Asymptotic distributions of statistics and parameter estimates for mixed Poisson processes

V. SAVANI & A. ZHIGLJAVSKY

Department of Mathematics, Cardiff University, Cardiff, CF24 4AG, U.K.

e-mail: SavaniV@Cardiff.ac.uk, ZhigljavskyAA@Cardiff.ac.uk

Abstract

Mixed Poisson processes have been used as natural models for events occurring in continuous or discrete time. Our main result is the derivation of the joint asymptotic distributions of statistics, including parameter estimators, computed in different time intervals from data generated by mixed Poisson processes. These distributions can be used, for example, to test the hypothesis about the adequacy of the mixed Poisson process against data. We provide some simulation results and test the model on actual market research data.

Keywords: Mixed Poisson process, mixed Poisson distribution, maximum likelihood, moment estimators.

AMS Subject Classification: 62M, 60K10, 62F, 90A40.

1 Introduction

Mixed Poisson processes have been used as natural models for events occurring in continuous or discrete time in many different fields including accident proneness (Greenwood et Yule (1920)), accidents and sickness (Lundberg (1964)), market research (Ehrenberg (1988)), risk theory (Grandell (1997)) and clinical trials (Cook et Wei (2003)). The main result of this paper is the derivation of the joint asymptotic distributions of statistics, including parameter estimators, computed in different time intervals from data generated by mixed Poisson processes.

The structure of the paper is as follows. In this section we introduce mixed Poisson processes and our main special case, the gamma-Poisson process. We also discuss the market research interpretation of this process which we will use throughout the paper. In Section 2 we give a general expression for the asymptotic covariance matrix of functionals of data from mixed Poisson processes. Using this general setup, in Section 3 we derive the asymptotic distributions between different statistics and estimators computed in different time intervals. These distributions allow hypothesis testing to assess goodness of fit of the mixed Poisson process. In Section 4 we apply general results of Section 3 to the case of the gamma-Poisson process. We provide some simulation results and test the model on actual market research data.

Mixed Poisson processes

Define the multivariate Poisson distribution as

$$\mathbb{P}\left(\boldsymbol{Z}=\boldsymbol{x}|\Lambda=\lambda\right)=\prod_{i=0}^{n-1}\frac{[\lambda(t_{i+1}-t_i)]^{x_{i+1}-x_i}}{(x_{i+1}-x_i)!}\exp\left(-\lambda(t_{i+1}-t_i)\right), \tag{1}$$

where $\lambda > 0$ is the intensity, $\boldsymbol{Z} = \{Z(t_1), Z(t_2), \ldots, Z(t_n)\}$ is a random vector, the set $\boldsymbol{x} = \{x_0, x_1, x_2, \ldots, x_n\}$ is a set of non-negative integers with $0 = x_0 \leqslant x_1 \leqslant \ldots \leqslant x_n$ and $0 = t_0 \leqslant t_1 \leqslant \ldots \leqslant t_n$ represents an increasing sequence of time points. The mixed Poisson process is then defined as a process whose finite-dimensional distributions are

$$\mathbb{P}\left(\boldsymbol{Z}=\boldsymbol{x}\right)=\int_{0-}^{\infty}\mathbb{P}\left(\boldsymbol{Z}=\boldsymbol{x}|\Lambda=\lambda\right)dU_\Lambda(\lambda;\boldsymbol{\theta}). \tag{2}$$

Here $U_\Lambda(\lambda;\boldsymbol{\theta})$ is the distribution function for the random variable Λ and $\boldsymbol{\theta}$ is a vector of unknown parameters. The function $U_\Lambda(\lambda;\boldsymbol{\theta})$ is commonly known as the structure distribution of the mixed Poisson process.

The most common distribution function for Λ is that of the gamma distribution with probability density function

$$g(\lambda;a,k)=\frac{1}{a^k\Gamma(k)}\lambda^{k-1}\mathrm{e}^{-\lambda/a},\quad a>0,\ k>0,\quad \lambda>0. \tag{3}$$

The mixed Poisson process in this case is often referred to as the gamma-Poisson process. The one-dimensional distribution of the gamma-Poisson process is the negative binomial distribution (NBD) with probabilities

$$p_x = \mathbb{P}\left(Z(t_1) = x\right) = \frac{\Gamma(k+x)}{\Gamma(k)x!}\left(\frac{1}{1+at_1}\right)^k \left(\frac{at_1}{1+at_1}\right)^x \quad x = 0, 1, 2, \dots . \tag{4}$$

In literature the parametrization (m, k), where $m = ak$, is often used. In addition to the gamma distribution, Grandell (1997) considered other distributions including beta, shifted-gamma, generalized inverse Gaussian and lognormal distributions as structure distributions for the mixed Poisson process.

Fitting the mixed Poisson process

The fitting of mixed Poisson processes to observed data has mainly focussed on fitting the one-dimensional mixed Poisson distribution when considering data observed over fixed time intervals. Fitting the one-dimensional distribution only gives partial information as to the adequacy of the process being fitted; in particular the dynamical behavior of the mixed Poisson process is not considered by fitting the one-dimensional distribution. The derivation of the joint asymptotic distributions of statistics and estimators allows testing the hypothesis as to whether parameter estimates computed in the two different time intervals could have been generated from the same process. This will allow us to verify the dynamical properties of the underlying model against data.

Note that it is easy to construct methods of testing the adequacy of mixed Poisson processes which are based on testing whether each individual realization follows the standard Poisson process; the distribution of the intensities of the individual Poisson processes can then be checked against a specified structure distribution. However, in practice the individual behavior basically never follows the pure Poisson model (see e.g. Cox et Hinkley (1978) ; Ehrenberg (1988)) and therefore the related tests would almost certainly reject the Poisson process assumption. At the same time, it is widely known that the Poisson and mixed Poisson models often work fairly well when the data is aggregated over either time or realizations, or both.

The asymptotic distributions derived in this paper allow us to test the mixed Poisson model hypotheses using the aggregated data, see Section 4.4. We are not aware of any other procedure of testing the dynamics of the mixed Poisson models, except those based on testing individual realizations (but these are not practical). In the practice of market research, when using panel data, we observe multiple realizations of data which can be aggregated.

When only a few events are registered in each individual realization, testing the pure Poisson hypothesis is meaningless as there is not enough data. However, the methodology described in this paper can be perfectly suitable for testing the mixed Poisson model if there are enough realizations in the multiple realization scheme and a suitable aggregation is made.

The mixed Poisson model for consumer buying behavior

Consumer purchase occasions represent the rate of recurrence with which households purchase products. Let $\{z_l(t_1), \ldots, z_l(t_n)\}$ represent the number of purchase occasions for household l up until the times $\{t_1, t_2, \ldots, t_n\}$. Assume that the purchasing process of a household follows a Poisson process with mean λ_l over a unit time interval. The distribution of purchases for a fixed household is then given by equation (1). If the intensity λ_l varies between individuals so that λ_l has the distribution function $U_\Lambda(\lambda; \boldsymbol{\theta})$ then, for fixed time points $\{t_1, \ldots, t_n\}$, the number of purchase occasions $\{z_l(t_1), \ldots, z_l(t_n)\}$ for a random household follows the mixed Poisson distribution (2). It is assumed that purchasing across households are independent events. The mixed Poisson model, when λ_l is gamma distributed, was applied to consumer buying behavior by many authors (see e.g. Goodhardt, Ehrenberg, et Chatfield (1984) ; Ehrenberg (1988)).

2 Asymptotic properties of a general estimator

This section considers the asymptotic distribution of a general class of estimators for a vector of parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)^T$ where the estimators satisfy the equation $G_i(\boldsymbol{\theta}, \overline{f_i}) = 0$ $(i = 1, \ldots, d)$, using d asymptotically normal statistics $\overline{f_i}$, with $G_i(\boldsymbol{\theta}, \overline{f_i}) = \mathbb{E} f_i(\zeta; \boldsymbol{\theta}) - \overline{f_i}$. The covariance matrix of the limiting normal distribution of the estimators is derived. Using this methodology for the gamma-Poisson process it is possible to derive, for example, the limiting normal distribution of the vector of parameter estimates $(\hat{k}^{(1)}, \hat{k}^{(2)}, \ldots, \hat{k}^{(t)})^T$, where $\hat{k}^{(i)}$ is an estimator for the shape parameter k of the NBD computed using data from the i'th time interval.

The results of this section can be considered as a reformulation of the results on M- and Z-estimators, see van der Vaart (1998, Chapters 3-5). We need this reformulation to unify notation and simplify exposition in the next sections.

2.1 General estimation scheme

Let ζ be a random variable taking values in some set $\mathcal{Z}$ and let ζ have probability mass function $p(z; \boldsymbol{\theta})$, $z \in \mathcal{Z}$, where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)^T$ $(d \geq 1)$ is a vector of parameters taking values in some set $\Theta \subseteq \mathbb{R}^d$ with non-empty interior $\mathrm{int}(\Theta)$. For the purpose of this paper we only need to consider $\mathcal{Z} = \{0, 1, \ldots\}$, but the results of this section can be extended to arbitrary sets $\mathcal{Z}$; in the case of continuous distributions, $p(z; \boldsymbol{\theta})$ is a density. We now define a general method of estimating $\boldsymbol{\theta}^* = (\theta_1^*, \ldots, \theta_d^*)^T \in \mathrm{int}(\Theta)$, the true parameter values of the sampling distribution, by using an i.i.d. sample $\{z_1, \ldots, z_N\}$ of values of ζ. An expression for the covariance matrix of the limiting normal distribution is also given.

Let $\boldsymbol{f} = (f_1, \ldots, f_d)^T \in \mathbb{R}^d$ where $f_i : \mathcal{Z} \times \Theta \to \mathbb{R}$ $(i = 1, \ldots, d)$ are some functions which are smooth enough and possibly depend on $\boldsymbol{\theta}$; set $\bar{\boldsymbol{f}} = (\bar{f}_1, \ldots, \bar{f}_d)^T \in \mathbb{R}^d$ with $\overline{f_i} = \frac{1}{N}\sum_{l=1}^{N} f_i(z_l; \boldsymbol{\theta})$. Since $\{z_1, \ldots, z_N\}$ form an i.i.d. sample of values of ζ we have $\mathbb{E}\overline{f_i} = \mathbb{E}f_i(\zeta; \boldsymbol{\theta})$. The estimator $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_d)^T$ is then defined to be the solution to the equations

$$G_i(\boldsymbol{\theta}, \overline{f_i}) = 0 \qquad i = 1, \ldots, d\,, \tag{5}$$

where $G_i(\boldsymbol{\theta}, \overline{f_i}) = \mathbb{E}f_i(\zeta; \boldsymbol{\theta}) - \overline{f_i}$.

Let us give several examples of possible functions f_i $(i = 1, \ldots, d)$:

Example 2.1.1. $f_i(z; \boldsymbol{\theta}) = \partial \log(p(z; \boldsymbol{\theta}))/\partial\theta_i$ implying $\mathbb{E}f_i(\zeta; \boldsymbol{\theta}) = 0$,

Example 2.1.2. $f_i(z; \boldsymbol{\theta}) = f_i(z)$ so that the functions f_i do not depend on $\boldsymbol{\theta}$,

Example 2.1.3. $f_i(z; \boldsymbol{\theta}) = z^i$ implying $\mathbb{E}f_i(\zeta; \boldsymbol{\theta}) = \mathbb{E}\zeta^i$,

2.2 Asymptotic normality of estimators

Theorem 2.1. *Assume that the function* $\boldsymbol{G}$ *is invertible as a function of* $\boldsymbol{\theta}$ *in some neighbourhood of* $(\boldsymbol{\theta}^*, \mathbb{E}\boldsymbol{f})$ *and let* $\hat{\boldsymbol{\theta}}$ *be the solution of* $\boldsymbol{G}(\boldsymbol{\theta}, \bar{\boldsymbol{f}}) = \boldsymbol{0}$. *Assume that* $\mathbb{E}\,|\partial g_i(\zeta, \boldsymbol{\theta})/\partial\theta_j| < \infty$ *for all* i, j. *Additionally, assume that the estimator* $\hat{\boldsymbol{\theta}}$ *is a consistent estimator of* $\boldsymbol{\theta}$ *and* $\sqrt{N}(\bar{\boldsymbol{f}} - \mathbb{E}\boldsymbol{f})$ *is asymptotically normally distributed* $\mathcal{N}(0, \mathbb{D}\boldsymbol{f})$, *where*

$$\mathbb{D}\boldsymbol{f} = \mathbb{E}(\boldsymbol{f} - \mathbb{E}\boldsymbol{f})(\boldsymbol{f} - \mathbb{E}\boldsymbol{f})^T = \|\mathrm{Cov}(f_i(\zeta; \boldsymbol{\theta}), f_j(\zeta; \boldsymbol{\theta}))\|_{i,j=1}^{d}\,.$$

Then asymptotically as $N \to \infty$,

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) \overset{\mathcal{D}}{\Rightarrow} \mathcal{N}(0, \boldsymbol{V}\,(\mathbb{D}\boldsymbol{f})\,\boldsymbol{V}^T) \tag{6}$$

where

$$\boldsymbol{V} = \left[\lim_{N\to\infty} \left.\frac{\partial \boldsymbol{G}(\boldsymbol{\theta}, \bar{\boldsymbol{f}})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right]^{-1}. \tag{7}$$

The proof can be found in van der Vaart (1998, Section 5.3).

Example 2.2.1. ***Maximum likelihood.*** We have $f_i(z; \boldsymbol{\theta}) = \partial \log(p(z; \boldsymbol{\theta}))/\partial\theta_i$ $(i = 1, \ldots, d)$ so that $\mathbb{E}\overline{f_i} = \mathbb{E}f_i(\zeta; \boldsymbol{\theta}) = 0$,

$$\mathbb{D}\boldsymbol{f} = \left\|\mathbb{E}\frac{\partial}{\partial\theta_i}\log p(\zeta; \boldsymbol{\theta})\frac{\partial}{\partial\theta_j}\log p(\zeta; \boldsymbol{\theta})\right\| = I(\boldsymbol{\theta})$$

$$\text{and}\quad V^{-1} = -\lim_{N\to\infty}\left\|\frac{1}{N}\sum_{l=1}^{N}\frac{\partial^2}{\partial\theta_j\partial\theta_i}\log p(z_l; \boldsymbol{\theta})\right\| = I(\boldsymbol{\theta}),$$

where $I(\boldsymbol{\theta})$ is the Fisher information matrix. The covariance matrix of the maximum likelihood estimators is therefore $\mathbb{D}\hat{\theta} = I(\boldsymbol{\theta})^{-1}I(\boldsymbol{\theta})I(\boldsymbol{\theta})^{-1} = I(\boldsymbol{\theta})^{-1}$.

Example 2.2.2. ***General method of moments.*** We have $f_i(z, \boldsymbol{\theta}) = f_i(z)$ $(i = 1, \dots, d)$ so that the functions f_i do not depend on the unknown parameters $\boldsymbol{\theta}$. This implies $\mathbb{D}\boldsymbol{f} = \|\text{Cov}(f_i(\zeta), f_j(\zeta))\|$ and $V^{-1} = \|\partial \mathbb{E}\boldsymbol{f}(\zeta)/\partial\boldsymbol{\theta}\|$.

3 Covariances of statistics and estimators

In this section we derive the asymptotic distributions of different statistics and estimators computed in two different time intervals. All of the results can be easily generalized to any number of intervals. In Section 3.1 we consider the covariance between the statistics $\bar{\boldsymbol{f}}$ defined in Section 2 when computed using data over different time intervals. In Section 3.2 we use the results of Theorem 2.1 and Section 3.1 to derive the joint asymptotic distribution of parameter estimates of mixed Poisson processes.

The general results of this section will be specialized, in Section 4, to the case of the gamma-Poisson process and then used to check the asymptotic covariance structure of the parameter estimators for real data.

3.1 Covariance between statistics

In Section 3.1.1 we consider the simpler case of computing the covariance between statistics for non-overlapping intervals. The results of this section will be used in Section 3.1.2 which covers the general case of possibly overlapping intervals.

3.1.1 Non-overlapping intervals

Note that since the Poisson process is a stationary process that is homogenous in time, considering covariances between two statistics computed over the intervals $[t_1, t_2)$ and $[t_3, t_4)$ with $0 \le t_1 < t_2 \le t_3 < t_4$ is equivalent to considering covariances between the same statistics over the time intervals $[0, t)$ and $[t, t+s)$, so that $t_1 = 0,\ t_2 = t_3 = t$ and $t_4 = t + s$. Let us consider the covariance between the statistics

$$\bar{\phi}_{0,t} = \frac{1}{N}\sum_{l=1}^{N} \phi(z_l(0,t)) \quad \text{and} \quad \bar{\psi}_{t,t+s} = \frac{1}{N}\sum_{l=1}^{N} \psi(z_l(t,t+s)),$$

where $\{z_1(0,t), \dots, z_N(0,t)\}$ and $\{z_1(t,t+s), \dots, z_N(t,t+s)\}$ are i.i.d. data from a mixed Poisson process observed over two adjacent time intervals $[0, t)$ and $[t, t+s)$ respectively $(t, s > 0)$. Here ϕ and ψ are some functions possibly dependent upon the vector of parameters $\boldsymbol{\theta}$.

We note that for fixed u and v the observations $z_l(u, u+v)$ $(l = 1, \ldots, N)$ are mutually independent. For fixed l, the observations $z_l(0, t)$ and $z_l(t, t+s)$ are conditionally independent and Poisson distributed with means $\lambda_l t$ and $\lambda_l s$ respectively. Here λ_l is random for $l = 1, \ldots, N$ (e.g. different households), but is the same for fixed l (e.g. fixed household) as time varies. The samples $\{z_1(0,t), \ldots, z_N(0,t)\}$ and $\{z_1(t, t+s), \ldots, z_N(t, t+s)\}$ are dependent since, for each l, $z_l(0,t)$ and $z_l(t, t+s)$ are Poisson distributed with common λ_l.

Let $\zeta_{u,v}$ be a random variable whose distribution is identical to the distribution of the i.i.d. random variables $z_l(u, v)$ $(l = 1, \ldots, N)$, the number of events occurring in the time interval $[u, v)$. Then

$$N\mathrm{Cov}\left[\bar{\phi}_{0,t}, \bar{\psi}_{t,t+s}\right] = \mathrm{Cov}\left[\phi(\zeta_{0,t}), \psi(\zeta_{t,t+s})\right]. \tag{8}$$

Let $I_{[z=0]}$ be the indicator function such that $I_{[z=0]} = 1$ if $z = 0$ and $I_{[z=0]} = 0$ otherwise. Let $p_{[u,v)}(z; \boldsymbol{\theta})$ denote the mixed Poisson distribution over the time interval $[u, v)$. Let c be a positive constant not equal to 1. Consider the following cases:

Case 1. $\phi(z) = z^\alpha, \psi(z) = z^\beta$:

$$\mathrm{Cov}\left[\phi(\zeta_{0,t}), \psi(\zeta_{t,t+s})\right] = \mathbb{E}\mu_\alpha(\Lambda t)\mu_\beta(\Lambda s) - \mathbb{E}\mu_\alpha(\Lambda t)\mathbb{E}\mu_\beta(\Lambda s),$$

where $\mu_\alpha(\nu) = \mathbb{E}\kappa_\nu^\alpha$ and κ_ν is a Poisson random variable with intensity ν.

Case 1a. $\phi(z) = z, \psi(z) = z$:

$$\mathrm{Cov}\left[\phi(\zeta_{0,t}), \psi(\zeta_{t,t+s})\right] = \mathbb{E}\zeta_{0,t}\zeta_{t,t+s} - \mathbb{E}\zeta_{0,t}\mathbb{E}\zeta_{t,t+s} = \mathbb{E}\Lambda^2 ts - \mathbb{E}\Lambda t\mathbb{E}\Lambda s = ts\mathrm{Var}\,\Lambda.$$

Case 1b. $\phi(z) = z, \psi(z) = z^2$:

$$\mathrm{Cov}\left[\phi(\zeta_{0,t}), \psi(\zeta_{t,t+s})\right] = \mathbb{E}\zeta_{0,t}\zeta_{t,t+s}^2 - \mathbb{E}\zeta_{0,t}\mathbb{E}\zeta_{t,t+s}^2 = ts^2\mathrm{Cov}[\Lambda, \Lambda^2] + ts\mathrm{Var}\Lambda.$$

Case 2. $\phi(z) = z, \psi(z) = c^z$:

$$\begin{aligned}\mathrm{Cov}\left[\phi(\zeta_{0,t}), \psi(\zeta_{t,t+s})\right] &= \mathbb{E}\zeta_{0,t}c^{\zeta_{t,t+s}} - \mathbb{E}\zeta_{0,t}\mathbb{E}c^{\zeta_{t,t+s}} = \mathbb{E}\Lambda t e^{-\Lambda s(1-c)} - \mathbb{E}\Lambda t\,\mathbb{E}e^{-\Lambda s(1-c)} \\ &= -t\left[\mathcal{L}'(s(1-c)) + \mathbb{E}\Lambda\,\mathcal{L}(s(1-c))\right].\end{aligned}$$

Here $\mathcal{L}(c) = \mathbb{E}e^{-c\Lambda}$ is the Laplace transform of the random variable Λ and $\mathcal{L}'(c) = \frac{\partial}{\partial c}\mathbb{E}e^{-c\Lambda} = -\mathbb{E}[\Lambda e^{-c\Lambda}]$.

3.1.2 Overlapping intervals

In this section we consider covariances between statistics in the most general case when the intervals are possibly overlapping. This includes the cases when the intervals do not overlap and also when the intervals coincide. Let us consider the covariance between the statistics

$$\bar{\phi}_{t_1,t_3} = \frac{1}{N}\sum_{l=1}^{N}\phi(z_l(t_1,t_3)) \quad \text{and} \quad \bar{\psi}_{t_2,t_4} = \frac{1}{N}\sum_{l=1}^{N}\psi(z_l(t_2,t_4)),$$

where $\{z_1(t_1,t_3),\ldots,z_N(t_1,t_3)\}$ and $\{z_1(t_2,t_4),\ldots,z_N(t_2,t_4)\}$ are data from a mixed Poisson process observed over two possibly overlapping intervals $[t_1,t_3)$ and $[t_2,t_4)$ with $0 \le t_1 \le t_2 \le t_3 \le t_4$. Similarly to (8) we have

$$N\mathrm{Cov}\left[\bar{\phi}_{t_1,t_3},\bar{\psi}_{t_2,t_4}\right] = \mathbb{E}\phi(\zeta_{t_1,t_3})\psi(\zeta_{t_2,t_4}) - \mathbb{E}\phi(\zeta_{t_1,t_3})\mathbb{E}\psi(\zeta_{t_2,t_4}) = \mathrm{Cov}[\phi(\zeta_{t_1,t_3}),\psi(\zeta_{t_2,t_4})].$$

Computing these covariances for different functions ϕ and ψ can be simplified using the fact that the Poisson process has stationary and independent increments. Some covariances are given below.

Case 1a. $\phi(z) = z, \psi(z) = z$:

$$\begin{aligned}
\mathrm{Cov}[\phi(\zeta_{t_1,t_3}),\psi(\zeta_{t_2,t_4})] &= \mathbb{E}\zeta_{t_1,t_3}\zeta_{t_2,t_4} - \mathbb{E}\zeta_{t_1,t_3}\mathbb{E}\zeta_{t_2,t_4} \\
&= \mathbb{E}(\zeta_{t_1,t_2}+\zeta_{t_2,t_3})(\zeta_{t_2,t_3}+\zeta_{t_3,t_4}) - (\mathbb{E}\zeta_{t_1,t_2}+\mathbb{E}\zeta_{t_2,t_3})(\mathbb{E}\zeta_{t_2,t_3}+\mathbb{E}\zeta_{t_3,t_4}) \\
&= \mathrm{Cov}(\zeta_{t_1,t_2},\zeta_{t_2,t_3}) + \mathrm{Cov}(\zeta_{t_1,t_2},\zeta_{t_3,t_4}) + \mathrm{Cov}(\zeta_{t_2,t_3},\zeta_{t_3,t_4}) + \mathrm{Var}(\zeta_{t_2,t_3})
\end{aligned}$$

and using the results of case 1a of Section 3.1 we obtain

$$\begin{aligned}
\mathrm{Cov}[\phi(\zeta_{t_1,t_3}),\psi(\zeta_{t_2,t_4})] &= (t_3-t_2)\,\mathbb{E}\Lambda + (t_3-t_2)^2\mathrm{Var}\Lambda \\
&\quad + [(t_2-t_1)(t_3-t_2) + (t_2-t_1)(t_4-t_3) + (t_3-t_2)(t_4-t_3)]\,\mathrm{Var}\Lambda \\
&= (t_4-t_2)(t_3-t_1)\mathrm{Var}\Lambda + (t_3-t_2)\mathbb{E}\Lambda.
\end{aligned}$$

Case 1b. $\phi(z) = z, \psi(z) = z^2$:

$$\begin{aligned}
\mathrm{Cov}[\phi(\zeta_{t_1,t_3}),\psi(\zeta_{t_2,t_4})] &= \mathbb{E}\zeta_{t_1,t_3}\zeta^2_{t_2,t_4} - \mathbb{E}\zeta_{t_1,t_3}\mathbb{E}\zeta^2_{t_2,t_4} \\
&= \mathbb{E}(\zeta_{t_1,t_2}+\zeta_{t_2,t_3})(\zeta_{t_2,t_3}+\zeta_{t_3,t_4})^2 - (\mathbb{E}\zeta_{t_1,t_2}+\mathbb{E}\zeta_{t_2,t_3})\mathbb{E}(\zeta_{t_2,t_3}+\zeta_{t_3,t_4})^2 \\
&= \mathrm{Cov}(\zeta_{t_1,t_2},\zeta^2_{t_2,t_3}) + 2\mathrm{Cov}(\zeta_{t_1,t_2},\zeta_{t_2,t_3}\zeta_{t_3,t_4}) + \mathrm{Cov}(\zeta_{t_1,t_2},\zeta^2_{t_3,t_4}) \\
&\quad + \mathrm{Cov}(\zeta_{t_2,t_3},\zeta^2_{t_2,t_3}) + 2\mathrm{Cov}(\zeta_{t_2,t_3},\zeta_{t_2,t_3}\zeta_{t_3,t_4}) + \mathrm{Cov}(\zeta_{t_2,t_3},\zeta^2_{t_3,t_4})
\end{aligned}$$

and using the results of case 1b of Section 3.1 we obtain

$$\begin{aligned}
\mathrm{Cov}[\phi(\zeta_{t_1,t_3}),\psi(\zeta_{t_2,t_4})] &= (t_4-t_2)^2(t_3-t_1)\mathrm{Cov}(\Lambda,\Lambda^2) + (t_4-t_2)(t_3-t_1)\mathrm{Var}\Lambda \\
&\quad + 2(t_4-t_2)(t_3-t_2)\mathbb{E}\Lambda^2 + (t_3-t_2)\mathbb{E}\Lambda.
\end{aligned}$$

Case 2. $\phi(z) = z, \psi(z) = c^z$:

$$\begin{aligned}
\mathrm{Cov}[\phi(\zeta_{t_1,t_3}), \psi(\zeta_{t_2,t_4})] &= \mathbb{E}\zeta_{t_1,t_3} c^{\zeta_{t_2,t_4}} - \mathbb{E}\zeta_{t_1,t_3} \mathbb{E} c^{\zeta_{t_2,t_4}} \\
&= \mathbb{E}(\zeta_{t_1,t_2} + \zeta_{t_2,t_3}) c^{\zeta_{t_2,t_4}} - \mathbb{E}(\zeta_{t_1,t_2} + \zeta_{t_2,t_3}) \mathbb{E} c^{\zeta_{t_2,t_4}} \\
&= \mathrm{Cov}\left(\zeta_{t_1,t_2}, c^{\zeta_{t_2,t_4}}\right) + \mathbb{E}\zeta_{t_2,t_3} c^{\zeta_{t_2,t_3}} c^{\zeta_{t_3,t_4}} - \mathbb{E}\zeta_{t_2,t_3} \mathbb{E} c^{\zeta_{t_2,t_4}}.
\end{aligned}$$

Using the result of case 2, Section 3.1 we have

$$\mathrm{Cov}\left(\zeta_{t_1,t_2}, c^{\zeta_{t_2,t_4}}\right) = -(t_2 - t_1)\left[\mathcal{L}'((t_4 - t_2)(1-c)) + \mathbb{E}\Lambda\mathcal{L}((t_4 - t_2)(1-c))\right].$$

Similarly, we obtain

$$\begin{aligned}
\mathbb{E}\zeta_{t_2,t_3} c^{\zeta_{t_2,t_3}} c^{\zeta_{t_3,t_4}} &= \mathbb{E}_\Lambda\left[\mathbb{E}\left(\zeta_{t_2,t_3} c^{\zeta_{t_2,t_3}} | \Lambda = \lambda\right) \mathbb{E}\left(c^{\zeta_{t_3,t_4}} | \Lambda = \lambda\right)\right] \\
&= \mathbb{E}_\Lambda\left[\left(\Lambda c(t_3 - t_2) e^{-\Lambda(t_3 - t_2)(1-c)}\right)\left(e^{-\Lambda(t_4 - t_3)(1-c)}\right)\right] \\
&= c(t_3 - t_2)\mathbb{E}_\Lambda\left[\Lambda e^{-\Lambda(t_4 - t_2)(1-c)}\right] \\
&= -c(t_3 - t_2)\mathcal{L}'\left((t_4 - t_2)(1-c)\right).
\end{aligned}$$

Noting that $\mathbb{E}\zeta_{t_2,t_3} \mathbb{E} c^{\zeta_{t_2,t_4}} = (t_3 - t_2)\mathbb{E}\Lambda\mathcal{L}\left((t_4 - t_2)(1-c)\right)$ we combine the results to obtain

$$\begin{aligned}
&\mathrm{Cov}[\zeta_{t_1,t_3}, c^{\zeta_{t_2,t_4}}] = -(t_2 - t_1)\left[\mathcal{L}'((t_4 - t_2)(1-c)) + \mathbb{E}\Lambda\mathcal{L}((t_4 - t_2)(1-c))\right] \\
&- c(t_3 - t_2)\mathcal{L}'\left((t_4 - t_2)(1-c)\right) - (t_3 - t_2)\mathbb{E}\Lambda\mathcal{L}\left((t_4 - t_2)(1-c)\right) \\
&= -\left[(t_2 - t_1) + c(t_3 - t_2)\right]\mathcal{L}'((t_4 - t_2)(1-c)) - (t_3 - t_1)\mathbb{E}\Lambda\mathcal{L}((t_4 - t_2)(1-c)).
\end{aligned}$$

In particular if $t_1 = 0$, $t_2 = t_3 = t$ and $t_4 = t + s$ we obtain the results of Section 3.1.1, i.e. the covariances between statistics in non-overlapping intervals, in all three cases.

3.2 Covariances between parameter estimators

Let $\hat{\boldsymbol{\theta}}^{(1)}$ and $\hat{\boldsymbol{\theta}}^{(2)}$ be estimators of $\boldsymbol{\theta}$ in the intervals $[t_1, t_3)$ and $[t_2, t_4)$ constructed using the general scheme of Section 2.1 with the sets of functions $\{f_i^{(1)}(z; \boldsymbol{\theta})\}_{i=1}^d$ and $\{f_i^{(2)}(z; \boldsymbol{\theta})\}_{i=1}^d$, respectively. Assume that Theorem 2.1 applies to $\hat{\boldsymbol{\theta}}^{(1)}$ and $\hat{\boldsymbol{\theta}}^{(2)}$ so that both estimators are asymptotically normal and let $\boldsymbol{V}^{(1)}$, $\boldsymbol{V}^{(2)}$, $\mathbb{D}\boldsymbol{f}^{(1)}$ and $\mathbb{D}\boldsymbol{f}^{(2)}$ be the matrices associated with $\hat{\boldsymbol{\theta}}^{(1)}$ and $\hat{\boldsymbol{\theta}}^{(2)}$. We have $\sqrt{N}(\bar{\boldsymbol{f}} - \mathbb{E}\boldsymbol{f})$ is asymptotically normal $\mathcal{N}(0, \mathbb{D}\boldsymbol{f})$, where

$$\boldsymbol{f}(z; \boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{f}^{(1)}(z; \boldsymbol{\theta}) \\ \boldsymbol{f}^{(2)}(z; \boldsymbol{\theta}) \end{pmatrix}, \quad \bar{\boldsymbol{f}} = \begin{pmatrix} \bar{\boldsymbol{f}}^{(1)} \\ \bar{\boldsymbol{f}}^{(2)} \end{pmatrix}, \quad \mathbb{E}\boldsymbol{f} = \begin{pmatrix} \mathbb{E}\boldsymbol{f}^{(1)}(\zeta_{t_1,t_3}; \boldsymbol{\theta}) \\ \mathbb{E}\boldsymbol{f}^{(2)}(\zeta_{t_2,t_4}; \boldsymbol{\theta}) \end{pmatrix}$$

and

$$\mathbb{D}\boldsymbol{f} = \begin{pmatrix} \mathbb{D}\boldsymbol{f}^{(1)} & \mathbb{C}(\boldsymbol{f}^{(1)}, \boldsymbol{f}^{(2)}) \\ \mathbb{C}(\boldsymbol{f}^{(1)}, \boldsymbol{f}^{(2)})^T & \mathbb{D}\boldsymbol{f}^{(2)} \end{pmatrix} \tag{9}$$

with

$$\mathbb{C}(\boldsymbol{f}^{(1)}, \boldsymbol{f}^{(2)}) = \left\| \mathrm{Cov}(f_i^{(1)}(\zeta_{t_1,t_3}; \boldsymbol{\theta}), f_j^{(2)}(\zeta_{t_2,t_4}; \boldsymbol{\theta})) \right\|_{i,j=1}^{d}.$$

The components of the matrix $\mathbb{C}(\boldsymbol{f}^{(1)}, \boldsymbol{f}^{(2)})$ are computed using the results of Section 3.1.

Consider the problem of estimating the vector

$$\boldsymbol{\theta}_* = \begin{pmatrix} \boldsymbol{\theta}^{(1)} \\ \boldsymbol{\theta}^{(2)} \end{pmatrix} \quad \text{with the estimator} \quad \hat{\boldsymbol{\theta}}_* = \begin{pmatrix} \hat{\boldsymbol{\theta}}^{(1)} \\ \hat{\boldsymbol{\theta}}^{(2)} \end{pmatrix},$$

where $\boldsymbol{\theta}^{(1)}$ and $\boldsymbol{\theta}^{(2)}$ are two different copies of $\boldsymbol{\theta}$. The fact that $\boldsymbol{\theta}^{(1)}$ and $\boldsymbol{\theta}^{(2)}$ are two different copies of $\boldsymbol{\theta}$ implies that the matrix of partial derivatives $\boldsymbol{V}$, defined by equation (7) with $\boldsymbol{\theta}_*$ substituted for $\boldsymbol{\theta}^*$, has a block diagonal structure

$$\boldsymbol{V} = \begin{pmatrix} \boldsymbol{V}^{(1)} & 0 \\ 0 & \boldsymbol{V}^{(2)} \end{pmatrix}. \tag{10}$$

Using Theorem 2.1, $\sqrt{N}(\hat{\boldsymbol{\theta}}_* - \boldsymbol{\theta}_*)$ is asymptotically normal $\mathcal{N}(0, \boldsymbol{V}(\mathbb{D}\boldsymbol{f})\boldsymbol{V}^T)$, where $\mathbb{D}\boldsymbol{f}$ and $\boldsymbol{V}$ are defined by (9) and (10). The asymptotic covariance matrix is therefore

$$\boldsymbol{V}(\mathbb{D}\boldsymbol{f})\boldsymbol{V}^T = \begin{pmatrix} \boldsymbol{V}^{(1)}\, \mathbb{D}\boldsymbol{f}^{(1)}\, (\boldsymbol{V}^{(1)})^T & \boldsymbol{V}^{(1)}\, \mathbb{C}(\boldsymbol{f}^{(1)},\boldsymbol{f}^{(2)})\, (\boldsymbol{V}^{(2)})^T \\ \boldsymbol{V}^{(2)}\, (\mathbb{C}(\boldsymbol{f}^{(1)},\boldsymbol{f}^{(2)}))^T\, (\boldsymbol{V}^{(1)})^T & \boldsymbol{V}^{(2)}\, \mathbb{D}\boldsymbol{f}^{(2)}\, (\boldsymbol{V}^{(2)})^T \end{pmatrix}. \tag{11}$$

We shall use formulae (9) and (11) in Section 4 for deriving the asymptotic distributions of the estimators of the gamma-Poisson process where we plot scatter-plots of different estimators together with 95% confidence ellipses constructed using these formulae.

4 The gamma-Poisson process

In this section we derive the covariances between statistics and estimators computed in different time periods for data observed from the gamma-Poisson process. We consider only the case of overlapping intervals $[t_1, t_3)$ and $[t_2, t_4)$ where $0 \le t_1 \le t_2 \le t_3 \le t_4$ since the case of non-overlapping intervals $[0, t)$ and $[t, t+s)$ can be obtained by setting $t_1 = 0$, $t_2 = t_3 = t$ and $t_4 = t + s$. We apply these results to market research data where we estimate parameters of the joint asymptotic normal distribution of estimators; this asymptotic distribution is then checked against the empirical distribution of estimators for real data.

4.1 Estimating parameters of the NBD

The one-dimensional distribution of the gamma-Poisson process is the NBD. The NBD may therefore be used to estimate parameters of the gamma-Poisson process. Fisher

(1941) and Haldane (1941) independently considered estimation of the parameter pair (m, k), where $m = ak$, using maximum likelihood and Anscombe (1950) considered the use of moment based estimators. Anscombe (1950) noted that the maximum likelihood and all natural moment based estimators for the parameter pair (m, k) are asymptotically uncorrelated for an i.i.d. NBD sample. The estimation of NBD parameters in literature has therefore justifiably focussed on estimating the parameters m and k.

In market research the method of moments (MOM) and zero term method (ZTM) are commonly used (see e.g. Ehrenberg (1988) ; Savani et Zhigljavsky (2006a)). In Savani et Zhigljavsky (2006b) we consider the problem of efficient estimation using a variation of the power method of estimation introduced by Anscombe (1950). In this paper we consider three methods in the estimation of (m, k). The estimators $\hat{m}$ and $\hat{k}$ are obtained as the solutions to the equations $\bar{f}_1 - \mathbb{E}\bar{f}_1 = 0$ and $\bar{f}_2 - \mathbb{E}\bar{f}_2 = 0$, where $\bar{f}_1 = \frac{1}{N}\sum_{l=1}^{N} f_1(z_l; m, k)$ and $\bar{f}_2 = \frac{1}{N}\sum_{l=1}^{N} f_2(z_l; m, k)$ and $\{z_1, \ldots, z_N\}$ is an i.i.d. NBD sample observed from the gamma-Poisson process in a fixed time interval. The methods are defined by the functions f_1, f_2 which are as follows:

- **Standard method of moments (MOM):** $f_1(z) = z$, $f_2(z) = z^2$;
- **Zero term method (ZTM):** $f_1(z) = z$, $f_2(z) = 1$ if $z = 0$ and 0 otherwise;
- **Power method (PM(c)):** $f_1(z) = z$, $f_2(z) = c^z$ where $c > 0$, $c \neq 1$.

The ML, MOM, ZTM and PM estimators for m are identical with $\hat{m} = \frac{1}{N}\sum_{l=1}^{N} z_l$. This implies that the asymptotic variances of $\hat{m}$ for all four methods are identical. Table 1 shows equations for estimating k using ML and moment based methods (MOM, ZTM and PM). In Table 1, n_i denotes the observed frequency of $i = 0, 1, 2, \ldots$ within the sample. Note that for $c = 0$ the PM estimator is equivalent to the ZTM estimator and as $c \to 1$ the PM estimator tends to the MOM estimator.

Method	Estimator	Equation (solution in x) or estimator for k
ML	$\hat{k}_{ML}$	$\log\left(1 + \frac{\hat{m}}{x}\right) = \sum_{i=1}^{\infty} \frac{n_i}{N} \sum_{j=0}^{i-1} \frac{1}{x+j}$
MOM	$\hat{k}_{MOM}$	$\frac{\hat{m}^2}{s^2 - \hat{m}}$, where $s^2 = \frac{1}{N}\sum_{l=1}^{N} z_l^2 - \hat{m}^2$
ZTM	$\hat{k}_{ZTM}$	$\frac{n_0}{N} = \left(1 + \frac{\hat{m}}{x}\right)^{-x}$
PM	$\hat{k}_{PM(c)}$	$\frac{1}{N}\sum_{l=1}^{N} c^{z_l} = \left(1 + \frac{\hat{m}(1-c)}{x}\right)^{-x}$

Table 1: Maximum likelihood and moment based estimators for k.

Efficiency of estimators

The variances and covariances of the estimators $(\hat{m}, \hat{k})$ computed in a fixed time interval for different methods of estimation have been previously derived by Fisher (1941), Haldane (1941) and Anscombe (1950). The asymptotic variances below also follow from Theorem 2.1. We have $\lim_{N\to\infty} N\text{Cov}(\hat{m}, \hat{k}) = 0$ for all four methods of estimation and $\lim_{N\to\infty} N\ \text{Var}(\hat{m}) = ka(1+a)$. The asymptotic variances of $\hat{k}$ are

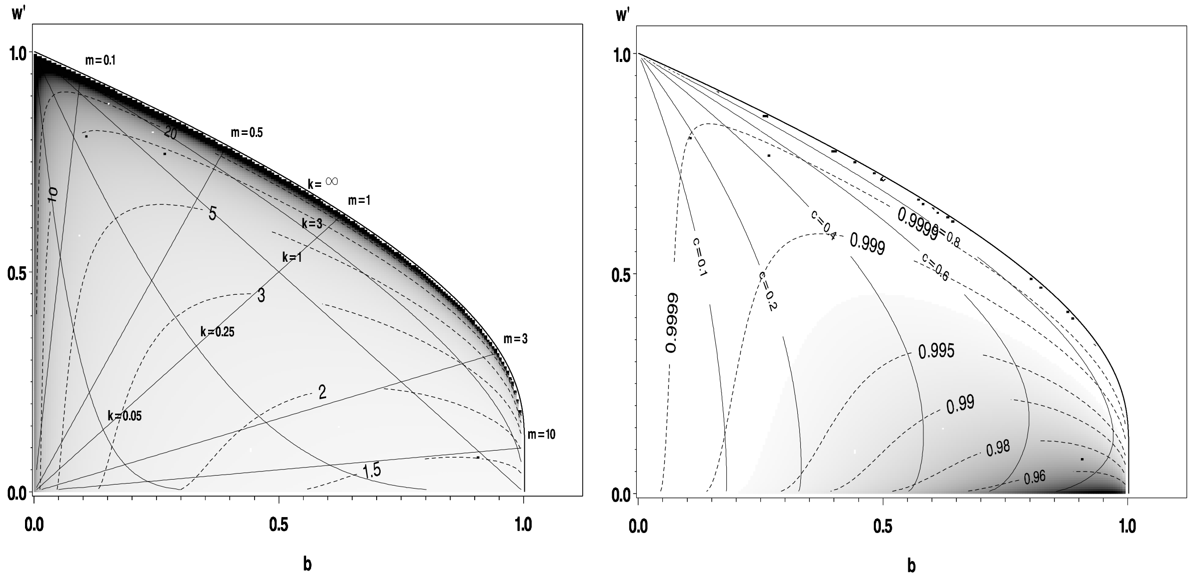
$$
\begin{aligned}
v_{ML} &= \lim_{N\to\infty} N\ \text{Var}\left(\hat{k}_{ML}\right) = \frac{2k(k+1)(1+a)^2}{a^2\left(1+2\sum_{j=2}^{\infty}\left(\frac{a}{1+a}\right)^{j-1}\frac{j!\Gamma(k+2)}{(j+1)\Gamma(k+j+1)}\right)},\\
v_{PM}(c) &= \lim_{N\to\infty} N\text{Var}\left(\hat{k}_{PM(c)}\right) = \frac{(1+a-ac^2)^{-k}r^{2k+2} - r^2 - ka(1+a)(1-c)^2}{\left[r\log(r) - r + 1\right]^2},\\
v_{MOM} &= \lim_{N\to\infty} N\text{Var}\left(\hat{k}_{MOM}\right) = \frac{2k(k+1)(1+a)^2}{a^2},\\
v_{ZTM} &= \lim_{N\to\infty} N\text{Var}\left(\hat{k}_{ZTM}\right) = \frac{(1+a)^{k+2} - (1+a)^2 - ka(1+a)}{\left[(1+a)\log(1+a) - a\right]^2}.
\end{aligned}
\tag{12}
$$

where $r = 1 + a - ac$. It is easy to see that the asymptotic variance of $\hat{k}_{MOM}$ can be obtained by taking the limit of $v_{PM}(c)$ as $c \to 1$ and $v_{ZTM} = v_{PM}(0)$.

Fig. 1(a) shows, using solid lines, the parameter values of m and k within the parameter space (b, w') where $b = (1 + m/k)^{-k} \in [0, 1]$ and $w' = b/m \in [0, 1]$. The parameter space (b, w') makes a visual comparison of the efficiency of different estimators easy. The parameters b and $w = 1/w'$ have a natural interpretation in consumer buying behavior in that b represents the probability of an individual to make at least one purchase and w represents the mean number of purchase occasions per buyer. The NBD is only defined for values of w' such that $w' < -b/\log(1-b)$.

Fig. 1(a) also shows, using dotted lines and grey shading, contour levels of the asymptotic normalized coefficient of variation $\sqrt{v_{ML}}/k$ of the ML estimator for k. Fig. 1(a) shows that increasing m for fixed k decreases the coefficient of variation $\sqrt{v_{ML}}/k$.

In Savani et Zhigljavsky (2006a) we note that there exists an optimum c, denoted by c_*, such that $v_{PM}(c)$ is minimum and that $v_{PM}(c_*)$ is very close to v_{ML}. The efficiency $v_{ML}/v_{PM}(c_*)$ is shown in Fig. 1(b), where the intensity of grey denotes the inefficiency (black shows efficiency< 0.95) and the level sets of the ratio $v_{ML}/v_{PM}(c_*)$ are shown by dotted lines. The contour levels of c_* are shown by solid lines. We can conclude from inspecting Fig. 1(b) that the power method with suitable c is practically as good as maximum likelihood; at the same time, the power method is simpler for implementing and studying. In Savani et Zhigljavsky (2006a) we demonstrate that the practically popular MOM and ZTM methods of estimation often produce rather inefficient estimators for k.



(a) $\sqrt{v_{ML}}/k$ (b) $v_{ML}/v_{PM}(c_*)$

Figure 1: (a) Coefficient of variation of $\hat{k}_{ML}$ (dotted lines) and values of m and k (solid lines) within the (b, w') parameter space, and (b) efficiency of the optimum PM estimator with respect to MLE (dotted lines) with contour levels of optimum c (solid lines).

4.2 Covariance between statistics

In this section we consider the covariance between the statistics $\bar{\phi}_{t_1,t_3} = \frac{1}{N}\sum_{l=1}^{N}\phi(z_l(t_1,t_3))$ and $\bar{\psi}_{t_2,t_4} = \frac{1}{N}\sum_{l=1}^{N}\psi(z_l(t_2,t_4))$, which are computed using data in the time intervals $[t_1, t_3)$ and $[t_2, t_4)$ respectively, for the following pairs of functions $(\phi(z), \psi(z))^T \in \{(z,z)^T, (z,z^2)^T, (z, I_{[z=0]})^T, (z, c^z)^T\}$. Here $I_{[z=0]}$ denotes the indicator function of the event $z = 0$ and $c > 0$ with $c \neq 1$. Note that the covariances between these statistic pairs will allow us to compute the joint asymptotic distribution for the MOM, ZTM and PM estimators computed in two different time periods, when estimating the parameters of the NBD. For the gamma distributed random variable Λ with density (3) we have:

$$\mathbb{E}\Lambda^\alpha = \frac{a^\alpha \Gamma(k+\alpha)}{\Gamma(k)} \quad (\alpha = 1,2,3)\,, \quad \operatorname{Var}\Lambda = a^2 k, \quad \mathcal{L}(c) = (1+ac)^{-k},$$
$$\mathcal{L}'(c) = -ak(1+ac)^{-k-1}, \quad \operatorname{Cov}(\Lambda, \Lambda^2) = 2a^3k(k+1).$$

Case 1a. $\phi(z) = z, \psi(z) = z$:

$$\operatorname{Cov}\left[\bar{\phi}, \bar{\psi}\right] = (t_4 - t_2)(t_3 - t_1)a^2k + ak(t_3 - t_2).$$

Case 1b. $\phi(z) = z, \psi(z) = z^2$:

$$\operatorname{Cov}\left[\bar{\phi}, \bar{\psi}\right] = 2(t_4 - t_2)^2(t_3 - t_1)a^3k(k+1) + (t_4 - t_2)(t_3 - t_1)a^2k$$
$$+ 2(t_4 - t_2)(t_3 - t_2)a^2k(k+1) + (t_3 - t_2)ak.$$

Case 2. $\phi(z) = z, \psi(z) = I_{[z=0]}$:

$$\mathrm{Cov}\left[\bar{\phi}, \bar{\psi}\right] = -\frac{ak[(t_3 - t_2) + a(t_3 - t_1)(t_4 - t_2)]}{(1 + a(t_4 - t_2))^{k+1}}.$$

Case 3. $\phi(z) = z, \psi(z) = c^z$:

$$\mathrm{Cov}\left[\bar{\phi}, \bar{\psi}\right] = -\frac{ak(1-c)[(t_3 - t_2) + a(t_3 - t_1)(t_4 - t_2)]}{(1 + a(t_4 - t_2)(1-c))^{k+1}}.$$

4.3 Covariances between estimators

This section considers the covariances between MOM, ZTM and PM estimators of the gamma-Poisson parameter pair (m, k) computed using data in the time interval $[t_1, t_3)$ and $[t_2, t_4)$. Consider estimating parameters using data in the time interval $[u, v)$; the estimator for m is identical for all three methods and is given by

$$\hat{m} = \frac{1}{N(v-u)} \sum_{l=1}^{N} z_l(u, v) \qquad v > u \geq 0.$$

The MOM, PM and ZTM methods use the respective statistics

$$\bar{\boldsymbol{f}}_{MOM} = \frac{1}{N} \sum_{l=1}^{N} \begin{pmatrix} z_l(u, v) \\ z_l^2(u, v) \end{pmatrix}, \quad \bar{\boldsymbol{f}}_{PM(c)} = \frac{1}{N} \sum_{l=1}^{N} \begin{pmatrix} z_l(u, v) \\ c^{z_l(u,v)} \end{pmatrix},$$
$$\text{and} \quad \bar{\boldsymbol{f}}_{ZTM} = \frac{1}{N} \sum_{l=1}^{N} \begin{pmatrix} z_l(u, v) \\ I_{[z_l(u,v)=0]} \end{pmatrix}. \tag{13}$$

The covariances of the statistics $\bar{\boldsymbol{f}}_{MOM}$, $\bar{\boldsymbol{f}}_{PM(c)}$ and $\bar{\boldsymbol{f}}_{ZTM}$ when computed over different time intervals were discussed in the previous section. In the computation of covariances between parameter estimates we therefore only require the matrix of partial derivatives $\boldsymbol{V}$ defined by equation (10). We have

$$\boldsymbol{V}_{MOM}^{-1} = \begin{bmatrix} \frac{1}{t} & 0 \\ \frac{1+2at(k+1)}{a^2t^2} & -\frac{1}{a^2t^2} \end{bmatrix}, \quad \boldsymbol{V}_{PM(c)}^{-1} = \begin{bmatrix} \frac{1}{t} & 0 \\ \frac{c-1}{r\log(r)-r+1} & -\frac{r^{k+1}}{r\log(r)-r+1} \end{bmatrix},$$
$$\boldsymbol{V}_{ZTM}^{-1} = \begin{bmatrix} \frac{1}{t} & 0 \\ -\frac{1}{(1+at)\log(1+at)-at} & -\frac{(1+at)^{k+1}}{(1+at)\log(1+at)-at} \end{bmatrix},$$

where $r = 1 + at(1-c)$ and $t = v - u$.

The expression of the covariance matrix is only simple for estimators computed in the case of non-overlapping time intervals $[0, t)$ and $[t, t+s)$, which can be obtained by setting

$t_1 = 0$, $t_2 = t_3 = t$ and $t_4 = t + s$. For the MOM, we have in the case of non-overlapping intervals

$$\mathbb{D}\begin{pmatrix} \hat{m}_t \\ \hat{k}_t \\ \hat{m}_s \\ \hat{k}_s \end{pmatrix} = \begin{bmatrix} \frac{a\,k\,(1+a\,t)}{t} & 0 & a^2k & 0 \\ 0 & \frac{2k(k+1)(1+a\,t)^2}{a^2t^2} & 0 & 2\,k\,(k+1) \\ a^2k & 0 & \frac{a\,k\,(1+a\,s)}{s} & 0 \\ 0 & 2\,k\,(k+1) & 0 & \frac{2k(k+1)(1+a\,s)^2}{a^2s^2} \end{bmatrix} \tag{14}$$

and the covariance matrix for the PM and ZTM estimators is

$$\mathbb{D}\begin{pmatrix} \hat{m}_t \\ \hat{k}_t \\ \hat{m}_s \\ \hat{k}_s \end{pmatrix} = \begin{bmatrix} \frac{a\,k\,(1+a\,t)}{t} & 0 & a^2k & 0 \\ 0 & v_{PM}(c;t) & 0 & \mathbb{D}_{2,4} \\ a^2k & 0 & \frac{a\,k\,(1+a\,s)}{s} & 0 \\ 0 & \mathbb{D}_{4,2} & 0 & v_{PM}(c;s) \end{bmatrix} \tag{15}$$

$$\mathbb{D}_{2,4} = \mathbb{D}_{4,2} = \frac{{r_t}^{k+1}{r_s}^{k+1}\,(r_t + r_s - 1)^{-k} - r_t\,r_s - (1-c)^2\,a^2t\,s\,k}{(r_t\,\log(r_t) - r_t + 1)\,(r_s\,\log(r_s) - r_s + 1)},$$

$$v_{PM}(c;u) = \frac{(1+au-auc^2)^{-k} r_u^{2k+2} - r_u^2 - kau(1+au)(1-c)^2}{[r_u\log(r_u) - r_u + 1]^2},$$

where $r_u = 1 + au(1-c)$. For the ZTM the matrix $\mathbb{D}$ can be computed using (15) with $c = 0$.

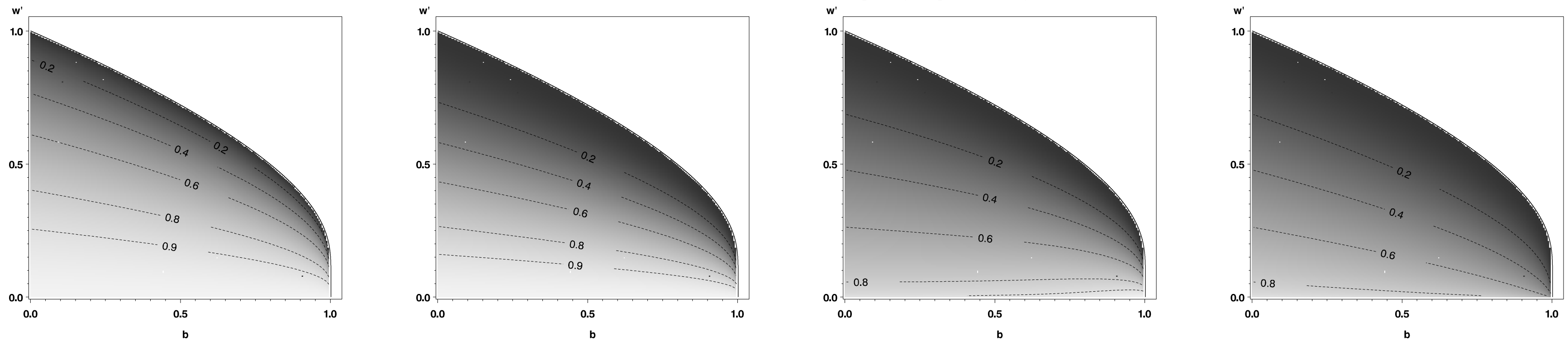
Fig. 2 and Fig. 3 show, for all valid NBD parameter values, correlations between $\rho(\hat{\theta}(t), \hat{\theta}(s))$, where $\hat{\theta}(t)$ and $\hat{\theta}(s)$ are different estimators of the same parameter θ^* computed using data in non-overlapping and overlapping time intervals. The correlations between estimators of the same parameter computed in different time intervals, for both estimators of m and k, increases as w increases for the MOM, PM and ZTM estimators.

4.4 Simulation results and application to market research data

Fig. 4 shows bivariate plots of various estimators $\hat{\theta}(t)$ and $\hat{\theta}(s)$ computed in different time intervals for 1000 replications of the gamma-Poisson process with sample size N=1000 with $m = 1$ and $k = 1$. A 95% confidence ellipse based on the covariance matrix (15) and constructed under the assumption of asymptotic normality is also shown. Figures are only shown for the case of non-overlapping intervals; this is the usual case encountered in practice. These figures confirm the validity of the expression (15).

Computing covariances between estimators in two different time intervals requires replications of estimators. In market research, data is usually collected in the form of panel data where realizations of the number of purchases of a product are observed for many households. Replications of estimators can be obtained by taking sub-samples of

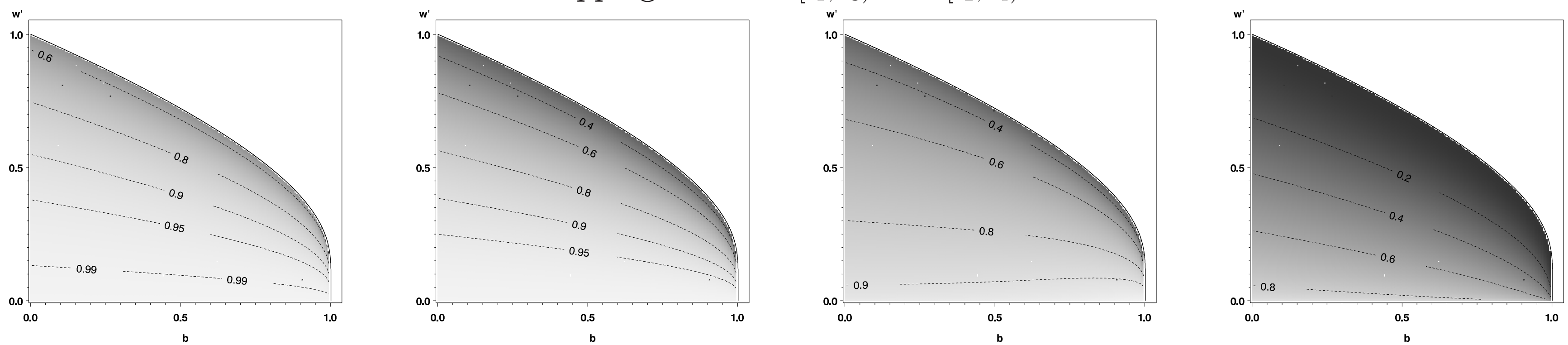
Correlations between estimators: Non-overlapping intervals $[0, t)$, $[t, t + s)$



(a) $\theta^* = m$ (b) $\theta^* = k$ (MOM) (c) $\theta^* = k$ (PM(c_*)) (d) $\theta^* = k$ (ZTM)

Figure 2: Correlation $\rho(\hat{\theta}(t), \hat{\theta}(s)) = \text{Cov}[\hat{\theta}(t), \hat{\theta}(s)]/\sqrt{\text{Var}[\hat{\theta}(t)]\text{Var}[\hat{\theta}(s)]}$, where $\hat{\theta}(t)$ and $\hat{\theta}(s)$ are different estimators of the same parameter θ^* computed using data in the time intervals $[0, t)$, $[t, t + s)$ respectively. Correlations are plotted for all NBD parameter values in the case $t = 1$ and $s = 1$ when sampling from the gamma-Poisson process.

Correlations between estimators: Overlapping intervals $[t_1, t_3)$ and $[t_2, t_4)$



(a) $\theta^* = m$ (b) $\theta^* = k$ (MOM) (c) $\theta^* = k$ (PM(c_*)) (d) $\theta^* = k$ (ZTM)

Figure 3: Correlation $\rho(\hat{\theta}(t), \hat{\theta}(s)) = \text{Cov}[\hat{\theta}(t), \hat{\theta}(s)]/\sqrt{\text{Var}[\hat{\theta}(t)]\text{Var}[\hat{\theta}(s)]}$, where $\hat{\theta}(t)$ and $\hat{\theta}(s)$ are different estimators of the same parameter θ^* computed using data in the time intervals $[t_1, t_3)$ and $[t_2, t_4)$ respectively. Correlations are plotted for all NBD parameter values in the case $t_1 = 0, t_2 = 1, t_3 = 2, t_4 = 3$ when sampling from the gamma-Poisson process.

the overall population and computing estimators for each sub-sample. We use panel data, kindly provided by the ACNielsen-BASES, which comprises 34500 households. We randomly split the whole panel into sub-samples of 500 households.

Note that the replications cannot be obtained by incrementing the time intervals and computing an estimator for each incremented time interval. The gamma-Poisson process is not an ergodic process and therefore the correlations between estimators obtained by considering different time intervals in a single realization are not equivalent to the correlations between estimators in the ensemble of realizations.

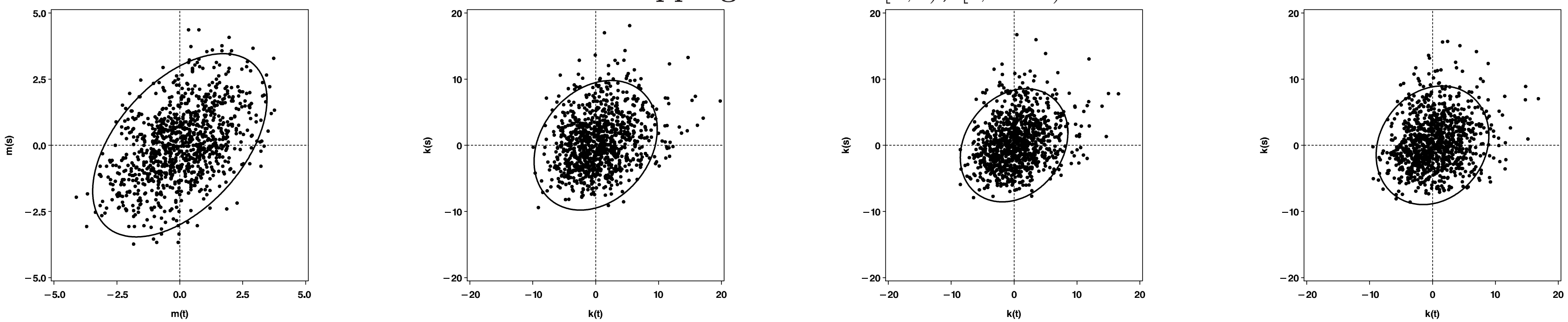
Fig. 5 shows normalized estimators of the gamma-Poisson parameters m and k computed in consecutive non-overlapping time intervals of length 26 weeks. Figures are shown for estimators of m computed in two different time intervals and estimators of k, using the MOM, PM and ZTM, computed in two different time intervals. In addition to the estimators a 95% confidence ellipses constructed using the covariance matrix (14) and (15) for estimators of m and k in non-overlapping time intervals are also shown. The values m and k required to construct the ellipses are replaced by the mean values of the estimators computed in both time intervals. The estimators for m and k are captured well by the 95% theoretical confidence ellipses.

These figures can be considered as a strong argument in favour of the suitability of the gamma-Poisson process for modelling the dynamical behaviour of purchasing in real markets.

Acknowledgement

We are grateful to the referee for carefull reading of the manuscript and a number of very constructive comments.

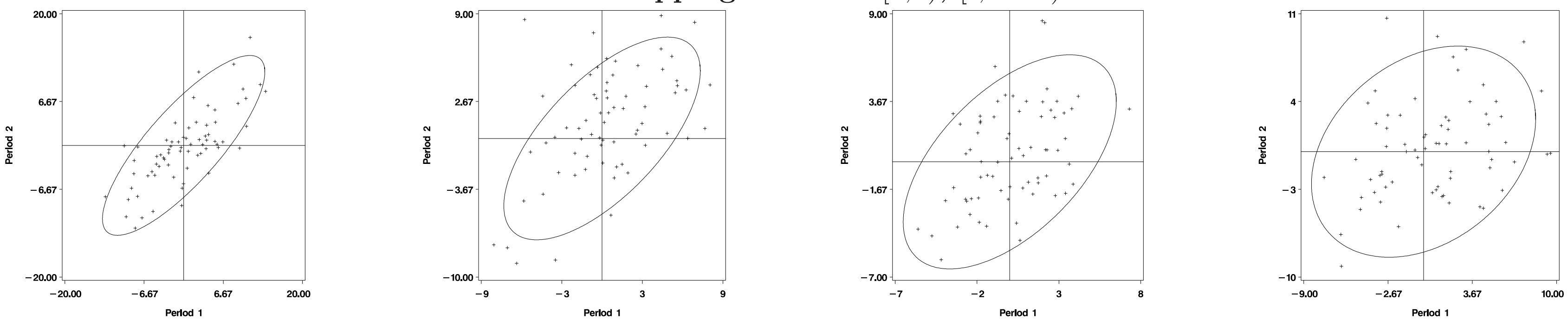
Joint distributions of Estimators: Non-overlapping intervals $[0, t)$, $[t, t+s)$ - Simulated data



(a) $\theta^* = m$ (b) $\theta^* = k$ (MOM) (c) $\theta^* = k$ (PM(c_*)) (d) $\theta^* = k$ (ZTM)

Figure 4: 1000 points of $\sqrt{N}(\hat{\theta}(t) - \theta^*(t))$ versus $\sqrt{N}(\hat{\theta}(s) - \theta^*(s))$ computed from data in the time intervals $[0, t)$ and $[t, t+s)$ respectively when sampling from the gamma-Poisson process with $m = 1, k = 1$ and samples of size $N = 1000$ in the case $t = 1, s = 1$. A 95% confidence ellipse based on the covariance matrix (11) and constructed under the assumption of asymptotic normality is also shown.

Joint distributions of Estimators: Non-overlapping intervals $[0, t)$, $[t, t+s)$ - Market Research Data



(a) $\theta^* = m$ (b) $\theta^* = k$ (MOM) (c) $\theta^* = k$ (PM(c_*)) (d) $\theta^* = k$ (ZTM)

Figure 5: Correlations between estimators when fitting the gamma-Poisson process to purchases of a product. Bivariate plots show normalized estimators computed in different time periods together with corresponding 95% confidence ellipses computed under the assumption of asymptotic normality. Here $t = s = 26$ weeks.

Références

Anscombe, F. J. (1950). Sampling theory of the negative binomial and logarithmic series distributions. *Biometrika*, *37*, 358–382.

Cook, R. J., & Wei, W. (2003). Conditional analysis of mixed poisson processes with baseline counts: implications for trial design and analysis. *Biostatistics*, *4*, 479–494.

Cox, D. R., & Hinkley, D. V. (1978). *Problems and solutions in theoretical statistics.* London: Chapman & Hall.

Ehrenberg, A. (1988). *Repeat-buying: Facts, theory and applications.* London: Charles Griffin & Company Ltd.; New York: Oxford University Press.

Fisher, R. A. (1941). The negative binomial distribution. *Ann. Eugenics*, *11*, 182–187.

Goodhardt, G., Ehrenberg, A., & Chatfield, C. (1984). The dirichlet: A comprehensive model of buying behaviour. *Journal of the Royal Statistical Society*, *A,147*, 621–655.

Grandell, J. (1997). *Mixed Poisson processes* (Vol. 77). London: Chapman & Hall.

Greenwood, M., & Yule, G. (1920). An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents. *Journal of the Royal Statistical Society*, *93*, 255–279.

Haldane, J. B. S. (1941). The fitting of binomial distributions. *Ann. Eugenics*, *11*, 179–181.

Lundberg, O. (1964). *On Random Processes and their Application to Sickness and Accident Statistics.* Uppsala: Almquist and Wiksells.

Savani, V., & Zhigljavsky, A. (2006a). Efficient estimation of parameters of the negative binomial distribution. *Communications in Statistics: Theory and Methods*, *35*(5), 767–783.

Savani, V., & Zhigljavsky, A. (2006b). Efficient parameter estimation for independent and INAR(1) negative binomial samples. *Metrika (in print).*

van der Vaart, A. W. (1998). *Asymptotic statistics* (Vol. 3). Cambridge: Cambridge University Press.