF binge-drinking question if G is expressed as $\{\{0\}, \{1\}, \{2\}, \{3,4,5\}, \{6,7,8,9\}, \{10,...\}\}$. Note that, for any given respondent in a MTF sample, there is only one component of $(\alpha_1, \cdots, \alpha_G)$ that equals 1 and all other α_j s are zero. This vector has a multinomial distribution $\mathrm{M}\,(1, \theta_1, \cdots, \theta_G)$,[1] where the parameter θ_j depends on the λ of the underlying $\mathrm{Poisson}(\lambda)$

[1] The first parameter denotes the number of trials in a multinomial distribution.

$$\theta_j(\lambda) = \sum_{y \in I_j} e^{-\lambda} \frac{\lambda^y}{y!}, \quad y = 0, 1, 2, \cdots, n. \tag{3}$$

If we refer the vector $\boldsymbol{\alpha}(X) = (\alpha_1(X), ..., \alpha_G(X))$ by α, its probability mass function is

$$f(\alpha \mid \lambda) \propto \theta_1^{\alpha_1} \theta_2^{\alpha_2} \cdots \theta_G^{\alpha_G}.$$

Now suppose that we have n observations $\{x_i\}_{i=1}^n$ independently drawn from the Poisson(λ) distribution. The likelihood function for estimating λ is:

$$L(\lambda) \propto \prod_{i=1}^{n} f(\alpha(x_i) \mid \lambda) \tag{4}$$

Based on Lehmann and Casella (1998: 447-449, Theorem 3.7 and Theorem 3.10), we next show that the maximum likelihood estimator, the MLE-GRC, is consistent and asymptotically normally distributed. Theorems 1 to 3 presented below are adapted from Theorem 3.7 and 3.10 in Lehmann and Casella (1998) and we verify *every* condition of Theorem 3.7 and 3.10 to show consistency and asymptotic normality of the MLE-GRC. In order to simplify the proof, we do not verify these conditions in exactly the same order as originally presented in Lehmann and Casella (1998). First, the consistency and asymptotic distribution are given by the theorems below.

Theorem 1. With probability tending to 1 as $n \to \infty$, the likelihood equation with $L(\lambda)$ given in (4)

$$\frac{\mathrm{d}}{\mathrm{d}\lambda} \log L(\lambda) = 0 \tag{5}$$

has a root $\hat{\lambda}_n$ which converges to the true value λ_0 in probability.

Theorem 2. Any consistent sequence $\hat{\lambda}_n$ of roots of the likelihood equation (5) satisfies $\sqrt{n}(\hat{\lambda}_n - \lambda_0) \to \mathrm{N}\left(0, \frac{1}{I(\lambda_0)}\right)$ in distribution.

We assume that the number of groups $G \geq 2$ for all possible grouping schemes so that λ is estimable. Several conditions of $f(\alpha \mid \lambda)$ and $L(\lambda)$ should be investigated to show that MLE-GRC is a consistent estimator.

(A0) The distributions $\{f(\alpha \mid \lambda) : 0 < \lambda < \infty\}$ are distinct. Namely, if

$f(\alpha \mid \lambda_1) = f(\alpha \mid \lambda_2)$ holds for all possible α, we have $\lambda_1 = \lambda_2$. Given that $\sum_{j=1}^{G} \alpha_j = 1$ and a_j ³ 0"j, $f(\alpha \mid \lambda_1) = f(\alpha \mid \lambda_2)$ implies $\theta_j(\lambda_1) = \theta_j(\lambda_2)$ for $j = 1, \cdots, G$. In particular, we have $\theta_1(\lambda_1) = \theta_1(\lambda_2)$. Since $G \geq 2$ and $1 \leq l_2 < \infty$, the following conclusion holds:

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\theta_1(\lambda) = \frac{\mathrm{d}}{\mathrm{d}\lambda}\sum_{y=l_1=0}^{l_2-1} \mathrm{e}^{-\lambda}\frac{\lambda^y}{y!} = -\mathrm{e}^{-\lambda}\frac{\lambda^{l_2-1}}{(l_2-1)!} < 0.$$

Thus, θ_1 is strictly decreasing on $\lambda \in (0,\infty)$, which implies $\lambda_1 = \lambda_2$.

(A1) The distributions $\{f(\alpha \mid \lambda) : 0 < \lambda < \infty\}$ have the same support. For any $0 < \lambda < \infty$ and $j = 1, \cdots, G$, it is easy to show

$$Pr(\alpha_j = 1 \mid \lambda) = \theta_j(\lambda) = \sum_{y \in I_j} \mathrm{e}^{-\lambda}\frac{\lambda^y}{y!} > 0.$$

(A2) The observations $\{\alpha(x_i)\}_{i=1}^{n}$ are iid draws from $f(\alpha \mid \lambda)$. This immediately follows the definition that x_i is an iid draw from the $\mathrm{Poisson}(\lambda)$ distribution.

(A3) The parameter space is an open interval. This assumption is also satisfied because we maximize $L(\lambda)$ on an open interval $(0,\infty)$, though we do not claim that the optimal value of $L(\lambda)$ is always unique.

Given that the MLE-GRC satisfies (A0)-(A3), Theorem 1 holds according to Theorem 3.7 in Lehmann and Casella (1998: 447). Based on Theorem 3.10 in Lehmann and Casella (1998: 449-450), we show that the MLE-GRC also is asymptotically efficient and unbiased by investigating three additional assumptions (A4)-(A6).

(A4) For every α, the probability mass function $f(\alpha \mid \lambda)$ is infinitely differentiable with respect to λ, and there exists some function $D(\alpha)$ such that the third derivative of $f(\alpha \mid \lambda)$ satisfies $\left|\frac{\partial^3 \log f(\alpha \mid \lambda)}{\partial \lambda^3}\right| \leq D(\alpha) < \infty$ for any α and all λ Î $[\frac{1}{2}\lambda_0, \frac{3}{2}\lambda_0]$ (see Appendix for proof).

(A5) The integral $\int f(\alpha \mid \lambda)\mathrm{d}\sigma(\alpha)$ is infinitely differentiable under the integral sign. The proof is trivial because $\mathrm{d}\sigma(\alpha)$ pertains to a discrete measure and the integral is just a finite sum.

(A6) The Fisher information $I(\lambda)\in(0,\infty)$. With regard to $f(\alpha\,|\,\lambda)$, its Fisher information $I(\lambda)$ is a function of λ and also depends on the grouping scheme G (see Appendix for proof):

$$I(\lambda)=I_{\mathrm{G}}(\lambda)=\mathrm{E}\left[\left(\frac{\mathrm{d}}{\mathrm{d}\lambda}\log f(\alpha\,|\,\lambda)\right)^2\right]=\mathrm{E}\left[\left(\sum_{j=1}^{G}\alpha_j\frac{\theta_j'}{\theta_j}\right)^2\right]=\sum_{j=1}^{G}\frac{(\theta_j')^2}{\theta_j}.$$

Given that (A0)-(A6) are satisfied, both the asymptotic efficiency and asymptotic unbiasedness of the MLE-GRC are guaranteed by Theorem 2, which follows Theorem 3.10 in Lehmann and Casella (1998: 449-450).

The Zero-Inflated Poisson Case

λ is both the mean and variance of the Poisson distribution. However, empirical frequency distributions of count data violate this "mean equals variance" assumption if the count data has excess zeros relative to a Poisson distribution (Barron 1992; Hall 2000; Lambert 1992; Zorn 1998). Instead, researchers often use the *zero-inflated Poisson (ZIP) distribution* which, for the count variable *y*, has the probability mass function

$$f(y\,|\,p,\lambda)=\begin{cases}1-p+pe^{-\lambda} & when\;\; y=0\\ pe^{-\lambda}\dfrac{\lambda^y}{y!} & when\;\; y>0\end{cases}\tag{6}$$

where p is the proportion of population exposed to the Poisson distribution and λ is the parameter of a corresponding Poisson distribution.

While the definition of α_j remains the same as (2), the G-dimensional random vector $(\alpha_1,\cdots,\alpha_G)$ defined in the last section now has a multinomial distribution $\mathrm{M}\,(1,\theta_1,\cdots,\theta_G)$, where

$$\theta_1(p,\lambda)=1-p+p\sum_{y\in I_1}\mathrm{e}^{-\lambda}\frac{\lambda^y}{y!}\text{ for } j=1 \text{ and } \theta_i(p,\lambda)=p\sum_{y\in I_j}\mathrm{e}^{-\lambda}\frac{\lambda^y}{y!}\text{ for } j=2,\cdots,G.\tag{7}$$

The probability mass function of α depends on p and λ,

$$f(\alpha \mid p,\lambda) \propto \theta_1^{\alpha_1} \cdots \theta_G^{\alpha_G}. \tag{8}$$

For independent observations $\{x_i\}_{i=1}^n$, the likelihood function is

$$L(p,\lambda) \propto \prod_{i=1}^{n} f(\alpha(x_i) \mid p,\lambda). \tag{9}$$

Based on Theorem 5.1 in Lehmann and Casella (1998: 443-444, 462-467), we show that the MLE-GRC remains consistent, asymptotically efficient and asymptotically unbiased for the ZIP case, whose asymptotic distribution is given by Theorem 3:

Theorem 3. With probability tending to 1 as $n \to \infty$, there exist solutions $(\hat{\lambda}_n, \hat{p}_n)$ of the likelihood equations $\dfrac{\partial L(p,\lambda)}{\partial p} = \dfrac{\partial L(p,\lambda)}{\partial \lambda} = 0$ such that

(1) $\hat{\lambda}_n \to \lambda_0$ and $\hat{p}_n \to p_0$ in probability as $n \to \infty$.

(2) $\sqrt{n}\left[\begin{pmatrix} \hat{p}_n \\ \hat{\lambda}_n \end{pmatrix} - \begin{pmatrix} p_0 \\ \lambda_0 \end{pmatrix}\right]$ is asymptotically normal with mean zero and variance matrix $I(p_0,\lambda_0)^{-1}$;

(3) $\hat{\lambda}_n$ and $\hat{p}_n$ are asymptotically efficient in the sense that in distribution,

$$\sqrt{n}(\hat{p}_n - p_0) \to \mathrm{N}\ (0, [I(p_0,\lambda_0)]_{11}^{-1}),$$
$$\sqrt{n}(\hat{\lambda}_n - \lambda_0) \to \mathrm{N}\ (0, [I(p_0,\lambda_0)]_{22}^{-1}).$$

To prove that the MLE-GRC is a consistent, asymptotically efficient and asymptotically unbiased estimator for the zero-inflated Poisson case, we next state and verify a series of assumptions similar to those of the previous section based on Theorem 5.1 from Lehmann and Casella (1998: 462-463). Now that there are two parameters in the likelihood function, we assume that $G \geq 3$ for all possible grouping schemes so that both p and λ are estimable. The assumptions (B0)-(B2) are virtually the same as assumptions (A0)-(A2).

(B0) The distributions $\{f(\alpha \mid p,\lambda): 0 < p < 1, 0 < \lambda < \infty\}$ are distinct. That is, if $f(\alpha \mid p_1,\lambda_1) = f(\alpha \mid p_2,\lambda_2)$ holds for all possible α, we have $\lambda_1 = \lambda_2$ and $p_1 = p_2$. To prove this, we note that $G \geq 3$ and the condition $f(\alpha \mid p_1, \lambda_1) = f(\alpha \mid p_2, \lambda_2)$ for all α implies $\theta_G(p_1, \lambda_1) = \theta_G(p_2, \lambda_2)$ and $\theta_{G-1}(p_1, \lambda_1) = \theta_{G-1}(p_2, \lambda_2)$. The two equations are expanded as follows.

$$\begin{cases} p_1 \sum_{k=l_{G-1}}^{l_G - 1} \mathrm{e}^{-\lambda_1} \frac{\lambda_1^k}{k!} - p_2 \sum_{s=l_{G-1}}^{l_G - 1} \mathrm{e}^{-\lambda_2} \frac{\lambda_2^s}{s!} = 0, \\ p_1 \sum_{k=l_G}^{\infty} \mathrm{e}^{-\lambda_1} \frac{\lambda_1^k}{k!} - p_2 \sum_{s=l_G}^{\infty} \mathrm{e}^{-\lambda_2} \frac{\lambda_2^s}{s!} = 0. \end{cases} \tag{E1}$$

Their matrix form is

$$A \begin{pmatrix} p_1 \\ -p_2 \end{pmatrix} = 0, \quad \text{where } A = \begin{pmatrix} \sum_{k=l_{G-1}}^{l_G - 1} \mathrm{e}^{-\lambda_1} \frac{\lambda_1^k}{k!} & \sum_{s=l_{G-1}}^{l_G - 1} \mathrm{e}^{-\lambda_2} \frac{\lambda_2^s}{s!} \\ \sum_{k=l_G}^{\infty} \mathrm{e}^{-\lambda_1} \frac{\lambda_1^k}{k!} & \sum_{s=l_G}^{\infty} \mathrm{e}^{-\lambda_2} \frac{\lambda_2^s}{s!} \end{pmatrix}.$$

As we assume $0 < p_1, p_2 < 1$, the equation above holds if and only if $\det A = 0$:

$$\begin{aligned} 0 = \det A &= \sum_{k=l_{G-1}}^{l_G - 1} \sum_{s=l_G}^{\infty} \mathrm{e}^{-\lambda_1 - \lambda_2} \frac{\lambda_1^k \lambda_2^s}{k!s!} - \sum_{k=l_G}^{\infty} \sum_{s=l_{G-1}}^{l_G - 1} \mathrm{e}^{-\lambda_1 - \lambda_2} \frac{\lambda_1^k \lambda_2^s}{k!s!} \\ &= \mathrm{e}^{-\lambda_1 - \lambda_2} \sum_{k=l_{G-1}}^{l_G - 1} \sum_{s=l_G}^{\infty} \frac{(\lambda_1 \lambda_2)^k}{k!s!} (\lambda_2^{s-k} - \lambda_1^{s-k}). \end{aligned} \tag{E2}$$

The last step of (E2) follows Tonelli's Theorem (see e.g., DiBenedetto 2002 for reference). When $\lambda_1 - \lambda_2 \neq 0$, we assume $\lambda_2 > \lambda_1$ without loss of generality. However, this will lead to the conclusion that $\mathrm{e}^{-\lambda_1 - \lambda_2} \sum_{k=l_{G-1}}^{l_G - 1} \sum_{s=l_G}^{\infty} \frac{(\lambda_1 \lambda_2)^k}{k!s!} (\lambda_2^{s-k} - \lambda_1^{s-k}) > 0$, which contradicts $\det A = 0$. Therefore, we have $\lambda_1 = \lambda_2$, which immediately implies $p_1 = p_2$ by any of the two equations in (E1).

(B1) The distributions $f(\alpha \mid p, \lambda)$ have common support. Actually, given that $0 < \lambda < \infty$ and $0 < p < 1$, the support set of $f(\alpha \mid p, \lambda)$ covers all the possible α 's for any $1 \leq j \leq G$.

$$\Pr(\alpha_j = 1 \mid p, \lambda) = \begin{cases} 1 - p + p \sum_{y \in I_1} \mathrm{e}^{-\lambda} \frac{\lambda^y}{y!} > 0, & \text{if j=1}, \\ p \sum_{y \in I_j} \mathrm{e}^{-\lambda} \frac{\lambda^y}{y!} > 0, & \text{if j} \geq 2. \end{cases}$$

(B2) The observations $\{\alpha(x_i)\}_{i=1}^{n}$ are iid draws from $f(\alpha \mid p,\lambda)$. Again, this condition immediately follows the specification that $\{x_i\}_{i=1}^{n}$ is an iid draw from the zero-inflated Poisson distribution.

(B3) The parameter space is open and the density $f(\alpha \mid p,\lambda)$ admits all third derivatives. This assumption is still satisfied as long as we maximize $L(p,\lambda)$ on an open space $(0,1)\times(0,\infty)$. Meanwhile, the fact that θ_j is analytic on p and λ for $j=1,\cdots,G$ implies that $f(\alpha \mid p,\lambda)$ is infinitely differentiable.

(B4) The first and second logarithmic derivatives of $f(\alpha \mid p,\lambda)$ satisfy the following equations (see Appendix for proof).

$$\mathrm{E}\left[\frac{\partial}{\partial p}\log f(\alpha \mid p,\lambda)\right]=\mathrm{E}\left[\frac{\partial}{\partial \lambda}\log f(\alpha \mid p,\lambda)\right]=0$$

$$\mathrm{E}\left[\left(\frac{\partial}{\partial \lambda}\log f(\alpha \mid p,\lambda)\right)^2\right]=-\mathrm{E}\left[\frac{\partial^2}{\partial \lambda^2}\log f(\alpha \mid p,\lambda)\right], \tag{E3}$$

$$\mathrm{E}\left[\frac{\partial}{\partial \lambda}\log f(\alpha \mid p\lambda)\frac{\partial}{\partial p}\log f(\alpha \mid p,\lambda)\right]=-\mathrm{E}\left[\frac{\partial^2}{\partial \lambda \partial p}\log f(\alpha \mid p,\lambda)\right], \tag{E4}$$

$$\mathrm{E}\left[\left(\frac{\partial}{\partial p}\log f(\alpha \mid p,\lambda)\right)^2\right]=-\mathrm{E}\left[\frac{\partial^2}{\partial p^2}\log f(\alpha \mid p,\lambda)\right]. \tag{E5}$$

(B5) The Fisher information matrix $I(p,\lambda)$ is finite (all four elements of the matrix are finite) and positive definite (see Appendix for proof). The Fisher information matrix is

$$I(p,\lambda)=\begin{pmatrix} I_{11}(p,\lambda) & I_{12}(p,\lambda) \\ I_{21}(p,\lambda) & I_{22}(p,\lambda) \end{pmatrix},$$

where

$$I_{11}(p,\lambda)=\mathrm{Var}[\frac{\partial}{\partial \lambda}\log \ f(\alpha \mid p,\lambda)],$$

$$I_{12}(p,\lambda)=I_{21}(p,\lambda)=\mathrm{Cov}\,[\frac{\partial}{\partial \lambda}\log \ f(\alpha \mid p,\lambda),\ \frac{\partial}{\partial p}\log \ f(\alpha \mid p,\lambda)],$$

$$I_{22}(p,\lambda)=\mathrm{Var}[\frac{\partial}{\partial p}\log\ f(\alpha\mid p,\lambda)].$$

Or, after some derivation, the Fisher information matrix is

$$I(p,\lambda)=\begin{pmatrix}\sum_j \frac{1}{\theta_j}\left(\frac{\partial\theta_j}{\partial\lambda}\right)^2 & \sum_j \frac{1}{\theta_j}\frac{\partial\theta_j}{\partial\lambda}\frac{\partial\theta_j}{\partial p} \\ \sum_j \frac{1}{\theta_j}\frac{\partial\theta_j}{\partial\lambda}\frac{\partial\theta_j}{\partial p} & \sum_j \frac{1}{\theta_j}\left(\frac{\partial\theta_j}{\partial p}\right)^2\end{pmatrix}. \tag{E6}$$

(B6) For any $p_0\in(0,1)$ and $\lambda_0\in(0,\infty)$, there exists a neighborhood where the absolute values of all the partial derivatives $\frac{\partial^3}{\partial p^3}\log f$, $\frac{\partial^3}{\partial p^2\partial\lambda}\log f$, $\frac{\partial^3}{\partial p\partial\lambda^2}\log f$ and $\frac{\partial^3}{\partial\lambda^3}\log f$ are uniformly bounded away from ∞. Similar to our proof of (A4) (see Appendix), this conclusion follows the fact that $f(\alpha\mid p,\lambda)$ is positive for all α, $p_0\in(0,1)$ and $\lambda_0\in(0,\infty)$, as already given in (B1).

With assumptions (B0)-(B6) satisfied, Theorem 3 holds based on Theorem 5.1 from Lehmann and Casella (1998: 462-463).

GOODNESS-OF-FIT TEST

While other test statistics (such as the log-likelihood ratio test, AIC and BIC) can also be applied to justify model selection (Burnham and Anderson 2004; Raftery 1995; Wong 1994), this section applies χ^2 goodness-of-fit tests to the MLE-GRC. As the χ^2 test for the Poisson case is straightforward, we only illustrate such tests for the ZIP case. Suppose that $(\hat{p}_n,\hat{\lambda}_n)$ is the MLE-GRC estimator pertaining to the likelihood function in (9), we define the test statistics for the ZIP case as

$$Q_{\mathrm{ZIP}}=\sum_{j=1}^{G}\frac{(\sum_{i=1}^{n}\alpha_j(x_i)-n\theta_j(\hat{p}_n,\hat{\lambda}_n))^2}{n\theta_j(\hat{p}_n,\hat{\lambda}_n)}, \tag{10}$$

where α_j and θ_j are defined by (2) and (7), respectively and n is the sample size. Now we test the null hypothesis H_0 that the observed vectors $(\sum_{i=1}^{n}\alpha_1(x_i),\cdots,\sum_{i=1}^{n}\alpha_G(x_i))$ are generated by the $M(n,\theta_1(p,\lambda),\cdots,\theta_G(p,\lambda))$ distribution for some p and λ against the alternative hypothesis H_1 that H_0 is incorrect. If we want to test the null hypothesis at a 95% level of confidence, the H_0 would be rejected when $Q_{ZIP} > F^{-1}_{\chi^2_{G-3}}(0.95)$. Note that the degrees of freedom of the χ^2 distribution is $G-3$ because the ZIP case has two parameters.

DATA SIMULATION AND EMPIRICAL ANALYSES

To test the validity of the MLE-GRC, we first constructed a series of hypothetical Poisson/ZIP distributions, grouped them according to a specific grouping scheme, fitted the GRC count data using the MLE-GRC and then compared different MLE-GRC estimators with true parameters chosen to generate the Poisson/ZIP distributions. First, we chose grouping schemes adopted by the binge drinking question and other alcohol drinking questions, and then constructed Poisson/ZIP distributions to test the MLE-GRC using different sample sizes, which are shown in Table 1.1 (its grouping scheme I is from the binge drinking question) and Table 1.2 (its grouping scheme II is from the other three alcohol drinking questions). The first two columns are the numerical values of the parameters we used to define the Poisson/ZIP distributions. As expected, the MLE-GRC estimators $\hat{\lambda}_{Poisson}$, $\hat{\lambda}_{ZIP}$ and $\hat{P}_{ZIP}$ largely converge to their true values as sample sizes increase. Given that the null hypothesis of a goodness-of-fit test is that an inferred model fits the data well, the test statistics tend to reject false null hypotheses as sample sizes increase. For Scheme I, ZIP distributions are constructed when $P=0.6$ and λ equals to 0.5. The false null hypotheses of Poisson distributions are not rejected with a smaller sample size (N=500) but successfully rejected with a larger sample size (N=5000).

The MLE-GRC of either the Poisson case or the ZIP case may provide a relatively poor fit to the count data if λ, the mean of the frequency distribution, is very small. When λ equals to 0.5, both $\hat{\lambda}_{Poisson}$ and $\hat{\lambda}_{ZIP}$, especially the latter, can deviate considerably from 0.5. One reason is that the grouping of very rare count events will lead to right-skewed Poisson distribution (e.g., most individuals reporting 3-5 times of binge drinking probably refer to 3 times instead of 4 or 5 times). For example, the grouping scheme I ("1" and "2" are included in separate groups) adopted by Table 1.1 appears to be a better choice than the grouping scheme II adopted by Table 1.2 ("1" and "2" are combined into one group) when rare events are analyzed. For $\lambda = 0.5$ in Table 1.2, $\hat{\lambda}$ tends to be overestimated due to this grouping effect. For rare count events (e.g., $\lambda \leq 1$), Table 1.1 and 1.2 also show that test statistics can reject true null hypotheses of ZIP distributions when sample sizes are large, or fail to reject false null hypotheses of Poisson distributions when sample sizes are small. These findings have several implications. First, it could be inappropriate to measure very rare count events by GRC responses, especially when the sample size is small. Second, if count data for rare events were collected using GRC responses, scholars should be cautious in using Poisson models and consider other alternative ways (such as logistic regression) for modeling such rare events. Third, although the grouping effect tends to exacerbate the differences between hypothetical Poisson/ZIP parameters used for simulation and MLE-GRC estimators, such bias can be greatly reduced if sample sizes are large or an appropriate grouping scheme is chosen.

[Table 1.1 and Table 1.2 about here]

Table 1 shows that MLE-GRC estimates tend to be close to true parameters. As results based on one simulation are subject to stochasticity, we further investigate the performance of MLE-GRC under different scenarios through repeated simulations.[2] All results reported in Table 2 are produced based on 1000 simulations. The simulated data were generated by Poisson and ZIP models separately and the corresponding models were fitted. Under a specific combination of sample size, grouping scheme and (p, λ), the average of estimates across simulations, their standard deviation and the coverage probability of 95% confidence intervals (the percentage that a 95% confidence interval contains the true parameter across the 1000 simulations) are reported. Compared to results based on the single simulation in Table 1, the average estimate is in general closer to its corresponding true parameter. The standard deviations of estimates are very small under different

[2] Due to space limits, we only present results when λ is 0.5, 3 and 10, respectively.

scenarios and tend to be even smaller with larger sample sizes, suggesting the consistency of estimation. $\hat{\lambda}_{ZIP}$ is often associated with a higher standard deviation than that of a corresponding $\hat{\lambda}_{Poisson}$ in a scenario where the true λ is small but such differences in standard deviations tend to diminish or even reverse when the true λ is larger. For $p = 0.6$ in Scheme I ($N = 100$), the standard deviation of $\hat{\lambda}_{Poisson}$ (0.062) is smaller than that of $\hat{\lambda}_{ZIP}$ (0.186) when $\lambda = 0.5$; yet the standard deviation of $\hat{\lambda}_{Poisson}$ (0.526) is larger than that of $\hat{\lambda}_{ZIP}$ (0.475) when $\lambda = 10$. In Table 2, results in bold are scenarios where Poisson and zero-inflated Poisson are nearly unidentifiable ($P = 1.0$). For these scenarios, $\hat{\lambda}_{ZIP}$ is also associated with a higher standard deviation than that of a corresponding $\hat{\lambda}_{Poisson}$: when the true λ is small (e.g., $\lambda = 0.5$), the use of ZIP models appears to introduce more variance in estimation. Yet, it appears that the choice of ZIP or Poisson models does not matter for the estimation of a large λ : both ways of estimation of data without zero inflation ($P = 1.0$) yield approximately the same $\hat{\lambda}_{Poisson}$, $\hat{\lambda}_{ZIP}$ and their corresponding standard deviations when λ is sufficiently large (e.g., $\lambda = 10$).

$\hat{P}_{ZIP}$ is associated with a larger standard deviation if the sample size or λ is small, while the standard deviations of $\hat{P}_{ZIP}$ do not vary systematically with different values of P .When the Poisson model for count data with or without zero inflation is correctly specified, the coverage probability deviates little from the 95% confidence level. Yet, we also notice that the coverage probability is sensitive to the value of λ and can deviate substantially from 0.95 if the true parameter λ is small (e.g., $\lambda = 0.5$). This finding again suggests that Poisson-based models do not fit rare count data well and analysts should consider alternative models.

[Table 2 about here]

As an illustration of this analytic strategy, we analyze alcohol drinking data from a nationally representative annual survey of youth in the United States, the *Monitoring the Future: A Continuing Study of the Lifestyle and Values of Youth* (MTF) study (http://monitoringthefuture.org/). Every year thousands of students from approximately 130 high schools nationwide participate in this survey and respond to a series of questions on values, behaviors, and characteristics of American adolescents. The MTF was initiated in 1975 with an annual survey of 12th graders in the United States. Instructed by MTF research staff, students participate in this annual survey by completing self-administered and machine-readable questionnaires at school. While we have previously discussed the binge drinking question, the other three questions are virtually the same except for the reference period.

"On how many occasions have you had alcoholic beverages to drink--more than just a few sips . . .

A: . . . in your lifetime? B: . . . during the last 12 months? C: . . . during the last 30 days?"

The GRC response categories for the three questions include the following: 0 Occasions, 1-2 Occasions, 3-5 Occasions 6-9 Occasions, 10-19 Occasions, 20-39 Occasions and 40 or More. This research considers MTF survey data from 1996 to 2012. The original counts of all four alcohol drinking questions were retrieved from data-description files prepared by the MTF survey team. A simple calculation based on Table 3 would show that the prevalence rates of lifetime drinking are higher than those of drinking in 12 months, while both are higher than those of drinking in the last 30 days. Among all four questions, binge drinking is most rare for 12th graders. Table 3 also suggests that there was an overall decline in estimated adolescent alcoholic consumption. Figure 1 shows these values of λ (left y-axis) and P (right y-axis) estimated by an application of the MLE-GRC method to grouped and censored counts of drinking.

We calculated these MLE-GRC estimates by writing an R function (grcmle) with following parameters:

grcmle (count, scheme, method)

where frequency distributions and the grouping scheme used are given by *count* and *scheme*, respectively. *Scheme* consists of an array represents the lowest integer included in each group. *Method* is either "Poisson" or "ZIP". For example, the λ of lifetime drinking (the Poisson case) in the 2012 wave is estimated by:

```
grcmle(counts = c(674, 199, 235, 208, 281, 224, 416), scheme = c(0, 1, 3, 6, 10, 20, 40), method = "Poisson").
```

Although this R function is not specifically designed for trend analysis and the estimates in Figure 1 are obtained by fitting each year's data separately, it can be seen that these parameter estimates are consistent with the descriptive patterns revealed in Table 1 and supportive of an inference of a decline in adolescent drinking from the late 1990s to recent years (Newes-Adeyi et al. 2005; Wallace Jr et al. 2002). Across all four panels in Figure 1, all λ s estimated in the year 2012 are lower than their counterparts in the year 1996. Meanwhile, λ s estimated using the ZIP distribution share the same pattern as their counterparts estimated using the Poisson distribution, although levels of the former are higher than those of the latter because, for the ZIP case, drinking behaviors as described by the Poisson process can only be attributable to individuals who are not non-drinkers. In general, both λ and P of four types of drinking behaviors decrease from 1996 to 2012 and all MLE-GRC estimates are significant at the 95% confidence level. As expected, results based on different statistics (AIC, BIC and log-likelihood statistics) show that ZIP models fit drinking data better than Poisson models do (results available upon request).

[Table 3 and Figure 1 about here]

DISCUSSION AND CONCLUSION

To estimate λ from grouped and right-censored count data, this paper proposes a data generating process for conceptualizing GRC count data, derives statistical properties of a maximum likelihood estimator, the MLE-GRC, for both Poisson and zero-inflated Poisson cases. Based on results from data simulation and empirical analyses, we discuss the application of MLE-GRC in (rare) count events and develop an R function.

By grouping and censoring count data, the GRC count data structure has properties of both categorical data and count data. If the number of groups is greatly expanded and each group of a grouping scheme is allowed to include only a single integer, the likelihood functions introduced in this paper would reduce to existing likelihood functions for Poisson and zero-inflated Poisson models. Therefore, the statistical models presented here could be regarded as a more general case for analyzing count data.[3]

[3] The Poisson-multinomial mixture model method presented here can be viewed as a special case of modeling count data that have been coarsened at random, where the coarsening mechanism is the same for each observation and determined by researchers, see Heitjan, Daniel F and Donald B Rubin. 1991. "Ignorability and Coarse Data." *The Annals of Statistics* 19(4): 2244-53. This mixture model method can

There are explicit likelihood functions available for count data without both grouping and right censoring, whose statistical properties have been investigated. In the presence of grouping and right censoring, however, additional complexities arise. By proposing a new Poisson-multinomial approach, this paper demonstrates that MLE-GRC estimators are consistent and have asymptotic normal distributions for Poisson and ZIP cases (Theorems 1 to 3). The new R function grcmle (count, scheme, method) developed in the present study are thus useful for epidemiologists and social scientists to monitor changes of a grouped and right-censored count outcomes. These asymptotic properties of the MLE-GRC also lay a foundation for subsequent multivariate regression analyses of grouped and right censored data, which can be implemented, for example, by an iterative procedure (McCullagh and Nelder 1989: 40-43). Moreover, the MLE-GRC method presented here motivates research on the choice of grouping schemes: given prior distributions of λ and P , an optimal grouping scheme should be the one that minimizes the variance (or maximizes the Fisher information) of the asymptotic distribution of the MLE-GRC. Drawing upon this observation, survey investigators may design optimal grouping schemes for GRC count responses.

also be implemented using a weighted EM algorithm, see Ibrahim, Joseph G, Ming-Hui Chen, Stuart R Lipsitz and Amy H Herring. 2005. "Missing-Data Methods for Generalized Linear Models: A Comparative Review." *Journal of the American Statistical Association* 100(469): 332-46. We are grateful to an anonymous reviewer for highlighting these points.

TABLES AND FIGURES

Table 1.1 Simulated results for MLE-GRC using hypothetical Poisson and Zero-inflated Poisson distributions: Grouping scheme I*

		N=500					N=1000					N=5000					N=10000				
λ	P	$\hat{\lambda}_{Poisson}$	P-Value	$\hat{\lambda}_{ZIP}$	$\hat{P}_{ZIP}$	P-Value	$\hat{\lambda}_{Poisson}$	P-Value	$\hat{\lambda}_{ZIP}$	$\hat{P}_{ZIP}$	P-Value	$\hat{\lambda}_{Poisson}$	P-Value	$\hat{\lambda}_{ZIP}$	$\hat{P}_{ZIP}$	P-Value	$\hat{\lambda}_{Poisson}$	P-Value	$\hat{\lambda}_{ZIP}$	$\hat{P}_{ZIP}$	P-Value
	0.2	0.140	0.139	0.503	0.278	0.852	0.090	0.000	0.405	0.222	0.670	0.114	0.000	0.503	0.228	0.967	0.103	0.000	0.587	0.176	0.998
	0.4	0.240	0.061	0.606	0.396	0.558	0.200	0.001	0.472	0.425	0.992	0.221	0.000	0.602	0.369	0.991	0.185	0.000	0.493	0.377	0.994
0.5	0.6	0.350	0.592	0.465	0.753	0.651	0.285	0.191	0.428	0.666	0.937	0.306	0.001	0.501	0.612	0.959	0.309	0.000	0.510	0.607	0.511
	0.8	0.554	0.902	0.584	0.950	0.815	0.407	0.074	0.592	0.689	0.984	0.443	0.489	0.513	0.864	0.853	0.392	0.000	0.518	0.758	0.448
	1.0	**0.500**	**0.531**	**0.518**	**0.966**	**0.387**	**0.473**	**0.804**	**0.477**	**0.991**	**0.658**	**0.496**	**0.999**	**0.501**	**0.990**	**0.995**	**0.500**	**0.476**	**0.514**	**0.973**	**0.411**
	0.2	0.160	0.000	0.835	0.194	1.000	0.221	0.000	1.126	0.201	0.983	0.177	0.000	1.158	0.156	0.599	0.203	0.000	1.071	0.193	0.695
	0.4	0.426	0.000	1.243	0.351	0.975	0.398	0.000	1.060	0.383	0.707	0.354	0.000	0.952	0.376	0.710	0.395	0.000	0.978	0.409	0.033
1	0.6	0.505	0.064	0.945	0.540	0.997	0.558	0.000	1.014	0.556	0.234	0.655	0.000	1.126	0.589	0.540	0.574	0.000	0.950	0.609	0.687
	0.8	0.840	0.265	0.923	0.913	0.209	0.832	0.069	0.970	0.860	0.349	0.809	0.001	0.994	0.818	0.912	0.776	0.000	0.949	0.821	0.766
	1.0	**1.195**	**0.876**	**1.195**	**1.000**	**0.749**	**0.975**	**0.285**	**0.994**	**0.981**	**0.182**	**1.016**	**0.962**	**1.027**	**0.989**	**0.916**	**0.962**	**0.331**	**0.969**	**0.993**	**0.223**
	0.2	0.472	0.000	4.264	0.132	0.769	0.534	0.000	3.063	0.195	0.432	0.521	0.000	2.962	0.200	0.194	0.528	0.000	2.972	0.200	0.001
	0.4	1.003	0.000	2.873	0.382	0.967	1.076	0.000	2.986	0.396	0.467	1.026	0.000	3.021	0.373	0.478	1.074	0.000	2.960	0.397	0.199
3	0.6	1.329	0.000	2.744	0.513	0.496	1.598	0.000	2.949	0.576	0.783	1.653	0.000	2.966	0.591	0.727	1.754	0.000	3.024	0.613	0.554
	0.8	2.641	0.002	3.042	0.882	0.504	2.438	0.000	3.279	0.769	0.798	2.392	0.000	2.976	0.823	0.668	2.360	0.000	3.067	0.792	0.804
	1.0	**2.909**	**0.639**	**2.976**	**0.980**	**0.566**	**3.013**	**0.598**	**3.017**	**0.999**	**0.431**	**3.064**	**0.216**	**3.064**	**1.000**	**0.123**	**3.004**	**0.579**	**3.004**	**1.000**	**0.411**
	0.2	0.647	0.000	4.598	0.162	0.131	0.880	0.000	4.914	0.212	0.168	0.877	0.000	5.150	0.199	0.873	0.811	0.000	4.948	0.192	0.523
	0.4	1.806	0.000	4.896	0.413	0.973	1.714	0.000	4.860	0.395	0.064	1.932	0.000	5.278	0.410	0.711	1.792	0.000	5.045	0.399	0.243
5	0.6	3.006	0.000	5.369	0.603	0.729	2.734	0.000	4.877	0.603	0.862	2.794	0.000	5.083	0.593	0.787	2.696	0.000	4.937	0.589	0.743
	0.8	3.961	0.000	4.803	0.847	0.832	3.707	0.000	4.914	0.784	0.684	3.859	0.000	5.008	0.798	0.647	3.867	0.000	4.954	0.807	0.371
	1.0	**5.326**	**0.909**	**5.359**	**0.995**	**0.931**	**5.090**	**0.469**	**5.101**	**0.998**	**0.354**	**4.972**	**0.772**	**4.990**	**0.997**	**0.944**	**4.951**	**0.169**	**4.953**	**1.000**	**0.095**
	0.2	2.098	0.000	9.219	0.270	0.978	1.921	0.000	9.955	0.234	0.740	1.755	0.000	9.987	0.214	0.894	1.563	0.000	9.918	0.192	0.620
	0.4	3.915	0.000	10.561	0.440	0.823	3.401	0.000	9.809	0.406	0.629	3.151	0.000	10.124	0.370	0.118	3.452	0.000	10.057	0.405	0.732
10	0.6	5.432	0.000	9.808	0.620	0.806	5.188	0.000	9.686	0.600	0.134	5.206	0.000	9.830	0.596	0.289	5.303	0.000	9.935	0.602	0.342
	0.8	7.088	0.000	9.584	0.790	0.452	7.630	0.000	10.177	0.806	0.400	7.414	0.000	9.999	0.796	0.813	7.447	0.000	10.122	0.793	0.832
	1.0	**9.468**	**0.084**	**9.468**	**1.000**	**0.042**	**10.255**	**0.270**	**10.255**	**1.000**	**0.159**	**10.032**	**0.865**	**10.032**	**1.000**	**0.734**	**10.023**	**0.009**	**10.023**	**1.000**	**0.004**

*Note: 1. the grouping scheme I used for Table 1.1 is [0, 1, 2, 3-5, 6-9, 10+].

2. P-values reported in this table are based on goodness-of-fit tests. A lower p-value (e.g., 0.05) means that a Poisson/ZIP model with selected parameters does not fit the simulated data well.

Table 1.2 Simulated results for MLE-GRC using hypothetical Poisson and Zero-inflated Poisson distributions: Grouping scheme II*

		N=500					N=1000					N=5000					N=10000				
λ	P	$\hat{\lambda}_{Poisson}$	P-Value	$\hat{\lambda}_{ZIP}$	$\hat{P}_{ZIP}$	P-Value	$\hat{\lambda}_{Poisson}$	P-Value	$\hat{\lambda}_{ZIP}$	$\hat{P}_{ZIP}$	P-Value	$\hat{\lambda}_{Poisson}$	P-Value	$\hat{\lambda}_{ZIP}$	$\hat{P}_{ZIP}$	P-Value	$\hat{\lambda}_{Poissor}$	P-Value	$\hat{\lambda}_{ZIP}$	$\hat{P}_{ZIP}$	P-Value
	0.2	0.051	1.000	0.236	0.233	1.000	0.074	0.000	0.561	0.152	0.992	0.072	0.000	0.521	0.160	0.997	0.082	0.000	0.278	0.320	0.339
	0.4	0.173	1.000	0.242	0.731	0.997	0.180	0.000	0.699	0.294	1.000	0.219	0.001	0.563	0.427	0.985	0.188	0.000	0.558	0.374	0.841
0.5	0.6	0.306	0.995	0.355	0.874	0.965	0.269	0.804	0.571	0.506	0.838	0.311	0.950	0.558	0.588	0.409	0.281	0.000	0.559	0.539	0.374
	0.8	0.375	0.671	0.704	0.574	1.000	0.345	0.999	0.338	0.996	0.990	0.372	1.000	0.576	0.703	0.365	0.390	0.763	0.576	0.716	0.008
	1.0	**0.508**	**0.999**	**0.704**	**0.748**	**0.993**	**0.495**	**0.999**	**0.576**	**0.843**	**0.615**	**0.523**	**1.000**	**0.637**	**0.840**	**0.682**	**0.503**	**0.021**	**0.613**	**0.831**	**0.022**
	0.2	0.176	0.000	2.141	0.102	0.988	0.172	0.000	0.889	0.228	1.000	0.179	0.000	0.867	0.241	0.999	0.172	0.000	1.100	0.187	0.270
	0.4	0.462	0.596	0.790	0.627	1.000	0.353	0.000	0.778	0.498	1.000	0.354	0.000	0.920	0.432	0.996	0.345	0.000	0.905	0.441	0.127
1	0.6	0.693	0.999	0.748	0.936	1.000	0.549	0.000	1.070	0.563	0.508	0.559	0.000	1.084	0.547	0.377	0.550	0.000	1.072	0.547	0.003
	0.8	0.789	0.805	1.066	0.770	0.999	0.715	0.003	1.080	0.695	0.967	0.719	0.008	0.905	0.836	0.879	0.755	0.000	1.088	0.716	0.000
	1.0	**0.932**	**0.993**	**1.055**	**0.898**	**0.999**	**0.999**	**0.869**	**0.999**	**1.000**	**0.763**	**0.976**	**0.187**	**1.082**	**0.915**	**0.766**	**1.024**	**1.000**	**1.055**	**0.975**	**0.798**
	0.2	0.719	0.000	3.426	0.248	0.997	0.563	0.000	3.267	0.206	0.461	0.534	0.000	3.077	0.208	0.542	0.498	0.000	2.941	0.202	0.908
	0.4	1.037	0.000	3.392	0.352	0.990	1.105	0.000	2.891	0.430	0.929	1.064	0.000	2.905	0.415	0.833	1.018	0.000	2.981	0.389	1.000
3	0.6	1.595	0.000	2.596	0.659	0.211	1.619	0.000	3.187	0.555	0.853	1.529	0.000	2.846	0.584	0.744	1.659	0.000	2.981	0.602	0.994
	0.8	2.253	0.000	3.078	0.765	0.975	2.347	0.000	3.089	0.790	0.757	2.295	0.000	2.917	0.814	0.997	2.321	0.000	2.991	0.804	0.648
	1.0	**3.083**	**0.883**	**3.103**	**0.995**	**0.793**	**2.948**	**0.955**	**2.973**	**0.993**	**0.954**	**3.084**	**0.993**	**3.084**	**1.000**	**0.976**	**2.974**	**0.917**	**2.974**	**1.000**	**0.833**
	0.2	1.247	0.000	4.765	0.303	0.760	0.853	0.000	5.009	0.201	0.514	0.833	0.000	5.115	0.193	1.000	0.888	0.000	5.168	0.204	0.886
	0.4	1.677	0.000	4.975	0.383	0.864	1.977	0.000	4.996	0.443	0.999	1.748	0.000	5.170	0.383	0.599	1.705	0.000	5.032	0.385	0.842
5	0.6	2.658	0.000	5.591	0.522	0.826	2.752	0.000	4.837	0.613	0.982	2.803	0.000	4.944	0.611	0.468	2.784	0.000	5.025	0.599	0.717
	0.8	3.900	0.000	4.798	0.837	0.886	3.889	0.000	5.033	0.801	0.953	3.798	0.000	5.015	0.787	0.994	3.865	0.000	5.038	0.796	0.922
	1.0	**4.620**	**0.501**	**4.620**	**1.000**	**0.361**	**5.065**	**0.904**	**5.088**	**0.996**	**0.978**	**5.045**	**0.979**	**5.048**	**0.999**	**0.950**	**5.007**	**0.225**	**5.007**	**1.000**	**0.139**
	0.2	1.019	0.000	9.557	0.130	0.840	1.555	0.000	10.516	0.182	0.823	1.687	0.000	9.863	0.207	0.840	1.671	0.000	9.883	0.205	0.427
	0.4	4.085	0.000	10.089	0.470	0.908	2.973	0.000	10.151	0.348	0.963	3.554	0.000	9.894	0.419	0.770	3.455	0.000	9.924	0.407	0.650
10	0.6	5.753	0.000	9.808	0.650	0.980	5.264	0.000	10.341	0.580	0.590	5.180	0.000	10.182	0.578	0.246	5.381	0.000	9.981	0.606	0.008
	0.8	7.888	0.000	10.048	0.830	0.590	7.715	0.000	10.136	0.812	0.105	7.580	0.000	10.020	0.807	0.935	7.481	0.000	9.952	0.802	0.825
	1.0	**9.898**	**0.966**	**9.898**	**1.000**	**0.917**	**9.968**	**0.209**	**9.968**	**1.000**	**0.128**	**10.019**	**0.997**	**10.019**	**1.000**	**0.988**	**9.950**	**0.739**	**9.950**	**1.000**	**0.601**

*Note: 1. the grouping scheme II used for Table 1.2 is [0, 1-2, 3-5, 6-9, 10-19, 20-39, 40+].

2. P-values reported in this table are based on goodness-of-fit tests. A lower p-value (e.g., 0.05) means that a Poisson/ZIP model with selected parameters does not fit the simulated data well.

Table 2 Results from repeated simulations for MLE-GRC using hypothetical Poisson and Zero-inflated Poisson distributions*

			[illegible]									N=500								
Scheme	λ	P	$\hat{\lambda}_{Poisson}$	Std. Dev.	Coverage	$\hat{\lambda}_{ZIP}$	Std. Dev.	Coverage	$\hat{P}_{ZIP}$	Std. Dev.	Coverage	$\hat{\lambda}_{Poisson}$	Std. Dev.	Coverage	$\hat{\lambda}_{ZIP}$	Std. Dev.	Coverage	$\hat{P}_{ZIP}$	Std. Dev.	Coverage
		0.2	0.100	0.037	0%	0.495	0.338	65%	0.323	0.234	72%	0.100	0.015	0%	0.497	0.142	75%	0.218	0.070	74%
		0.4	0.197	0.048	0%	0.497	0.232	68%	0.479	0.216	72%	0.200	0.023	0%	0.496	0.100	71%	0.418	0.081	73%
	0.5	0.6	0.300	0.062	11%	0.511	0.186	70%	0.647	0.202	70%	0.300	0.027	0%	0.500	0.088	67%	0.617	0.102	68%
		0.8	0.400	0.065	64%	0.519	0.142	79%	0.808	0.166	85%	0.399	0.030	8%	0.501	0.070	70%	0.808	0.102	68%
		1.0	**0.501**	**0.073**	**94%**	**0.549**	**0.107**	**87%**	**0.927**	**0.109**	**85%**	**0.499**	**0.033**	**93%**	**0.522**	**0.047**	**84%**	**0.960**	**0.057**	**83%**
		0.2	0.528	0.135	0%	2.996	0.475	96%	0.199	0.042	93%	0.529	0.057	0%	2.997	0.208	94%	0.199	0.018	94%
		0.4	1.097	0.172	0%	2.990	0.323	95%	0.402	0.053	94%	1.091	0.076	0%	2.990	0.141	94%	0.399	0.022	96%
I	**3**	0.6	1.704	0.020	0%	3.001	0.263	97%	0.602	0.054	94%	1.698	0.088	0%	3.003	0.117	94%	0.599	0.023	96%
		0.8	2.350	0.204	7%	3.011	0.235	94%	0.803	0.047	94%	2.335	0.092	0%	2.998	0.103	94%	0.801	0.021	94%
		1.0	**2.996**	**0.184**	**94%**	**3.028**	**0.192**	**94%**	**0.991**	**0.015**	**99%**	**2.999**	**0.081**	**95%**	**3.015**	**0.083**	**95%**	**0.995**	**0.007**	**98%**
		0.2	1.638	0.348	0%	10.076	0.902	96%	0.200	0.040	93%	1.638	0.158	0%	10.028	0.380	96%	0.200	0.018	95%
		0.4	3.382	0.466	0%	10.017	0.598	95%	0.398	0.047	96%	3.404	0.203	0%	10.019	0.281	93%	0.401	0.021	96%
	10	0.6	5.320	0.526	0%	10.002	0.475	97%	0.601	0.048	96%	5.321	0.235	0%	9.995	0.225	94%	0.601	0.022	94%
		0.8	7.469	0.533	0%	9.989	0.433	94%	0.801	0.040	93%	7.457	0.234	0%	10.008	0.185	95%	0.799	0.018	96%
		1.0	**10.028**	**0.392**	**94%**	**10.029**	**0.392**	**94%**	**1.000**	**0.001**	**100%**	**10.008**	**0.175**	**93%**	**10.009**	**0.175**	**93%**	**1.000**	**0.000**	**100%**
		0.2	0.088	0.033	0%	0.442	0.386	24%	0.301	0.139	80%	0.087	0.015	0%	0.500	0.205	58%	0.233	0.097	52%
		0.4	0.183	0.048	0%	0.485	0.290	44%	0.516	0.218	66%	0.182	0.021	0%	0.502	0.173	52%	0.453	0.179	44%
	0.5	0.6	0.282	0.058	6%	0.527	0.224	59%	0.640	0.197	77%	0.283	0.025	0%	0.540	0.124	65%	0.593	0.151	48%
		0.8	0.387	0.068	55%	0.590	0.154	74%	0.712	0.135	79%	0.390	0.030	5%	0.583	0.053	75%	0.707	0.061	51%
		1.0	**0.502**	**0.079**	**94%**	**0.653**	**0.101**	**77%**	**0.795**	**0.095**	**49%**	**0.502**	**0.033**	**96%**	**0.625**	**0.044**	**26%**	**0.820**	**0.043**	**1%**
		0.2	0.507	0.131	0%	2.983	0.477	94%	0.204	0.043	93%	0.499	0.056	0%	2.990	0.206	95%	0.200	0.018	95%
		0.4	1.068	0.175	0%	3.010	0.316	96%	0.403	0.052	94%	1.057	0.080	0%	2.997	0.150	94%	0.400	0.023	94%
II	**3**	0.6	1.663	0.197	0%	2.996	0.257	95%	0.601	0.053	95%	1.661	0.091	0%	3.001	0.125	94%	0.599	0.023	95%
		0.8	2.325	0.214	7%	3.016	0.240	93%	0.800	0.046	94%	2.318	0.093	0%	3.001	0.096	97%	0.801	0.020	94%
		1.0	**2.999**	**0.189**	**95%**	**3.035**	**0.195**	**94%**	**0.990**	**0.015**	**99%**	**2.999**	**0.083**	**96%**	**3.017**	**0.086**	**94%**	**0.995**	**0.007**	**97%**
		0.2	1.651	0.363	0%	10.046	0.875	95%	0.200	0.040	92%	1.643	0.166	0%	10.009	0.391	95%	0.200	0.018	95%
		0.4	3.460	0.469	0%	10.038	0.587	96%	0.405	0.048	95%	3.401	0.218	0%	10.006	0.273	95%	0.399	0.022	95%
	10	0.6	5.322	0.533	0%	10.014	0.491	96%	0.598	0.048	95%	5.345	0.244	0%	10.003	0.207	97%	0.601	0.022	94%
		0.8	7.467	0.526	0%	9.989	0.409	95%	0.799	0.040	93%	7.475	0.242	0%	10.003	0.191	94%	0.799	0.018	94%
		1.0	**10.016**	**0.366**	**95%**	**10.017**	**0.366**	**95%**	**1.000**	**0.001**	**100%**	**9.998**	**0.172**	**94%**	**9.999**	**0.173**	**94%**	**1.000**	**0.000**	**100%**

* Note: the grouping scheme I and II used for Table 2 are [0, 1, 2, 3-5, 6-9, 10+] and [0, 1-2, 3-5, 6-9, 10-19, 20-39, 40+], respectively. Results are based on 1000 simulations.

(continued)

Table 2 Results from repeated simulations for MLE-GRC using hypothetical Poisson and Zero-inflated Poisson distributions*

			N=1000									N=5000								
Scheme	λ	P	$\hat{\lambda}_{Poisson}$	Std. Dev.	Coverage	$\hat{\lambda}_{ZIP}$	Std. Dev.	Coverage	$\hat{P}_{ZIP}$	Std. Dev.	Coverage	$\hat{\lambda}_{Poisson}$	Std. Dev.	Coverage	$\hat{\lambda}_{ZIP}$	Std. Dev.	Coverage	$\hat{P}_{ZIP}$	Std. Dev.	Coverage
I	**0.5**	0.2	0.100	0.012	0%	0.496	0.110	70%	0.210	0.047	71%	0.100	0.005	0%	0.500	0.047	72%	0.201	0.018	71%
		0.4	0.199	0.016	0%	0.503	0.071	74%	0.404	0.055	71%	0.199	0.007	0%	0.501	0.034	69%	0.400	0.025	69%
		0.6	0.301	0.019	0%	0.500	0.059	69%	0.609	0.066	70%	0.299	0.008	0%	0.499	0.028	68%	0.602	0.031	67%
		0.8	0.399	0.021	0%	0.500	0.053	69%	0.807	0.076	68%	0.399	0.009	0%	0.499	0.023	58%	0.802	0.034	64%
		1.0	**0.501**	**0.023**	**94%**	**0.518**	**0.034**	**81%**	**0.970**	**0.044**	**84%**	**0.500**	**0.010**	**95%**	**0.508**	**0.015**	**81%**	**0.985**	**0.021**	**80%**
	3	0.2	0.532	0.040	0%	2.992	0.145	95%	0.201	0.013	96%	0.531	0.018	0%	2.999	0.067	95%	0.200	0.006	95%
		0.4	1.101	0.056	0%	3.009	0.102	95%	0.400	0.016	95%	1.098	0.025	0%	3.000	0.047	95%	0.400	0.007	95%
		0.6	1.702	0.063	0%	2.999	0.080	96%	0.601	0.016	95%	1.699	0.027	0%	2.999	0.038	95%	0.600	0.007	96%
		0.8	2.335	0.064	0%	3.001	0.073	94%	0.800	0.015	94%	2.336	0.028	0%	3.002	0.031	95%	0.800	0.006	95%
		1.0	**2.999**	**0.058**	**96%**	**3.009**	**0.060**	**95%**	**0.997**	**0.005**	**98%**	**3.000**	**0.027**	**95%**	**3.004**	**0.028**	**93%**	**0.999**	**0.002**	**97%**
	10	0.2	1.639	0.110	0%	9.987	0.249	97%	0.200	0.013	95%	1.641	0.049	0%	10.005	0.118	95%	0.200	0.005	96%
		0.4	3.400	0.146	0%	10.009	0.192	96%	0.400	0.015	96%	3.396	0.067	0%	9.997	0.083	96%	0.400	0.007	95%
		0.6	5.306	0.168	0%	9.993	0.155	95%	0.600	0.015	95%	5.313	0.077	0%	9.998	0.069	96%	0.600	0.007	94%
		0.8	7.468	0.165	0%	10.007	0.137	95%	0.800	0.013	94%	7.458	0.073	0%	10.003	0.059	96%	0.800	0.006	96%
		1.0	**9.996**	**0.120**	**95%**	**9.997**	**0.120**	**95%**	**1.000**	**0.000**	**100%**	**10.002**	**0.055**	**92%**	**10.002**	**0.055**	**94%**	**1.000**	**0.000**	**100%**
II	**0.5**	0.2	0.087	0.010	0%	0.509	0.164	61%	0.220	0.085	51%	0.088	0.005	0%	0.531	0.085	64%	0.196	0.044	41%
		0.4	0.182	0.015	0%	0.526	0.130	61%	0.410	0.127	43%	0.182	0.007	0%	0.555	0.049	2%	0.368	0.054	9%
		0.6	0.283	0.019	0%	0.553	0.091	69%	0.567	0.117	36%	0.282	0.008	0%	0.566	0.015	0%	0.537	0.025	1%
		0.8	0.388	0.022	0%	0.581	0.026	17%	0.701	0.032	12%	0.388	0.010	0%	0.575	0.007	0%	0.710	0.013	0%
		1.0	**0.501**	**0.024**	**95%**	**0.623**	**0.036**	**0%**	**0.822**	**0.033**	**0%**	**0.500**	**0.010**	**95%**	**0.618**	**0.017**	**0%**	**0.829**	**0.015**	**0%**
	3	0.2	0.502	0.041	0%	2.999	0.152	94%	0.200	0.013	95%	0.502	0.019	0%	3.001	0.066	95%	0.200	0.006	95%
		0.4	1.058	0.057	0%	3.001	0.105	94%	0.400	0.017	95%	1.058	0.024	0%	3.000	0.046	95%	0.400	0.007	96%
		0.6	1.667	0.064	0%	3.002	0.086	95%	0.601	0.017	94%	1.663	0.029	0%	3.000	0.039	95%	0.600	0.007	94%
		0.8	2.312	0.064	0%	2.999	0.075	94%	0.800	0.015	94%	2.315	0.030	0%	3.001	0.032	95%	0.800	0.007	94%
		1.0	**3.001**	**0.059**	**95%**	**3.012**	**0.061**	**94%**	**0.997**	**0.005**	**97%**	**3.000**	**0.026**	**95%**	**3.005**	**0.028**	**94%**	**0.999**	**0.002**	**97%**
	10	0.2	1.649	0.112	0%	10.005	0.262	95%	0.200	0.013	95%	1.645	0.050	0%	10.003	0.116	96%	0.200	0.006	96%
		0.4	3.410	0.149	0%	10.016	0.180	97%	0.400	0.015	97%	3.410	0.066	0%	10.003	0.083	95%	0.400	0.007	96%
		0.6	5.333	0.175	0%	10.001	0.150	95%	0.600	0.016	94%	5.336	0.074	0%	9.999	0.066	96%	0.600	0.007	95%
		0.8	7.482	0.170	0%	9.994	0.128	96%	0.800	0.013	94%	7.487	0.075	0%	10.000	0.059	95%	0.800	0.006	96%
		1.0	**10.006**	**0.117**	**95%**	**10.007**	**0.117**	**95%**	**1.000**	**0.000**	**100%**	**10.001**	**0.053**	**94%**	**10.001**	**0.053**	**94%**	**1.000**	**0.000**	**100%**

*Note: the grouping scheme I and II used for Table 2 are [0, 1, 2, 3-5, 6-9, 10+] and [0, 1-2, 3-5, 6-9, 10-19, 20-39, 40+], respectively. Results are based on 1000 simulations.

Table 3 Frequency distributions of adolescent alcoholic drinking, MTF, 1996-2012

	Lifetime Drinking							Drinking in 12 months						
Year	**0**	**1-2**	**3-5**	**6-9**	**10-19**	**20-39**	**40+**	**0**	**1-2**	**3-5**	**6-9**	**10-19**	**20-39**	**40+**
2012	674	199	235	208	281	224	416	795	335	322	225	231	160	163
2011	675	191	271	211	294	211	422	825	384	297	237	237	136	151
2010	671	193	265	216	281	244	460	784	397	321	231	250	137	199
2009	582	191	239	223	270	245	446	701	385	292	222	262	158	171
2008	634	160	251	210	292	222	470	759	380	288	218	241	165	187
2007	676	195	227	204	306	229	531	808	354	277	243	271	157	239
2006	626	169	233	226	304	242	497	756	349	302	249	262	178	196
2005	610	197	272	243	244	246	544	745	390	308	237	265	197	212
2004	571	187	245	210	307	254	583	690	363	314	246	295	214	218
2003	535	187	232	260	295	257	595	687	373	322	282	256	171	261
2002	465	151	228	214	267	211	564	592	318	311	214	239	171	240
2001	447	145	188	185	298	287	514	574	298	269	263	272	179	217
2000	412	157	246	196	283	248	538	533	303	318	236	277	190	209
1999	420	178	215	192	311	261	632	558	364	265	247	256	238	274
1998	442	168	295	233	295	307	724	588	396	327	256	331	213	350
1997	492	171	247	233	340	310	694	638	367	338	261	325	262	285
1996	515	155	226	206	282	314	625	642	350	298	254	297	219	250

	Drinking in 30 days							Binge drinking					
Year	**0**	**1-2**	**3-5**	**6-9**	**10-19**	**20-39**	**40+**	**0**	**1**	**2**	**3-5**	**6-9**	**10+**
2012	1264	462	254	123	76	28	24	1665	197	150	142	29	19
2011	1346	432	234	122	74	22	34	1730	190	134	124	32	28
2010	1323	457	254	132	86	25	38	1716	220	158	129	35	27
2009	1224	438	257	138	85	24	24	1586	222	145	147	34	32
2008	1262	454	240	137	78	26	41	1630	215	160	135	44	24
2007	1293	451	275	144	107	37	45	1688	217	164	149	55	39
2006	1244	469	262	154	100	35	27	1661	211	160	145	45	32
2005	1257	459	304	157	104	33	32	1674	243	178	165	37	33
2004	1201	466	296	193	103	31	40	1635	253	171	170	61	26
2003	1193	504	247	196	123	41	37	1636	234	180	202	42	22
2002	1062	446	223	151	136	30	39	1456	200	161	166	47	25
2001	1036	430	239	173	114	40	27	1438	209	158	144	62	37
2000	1022	437	288	169	94	31	30	1425	202	186	150	40	36
1999	1041	425	285	214	147	40	53	1460	213	176	217	63	54
1998	1162	453	347	220	155	58	63	1634	228	224	228	61	57
1997	1153	527	324	207	170	46	47	1664	253	198	238	62	53
1996	1134	436	321	193	142	34	53	1572	236	191	202	55	47

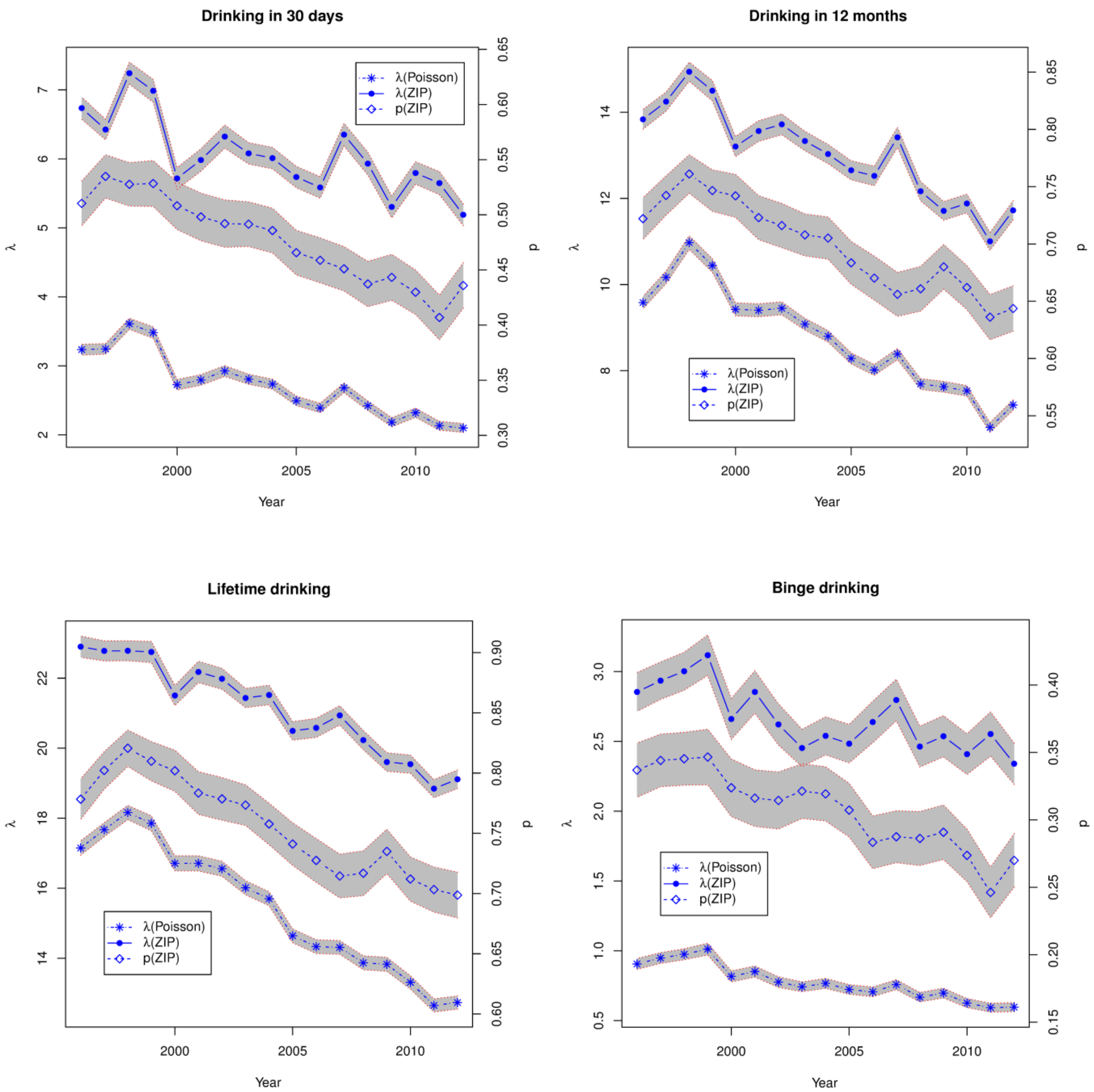

Figure 1 Adolescent alcoholic drinking estimated by MLE-GRC, MTF, 1996-2012

Note: the shaded area is the 95% confidence region.

APPENDIX

The proof of (A4)

To prove (A4), it is easy to see that $f(\alpha|\lambda)$ is infinitely differentiable: again, there is one and only one component of $(\alpha_1,\cdots,\alpha_G)$ that equals 1 so $f(\alpha|\lambda)=\theta_j(\lambda)$ if $\alpha_j=1$. $\theta_j(\lambda)$ is already defined in equation (3) and is infinitely differentiable with continuous derivatives for any order.

To investigate the third derivative of $f(\alpha|\lambda)$, we now introduce a function $D(\alpha)$ with its expectation $\mathrm{E}_{\lambda_0}[D(\alpha)] < \infty$ to bound the third derivative of $f(\alpha \mid \lambda)$:

$$\left| \frac{\partial^3 \log f(\alpha \mid \lambda)}{\partial \lambda^3} \right| \le D(\alpha) < \infty .$$

Next we try to find a $D(\alpha)$ such that the inequality above holds uniformly for any α and all $\lambda \in [\frac{1}{2}\lambda_0, \frac{3}{2}\lambda_0]$. We note that

$$\frac{\partial^3}{\partial \lambda^3} \log f = \frac{f''' f^2 - 3 f f' f'' + 2(f')^3}{f^3},$$

and for any $j = 1,\dots,G$ we can subsequently define $D(\alpha)$ as $\max\{D_1,\cdots,D_G\}$, where

$$D_j = \max_{\lambda \in \left[\frac{1}{2}\lambda_0, \frac{3}{2}\lambda_0\right]} \left| \frac{\theta'''_j \theta_j^2 - 3\theta_j \theta'_j \theta''_j + 2(\theta'_j)^3}{\theta_j^3} \right| .$$

$D(\alpha)$ is now independent of α and $D_j < \infty$ for any j. This proves $\mathrm{E}_{\lambda_0}[D(\alpha)] < \infty$.

The proof of (A6)

The proof of (A6) is trivial because $\theta_j > 0$ for any j. We also have

$$\mathrm{E}\frac{d}{d\lambda}\log f(\alpha \mid \lambda) = \sum_{j=1}^{G} \frac{d\theta_j}{d\lambda} = \frac{d}{d\lambda}\sum_{j=1}^{G}\theta_j = 0, \text{ and}$$

$$-\mathrm{E}\left[\frac{d^2}{d\lambda^2}\log f(\alpha \mid \lambda)\right] = \mathrm{E}\sum_{j=1}^{G}\frac{\alpha_j}{\theta_j^2}\left(\frac{d\theta_j}{d\lambda}\right)^2 - \mathrm{E}\sum_{j=1}^{G}\frac{\alpha_j}{\theta_j}\frac{d^2\theta_j}{d\lambda^2}$$

$$= \sum_{j=1}^{G}\frac{1}{\theta_j}\left(\frac{d\theta_j}{d\lambda}\right)^2 - \frac{d^2}{d\lambda^2}\sum_{j=1}^{G}\theta_j = \sum_{j=1}^{G}\frac{1}{\theta_j}\left(\frac{d\theta_j}{d\lambda}\right)^2 = I(\lambda).$$

The proof of (B4)

To prove (B4), we first demonstrate that

$$\frac{\partial}{\partial p}\log f(\alpha \mid p,\lambda) = \frac{\partial}{\partial p}\log(\theta_1^{\alpha_1}\cdots\theta_G^{\alpha_G}) = \frac{\partial}{\partial p}\sum_{j=1}^{G}\alpha_j \log\theta_j = \sum_{j=1}^{G}\frac{\alpha_j}{\theta_j}\frac{\partial\theta_j}{\partial p}, \text{ and}$$

$$\frac{\partial}{\partial\lambda}\log f(\alpha \mid p,\lambda) = \sum_{j=1}^{G}\frac{\alpha_j}{\theta_j}\frac{\partial\theta_j}{\partial\lambda}.$$

Next, we have

$$\mathrm{E}\frac{\partial}{\partial p}\log f(\alpha\,|\,p,\lambda)=\sum_{j=1}^{G}\frac{\partial\theta_j}{\partial p}=\frac{\partial}{\partial p}\sum_{j=1}^{G}\theta_j=0,\text{ and similarly }\mathrm{E}\frac{\partial}{\partial\lambda}\log f(\alpha\,|\,p,\lambda)=0.$$

Because one and only one of $\alpha_1,\cdots,\alpha_G$ is 1 and the rest are all 0 and $\mathrm{E}\left(\alpha_j\right)=\theta_j$, we also have

$$\mathrm{E}\left[\left(\frac{\partial}{\partial\lambda}\log f(\alpha\,|\,p,\lambda)\right)^2\right]=\mathrm{E}\left[\left(\sum_{j=1}^{G}\frac{\alpha_j}{\theta_j}\frac{\partial\theta_j}{\partial\lambda}\right)^2\right]=\mathrm{E}\sum_{j=1}^{G}\frac{\alpha_j}{\theta_j^2}\left(\frac{\partial\theta_j}{\partial\lambda}\right)^2=\sum_{j=1}^{G}\frac{1}{\theta_j}\left(\frac{\partial\theta_j}{\partial\lambda}\right)^2.$$

On the other hand,

$$\begin{aligned}-\mathrm{E}\left[\frac{\partial^2}{\partial\lambda^2}\log f(\alpha\,|\,p,\lambda)\right]&=\mathrm{E}\sum_{j=1}^{G}\frac{\alpha_j}{\theta_j^2}\left(\frac{\partial\theta_j}{\partial\lambda}\right)^2-\mathrm{E}\sum_{j=1}^{G}\frac{\alpha_j}{\theta_j}\frac{\partial^2\theta_j}{\partial\lambda^2}\\&=\sum_{j=1}^{G}\frac{1}{\theta_j}\left(\frac{\partial\theta_j}{\partial\lambda}\right)^2-\frac{\partial^2}{\partial\lambda^2}\sum_{j=1}^{G}\theta_j=\sum_{j=1}^{G}\frac{1}{\theta_j}\left(\frac{\partial\theta_j}{\partial\lambda}\right)^2.\end{aligned}$$

This proves equation (E3). Equations (E4) and (E5) are proved similarly.

The proof of (B5)

To prove (B5), all the four elements in the matrix (E6) are finite because $\theta_j>0$ for all j. For each j and any constants c_1 and c_2, one has

$$\frac{1}{\theta_j}c_1^2\left(\frac{\partial\theta_j}{\partial\lambda}\right)^2+2c_1c_2\frac{1}{\theta_j}\frac{\partial\theta_j}{\partial\lambda}\frac{\partial\theta_j}{\partial p}+c_2^2\frac{1}{\theta_j}\left(\frac{\partial\theta_j}{\partial p}\right)^2=\frac{1}{\theta_j}\left(c_1\frac{\partial\theta_j}{\partial\lambda}+c_2\frac{\partial\theta_j}{\partial p}\right)^2\geq 0,$$

which implies that $I(p,\lambda)$ is positive semi-definite. Moreover, the proof in Appendix shows that $I(p,\lambda)$ is strictly positive definite. To prove that $I(p,\lambda)$ is strictly positive definite, we need only to show that

$$\det\begin{pmatrix}\dfrac{\partial\theta_G}{\partial\lambda} & \dfrac{\partial\theta_{G-1}}{\partial\lambda}\\ \dfrac{\partial\theta_G}{\partial p} & \dfrac{\partial\theta_{G-1}}{\partial p}\end{pmatrix}\neq 0.$$

Since $G-1>1$, we define a function $\mu_j=\mu_j(\lambda)=\sum_{k\in I_j}\mathrm{e}^{-\lambda}\dfrac{\lambda^k}{k!}$ to obtain $\theta_G=p\mu_G$ and $\theta_{G-1}=p\mu_{G-1}$.

Now consider

$$\Delta = \det\begin{pmatrix} \mu'_G & \mu'_{G-1} \\ \mu_G & \mu_{G-1} \end{pmatrix} = e^{-2\lambda}\det\begin{pmatrix} \dfrac{\lambda^{l_G-1}}{(l_G-1)!} & \dfrac{\lambda^{l_{G-1}-1}}{(l_{G-1}-1)!} - \dfrac{\lambda^{l_G-1}}{(l_G-1)!} \\ \sum_{k=l_G}^{\infty}\dfrac{\lambda^k}{k!} & \sum_{k=l_{G-1}}^{l_G-1}\dfrac{\lambda^k}{k!} \end{pmatrix}$$

$$= e^{-2\lambda}\det\begin{pmatrix} \dfrac{\lambda^{l_G-1}}{(l_G-1)!} & \dfrac{\lambda^{l_{G-1}-1}}{(l_{G-1}-1)!} \\ \sum_{k=l_G-1}^{\infty}\dfrac{\lambda^k}{k!} & \sum_{k=l_{G-1}-1}^{\infty}\dfrac{\lambda^k}{k!} \end{pmatrix}, \textit{ note that } \det\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \det\begin{pmatrix} a & b+a \\ c & d+c \end{pmatrix}.$$

We let $a = l_G - 1$ and $b = l_{G-1} - 1$ so that

$$e^{2\lambda}\Delta = \frac{\lambda^a}{a!}\sum_{k=b}^{\infty}\frac{\lambda^k}{k!} - \frac{\lambda^b}{b!}\sum_{k=a}^{\infty}\frac{\lambda^k}{k!} = \sum_{s=0}^{\infty}\left\{\frac{\lambda^{a+b+s}}{(b+s)!a!} - \frac{\lambda^{a+b+s}}{b!(a+s)!}\right\} = \sum_{s=1}^{\infty}\frac{\lambda^{a+b+s}}{a!b!}\left(\prod_{t=1}^{s}\frac{1}{b+t} - \prod_{t=1}^{s}\frac{1}{a+t}\right).$$

Then we have $\Delta > 0$ since $a > b$.

References

Allison, Paul D and Richard P Waterman. 2002. "Fixed–Effects Negative Binomial Regression Models." Sociological methodology 32(1): 247-65

Barron, David. 1992. "The Analysis of Count Data: Over-Dispersion and Autocorrelation." Sociological methodology 22: 179-220

Burnham, Kenneth P and David R Anderson. 2004. "Multimodel Inference Understanding Aic and Bic in Model Selection." Sociological Methods & Research 33(2): 261-304

Cameron, A.Colin and Pravin K. Trivedi. 1998. Regression Analysis of Count Data. Cambridge University Press.

Chen, Ding-Geng, Jianguo Sun and Karl E Peace. 2012. Interval-Censored Time-to-Event Data: Methods and Applications. Chapman and Hall/CRC.

DiBenedetto, Emmanuele. 2002. Real Analysis. Springer.

Dickman, Paul W, Andy Sloggett, Michael Hills and Timo Hakulinen. 2004. "Regression Models for Relative Survival." Statistics in medicine 23(1): 51-64

Fay, Michael P and Pamela A Shaw. 2010. "Exact and Asymptotic Weighted Logrank Tests for Interval Censored Data: The Interval R Package." Journal of Statistical Software 36(2)

Greene, William H. 1997. Econometric Analysis. Prentice Hall.

Groves, Robert M., Jr. Floyd J. Fowler, Mick P. Couper, James M. Lepkowski, Eleanor Singer and Roger Tourangeau. 2009. Survey Methodology. John Wiley & Sons.

Hall, Daniel B. 2000. "Zero-Inflated Poisson and Binomial Regression with Random Effects: A Case Study." Biometrics 56(4): 1030-39

Heitjan, Daniel F and Donald B Rubin. 1991. "Ignorability and Coarse Data." The Annals of Statistics 19(4): 2244-53

Ibrahim, Joseph G, Ming-Hui Chen, Stuart R Lipsitz and Amy H Herring. 2005. "Missing-Data Methods for Generalized Linear Models: A Comparative Review." Journal of the American Statistical Association 100(469): 332-46

Lambert, Diane. 1992. "Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing." Technometrics 34(1): 1-14

Lehmann, Erich Leo and George Casella. 1998. Theory of Point Estimation. Springer.

Lindsey, Jane C and Louise M Ryan. 1998. "Methods for Interval-Censored Data." Statistics in medicine 17(2): 219-38

Long, J. Scott. 1997. Regression Models for Categorical and Limited Dependent Variables. Sage Publications, Incorporated.

Long, J. Scott and Jeremy Freese. 2006. Regression Models for Categorical Dependent Variables Using Stata. Stata Press.

McCullagh, Peter and John Ashworth Nelder. 1989. Generalized Linear Models. Chapman and Hall.

McGinley, James S, Patrick J Curran and Donald Hedeker. 2015. "A Novel Modeling Framework for Ordinal Data Defined by Collapsed Counts." Statistics in medicine 34(15): 2312-24

Newes-Adeyi, Gabriella, Chiung M Chen, Gerald D Williams and Vivian B Faden. 2005. "Surveillance Report: Trends in Underage Drinking in the United States, 1991-2003." National Institute on Alcohol Abuse and Alcoholism (Division of Epidemiology and Prevention Research).

Radloff, L.S. 1977. "The Ces-D Scale: A Self-Report Depression Scale for Research in the General Population." Applied psychological measurement 1(3): 385-401

Raftery, Adrian E. 1995. "Bayesian Model Selection in Social Research." Sociological methodology 25: 111-64

Rao, JNK and AJ Scott. 1999. "A Simple Method for Analysing Overdispersion in Clustered Poisson Data." Statistics in medicine 18(11): 1373-85

Saffari, Seyed Ehsan and Robiah Adnan. 2011. "Zero-Inflated Poisson Regression Models with Right Censored Count Data." Matematika 27(1)

Saffari, Seyed Ehsan, Robiah Adnan and William Greene. 2012. "Hurdle Negative Binomial Regression Model with Right Censored Count Data." Sort: Statistics and Operations Research Transactions 36(2): 181-94

Wallace Jr, John M, Jerald G Bachman, Patrick M O'Malley, Lloyd D Johnston, John E Schulenberg and Shauna M Cooper. 2002. "Tobacco, Alcohol, and Illicit Drug Use: Racial and Ethnic Differences among Us High School Seniors, 1976-2000." Public health reports 117(Suppl 1): S67-S75

Wong, Raymond Sin-Kwok. 1994. "Model Selection Strategies and the Use of Association Models to Detect Group Differences." Sociological Methods & Research 22(4): 460-91

Zorn, Christopher JW. 1998. "An Analytic and Empirical Examination of Zero-Inflated and Hurdle Poisson Specifications." Sociological Methods & Research 26(3): 368-400