**George Tzougas**

# Confidence intervals of the premiums of optimal Bonus Malus Systems

## Article (Accepted version) (Refereed)

**Original citation:** Tzougas, George (2017) *Confidence intervals of the premiums of optimal Bonus Malus Systems.* Scandinavian Actuarial Journal . ISSN 0346-1238
DOI: 10.1080/03461238.2017.1307267

# Confidence Intervals of the Premiums of Optimal Bonus Malus Systems

Dimitris Karlis*
Department of Statistics, Athens University of Economics and Business
Greece

George Tzougas†
Department of Statistics, London School of Economics, UK

Nicholas Frangos
Department of Statistics, Athens University of Economics and Business
Greece

**Abstract**

In view of the economic importance of motor third party liability insurance in developed countries the construction of optimal BMS has been given considerable interest. However, a major drawback in the construction of optimal BMS is that they fail to account for the variability on premium calculations which are treated as point estimates. The present study addresses this issue. Specifically, nonparametric mixtures of Poisson laws are used to construct an optimal BMS with a finite number of classes. The mixing distribution is estimated by nonparametric maximum likelihood (NPML). The main contribution of this paper is the use of the NPML estimator for the construction of confidence intervals for the premium rates derived by updating the posterior mean claim frequency. Furthermore, we advance one step further by improving the performance of the confidence intervals based on a bootstrap procedure where the estimated mixture is used for resampling. The construction

*Athens University of Economics and Business, 76 Patission Str., 104 34, Athens, Greece, Tel: +30 210 8203 920; fax: +30 210 8203 920. E-mail address: karlis@aueb.gr (D. Karlis)

†Part of the work was done while Dr Tzougas was a postdoctoral researcher at Department of Statistics, Athens University of Economics and Business

of confidence intervals for the individual premiums based on the asymptotic maximum likelihood theory is beneficial for the insurance company as it can result in accurate and effective adjustments to the premium rating policies from a practical point of view.

**Keywords:** Optimal BMS; Nonparametric maximum likelihood; Asymptotic Normality; Wald type two-sided confidence intervals; Efron percentile bootstrap confidence intervals

# 1 Introduction

Bonus-Malus Systems, BMS in short, are experience rating mechanisms which impose penalties on policyholders responsible for one or more accidents by premium surcharges (or maluses) and reward discounts (or bonuses) to policyholders who had a claim-free year. Optimal BMS are financially balanced for the insurer, i.e. the total amount of bonuses must be equal to the total amount of maluses, and fair for the policyholder, i.e. the premium paid for each policyholder is proportional to the risk that they impose on the pool. The design of such systems is achieved through Bayesian analysis and the form of the mixed Poisson distributions which capture the unobserved heterogeneity of claim count data with respect to the simplistic Poisson law. Over the years numerous articles have been devoted to this topic as this is an area of applied statistics that has close ties with theoretical statistics, notably Bayesian Analysis, nonparametric maximum likelihood estimation, advanced regression models and credibility theory, which is the cornerstone of contemporary insurance mathematics. An excellent account of BMS can be found in Lemaire (1995). Also, references for BMS include, among others, Dionne and Vanasse (1989, 1992), Coene and Doray (1996), Walhin and Paris (1999), Pinquet (1998), Pinquet et al. (2001), Denuit and Lambert (2001), Brouhns et al. (2003), Denuit et al. (2007), Pitrebois et al. (2005), Boucher et al. (2008), Tzougas and Frangos (2014) and Tzougas et al. (2014).

However, even though the construction of optimal BMS has been a basic interest of actuarial literature for over four decades, scientific attention has only now focused on deriving credibility updates of the claim frequency based on the employment of an abundance of alternative parametric distributions, nonparametric distributions and advanced regression models. In this respect, a major drawback in the design of such systems was neglected: namely the fact that they do not give a measure of uncertainty of the resulting premium estimates by providing a confidence interval that contains plausible values.

In a competitive insurance market, in order to avoid lapses, actuaries do not only have to construct optimal BMS that will fairly distribute the burden of claims among policyholders, as was the usual practice, but their designs also have to be able to adjust the individual premiums from a practical point of view . Moreover, taking into account that according to a 2015 report by Insurance Europe (Insurance Europe, 2015), an insurance and reinsurance federation with 34 member bodies, the largest non-life insurance market, motor insurance totaled 130.8bn Euros in premiums (stable in 2014), it becomes clear that the problem briefly described above can result in great losses for insurance companies operating in Europe.

Let us now explain how the present study addresses the aforementioned problem. In most settings involving count data, one of the biggest challenges that a researcher can come across is reliably estimating or building confidence intervals, CIs, for small and tail probabilities. In the majority of cases the available data are either insufficient to allow for asymptotic arguments or they need to be smoothed to render them useful. In motor third party liability (MTPL) insurance, the interest of actuaries lies in identifying customers with high claim frequency but they normally represent very few observations. A simple and intuitive approach could be to resort to the use of the empirical proportion as an estimate of the event probability. However, a serious drawback of this method is the heavy data requirement. That is, if the event is not observed with sufficient frequency, tail probabilities cannot be estimated with accuracy. Therefore, smoother estimates for tail probabilities are demanded in order to produce useful results. As a solution to the aforementioned problem, one could consider a model where the small probabilities are connected to other parts of the probability distribution. However, in this case inference is vulnerable to model assumptions.

Karlis and Patilea (2008) proposed a satisfactory trade-off between the flexibility of that model which guards against misspecification and the ability to provide non-degenerated estimates with finite samples. Specifically, following Böhning and Patilea (2005), these authors considered nonparametric mixtures of power series distributions and built CIs for small probabilities with count data based on the use of the nonparametric maximum likelihood estimator, NPMLE, of the mixing distribution. Also, they constructed bootstrap two-sided confidence intervals based on a bootstrap from the NPMLE of the mixture. Furthermore, Karlis and Patilea (2007) constructed NPMLE and bootstrap based CIs for the hazard rate function of the discrete lifetime distribution.

In this paper we extend BMS literature research by addressing the problem of building confidence intervals for the premiums determined by an optimal BMS in the following ways.

- Firstly, following Walhin and Paris (1999) and Denuit and Lambert (2001), we consider a flexible class of nonparametric mixtures of Poisson distributions for assessing claim frequency. An algorithm which is a variant of the EM algorithm adjusted for jumping between different numbers of components is proposed in order to estimate the unknown mixing, or risk, distribution based on nonparametric maximum likelihood estimation. The use of the nonparametric estimate of the risk distribution allows for a rich family of claim frequency distributions instead of restricting attention to particular laws such as the negative binomial distribution that has been widely applied for modelling claim count data. On the path toward actuarial relevance the Bayesian view is taken and the NPMLE of the risk distribution is used to calculate premiums as posterior means. Following Lambert and Tierney (1984) and Böhning and Patilea (2005), it is shown that the NPMLE based posterior mean claim frequency behaves asymptotically normal. Based on the asymptotic normality of the posterior mean claim frequency Wald type two-sided confidence intervals are constructed. The Wald CIs are not degenerated and therefore are more useful than the corresponding intervals based on model analogy or ad hoc reasoning.

- Secondly, we develop bootstrap two-sided confidence intervals for the individual premiums based on bootstrap from the NPMLE of the mixing distribution. This NPMLE based resampling procedure is a common method encountered in the literature, see for example Laird and Louis (1987) and Böhning (2000). Refer also to Karlis and Patilea (2008) for the proof of its asymptotic validity. Specifically, Efron percentile bootstrap confidence intervals are investigated and compared to the Wald Type confidence intervals obtained directly from the NPML estimates. Our analysis reveals that Efron percentile bootstrap intervals on certain occasions improve the asymptotic normal approximation used by Wald intervals. The aforementioned constructions of NPMLE and bootstrap based CIs account for the uncertainty as well as the fluctuations of the individual premium estimates.

In an experience ratemaking scheme the use of such intervals leaves room for the informed judgment of the actuary to select the final premiums to be charged to each policyholder based on the fluctuations that occur equally on either side of the credibility updates of their claim frequency. In this respect, the insurance company can be responsive to the needs of different constituencies, such as broader economic trends for the insurance market in which it operates or mounting regulatory requirements, in order to make more accurate and effective adjustments to the tariff from a practical point of view.

The rest of the paper is as follows. Section 2 presents the general background on mixtures of Poisson distributions. Section 3 provides the computational details for the algorithm used for the NPMLE. Section 4 describes the design of an optimal BMS with a finite number of classes based on the NPMLE of the risk distribution. Section 5 provides the main results for the NPMLE based intervals and the bootstrap intervals respectively. Section 6 contains an application to a data set concerning car-insurance claims at fault. Finally, Section 7 presents the concluding remarks of the paper.

# 2 Mixtures of Poisson Distributions

Let us consider a Poisson mixture with probability mass function (pmf) given by

$$P(x; F_\Lambda) = \pi_{F_\Lambda}(k) = \int_\Lambda P(x;\lambda) F_\Lambda(d\lambda), \tag{1}$$

for $k \in \mathbb{N}$, where $P(x;\lambda)$ is the probability distribution function of the Poisson distribution and where $F_\Lambda$ is the mixing distribution, that is a probability measure on $\Lambda$, whose support is $\mathbb{R}^+$. Assume that the independent observations distributed according to the mixture $\pi_{F_{\Lambda_0}}$ with individual probabilities $\pi_{F_{\Lambda_0}}(k), k \in \mathbb{N}$. The true mixing distribution $F_{\Lambda_0}$ is unknown but its support is included in a known compact interval $[0, M] \subset \mathbb{R}^+$. In practice one can choose $M$ to arbitrarily large. It is quite typical to assume a certain parametric form for $F_{\Lambda_0}(\cdot)$ and fit a parametric model. However, to gain more flexibility we prefer not to assume any parametric form for the mixing distribution and leave $F_{\Lambda_0}$ to be a general mixing distribution. There are a wide range of practical applications for this type of model, as for example, population heterogeneity studies, non-parametric empirical Bayes estimation and semiparametric density estimation; see Lindsay (1995), Lindsay and Lesperance (1995), Böhning (2000) and the references therein.

By definition, $F_{\Lambda_0}$ is identifiable if $F_{\Lambda_0} = F_\Lambda$ implies that $\Lambda_0 = \Lambda$. Lambert and Tierney (1984) and Böhning and Patilea (2005) showed that, because

$\sum_{k>0, k\in\mathbb{N}} k^{-1} = \infty$ holds, $F_{\Lambda_0}$ is identifiable among all the mixing distributions with the support in $\Lambda$.

In this study, we estimate $F_{\Lambda_0}$ in a nonparametric way and then use the Poisson mixture for constructing an optimal BMS. Let $X_1, \ldots, X_n \in \mathbb{N}$ be an i.i.d. sample with distribution $\pi_{F_{\Lambda_0}}$. The log-likelihood function (denoted as a function of the mixing distribution) is

$$\ell(F_\Lambda) = \sum_{i=1}^{n} \log\left\{\int_\Lambda P(x_i;\lambda) dF_\Lambda(\lambda)\right\}. \tag{2}$$

We want to maximize $\ell(\cdot)$ with respect to all distribution functions defined on $\Lambda$, this is called the nonparametric maximum likelihood estimator (NPMLE) and it is known to be a distribution with discrete support, i.e. giving positive probability to a finite number of points.

Hereafter, let $\hat{F}_\Lambda$ be the NPMLE of $F_{\Lambda_0}$.There are results on the maximum number of support points $\hat{q}$ (see Simar, 1976 and Lindsay, 1983), which cannot exceed the number of distinct values in the sample. Specifically, Simar (1976) was the first to show that the NPMLE will be unique under the following condition

$$\hat{q} \leq \min\left(\left[\frac{k_{\max}+1}{2}\right], \kappa\right),$$

where $k_{\max}$ is the maximum number of claims per risk and $\kappa$ is the number of classes for with non-zero frequency. Furthermore, existence, support size, and other finite sample properties of $\hat{F}_\Lambda$ can be found in Simar (1976) and Lindsay (1995). Concerning consistency, with probability one $\hat{F}_\Lambda \to F_{\Lambda_0}$ weakly, since $F_{\Lambda_0}$ is identifiable (see for instance, van de Geer, 1993, Lemma 5.2). Furthermore, existence, support size, and other finite sample properties of $\hat{F}_\Lambda$ can be found in Simar (1976) and Lindsay (1995). Since the NPMLE $\hat{F}_\Lambda$ is discrete the model resembles the finite mixture model.

Methods like the widely used EM algorithm could be used towards the derivation the NPMLE. Lambert and Tierney (1984) showed the asymptotic normality of the NPMLE for Poisson mixtures while Böhning and Patilea (2005) showed the asymptotic normality of the NPMLE for mixture of power series family (and hence for the Poisson case since it is a member of the power series family of distributions). Karlis and Patilea (2007, 2008) showed the consistency in probability of bootstrap confidence intervals and they applied this to the case of hazard function which is related to what follows here, since it involves ratio of probabilities as we will do for the BMS case. Being able to derive such confidence intervals for the BMS we are able to account for the uncertainty around the premium calculated.

# 3 Computational Details

In this section we describe the algorithm used to derive the NPMLE. Algorithms for finding the NPMLE has been proposed by Laird (1978), Dersimonian (1986), Lesperance and Kalbfleisch (1992)(see also the book of Böhning (2000) for a broad review on these algorithms). Recent work can be found in Wang (2007). They make use of the gradient function in order to decide where to add new supports points and which one can be removed. The gradient function is defined as

$$d(\lambda; F_\Lambda) = \sum_{i=1}^{n} \frac{P(x_i; \lambda)}{P(x_i; F_\Lambda)} - n \tag{3}$$

For the NPMLE it holds that $\sup_\lambda d(\lambda, \hat{F}_\Lambda) = 0$ (see Lindsay, 1995) and this provides a diagnostic whether the NPMLE has been found. Alternatively one may use algorithms for fixed number of support points $k$ for different values of $k$. These algorithms are feasible for count data because the number of support points in the NPMLE is usually small (see the results of Lindsay, 1995).

The algorithm used in the present paper for finding the NPMLE is a variant of the EM algorithm adjusted for jumping between different numbers of components. Namely the algorithm starts with the maximum possible number of components (see Simar, 1976 and Lindsay, 1983). Then we keep iterating using the EM algorithm until either satisfaction of the convergence criterion (measured by the change of the relative likelihood) or until a redundant support points is found. A support point is redundant either if a) two points are close together or b) one mixing proportion is close to 0. Two components with parameters, say $\lambda_j$ and $\lambda_k$ are considered as being close together if $|\lambda_j - \lambda_k| < 10^{-6}$. If this is the case, then, we check if combining these components in a single component with value the weighted average of the two components and mixing proportion the sum of the two proportions, we can improve the likelihood. If the likelihood can be improved, the components are merged, otherwise we keep iterating retaining both the components. The idea for this step is that if the components are close together then this implies either that the components must be merged or that the likelihood will remain trapped in this area for a long time. If the second is true our algorithm can fail, but in any case we will not waste our time waiting to pass over the flat point. Note that our experience was that typically two point so close should be merged. A mixing proportion is close to 0 if its value is smaller than a threshold like $10^{-6}$. In the case of a small mixing proportions we just remove this point by rescaling the other mixing proportions to sum to 1. Note that, very small mixing proportions are expected only for large sample sizes.

When the algorithm converges (i.e. the relative change of the log-likelihood is smaller than $10^{-12}$, we check whether the NPMLE has been found by using the conditions given in Lindsay (1995). These conditions were based on the gradient function defined in (3). We calculated the gradient function over a grid of 1000 points in a large interval from 0 to $1.2\lambda_{max}$, where $\lambda_{max}$ is she largest support point and we checked whether for all the points the gradient was less than 0.0001. If the solution was not truly a NPMLE (i.e. the function lies above zero) then we rerun the EM algorithm described above from different initial values.

For every repetition, initial values were chosen randomly over the interval of admissible values. For each sample 20 different initial values were considered. If the NPMLE was not found after 20 runs then we rejected this sample. The rate of rejecting samples was smaller than 2% for the Poisson case. Note also that, since we are not interested about reporting the number of support points, redundant points in the NPMLE do not cause any problem since the probabilities estimated by the NPMLE will coincide. Our algorithm is similar to running an EM with fixed support size equal to the maximum possible for each sample. Our algorithm improved on this approach by reducing the dimensionality between iterations and thus removing redundant calculations at each iteration.

A step by step description of the algorithm follows. Technical details are not repeated.

Step 0: Start with $k$ support points. Choose initial values for the parameters.

Step 1: Run a number of EM iterations, say $M$ ($M$ can be one but usually a larger value improves speed)

Step 2: Check if there are redundant support points: i.e. points, with $\lambda$ close together or mixing proportion close to 0.

Step 3a: If redundant points are found then merge them (or discard the one with a almost zero mixing proportion).

Step 3b: If the loglikelihood after merging is improved then keep going with the merged components and go back to Step 1, else keep going with the same number of components.

Step 4: Check if convergence is detected and stop otherwise go back to Step 1.

Step 5: Use the gradient function to ensure that the NPMLE is found. If not use other initial values and go back to Step 0.

# 4 The Design of an Optimal Bonus-Malus System

We assume that all policyholders have constant but unequal underlying risks of having an accident. Consider a policyholder and denote by $N_j$ the number of claims in which they were at fault during the $j$th year that the policy was in force. We assume that the claim frequency does not change over time and that $N_j$ are independent and identically distributed (i.i.d) random variables according to a mixed Poisson process with mass function given by

$$P(N_j = k) = \pi_{F_{\Lambda_0}}(k) = \int_{\lambda \in \mathbb{R}^+} \frac{e^{-\lambda}\lambda^k}{k!} F_{\Lambda_0}(d\lambda), \tag{4}$$

where $k \in \mathbb{N}$ and $\lambda$ is the observed value of a random variable $\Lambda$ whose support is $\mathbb{R}^+$ and where $F_{\Lambda_0}$ is the mixing distribution, called the structure function, which, as we have previously mentioned, is unknown but its support is included in a known compact interval $[0, M] \subset \mathbb{R}^+$. Depending on the chosen form of the mixing distribution, (4) will lead to different models. Two kinds of models can be distinguished in actuarial literature for the choice of the structure function, the parametric and nonparametric cases. The former consists of families where $F_{\Lambda_0}$ is approximated by some well known parametric distribution and the latter consists of choosing to estimate $F_{\Lambda_0}$ nonparametrically. Firstly, with respect to the parametric case, a traditional choice for the distribution of $\lambda$ values among all policyholders is the gamma distribution which gives the negative binomial distribution, see for instance, Lemaire (1995). Alternative choices are the inverse Gaussian (see Willmot, 1987 and Tremblay, 1963) and the generalized inverse Gaussian (see Tzougas and Frangos, 2014), which result in the Poisson-inverse Gaussian and Sichel laws respectively, and Hoffman's distributions (see Kestemont and Paris, 1985 and Walhin and Paris, 1999). The structure function can also be a finite point discrete distribution. In this case the portfolio heterogeneity is accounted for by choosing a finite number of unobserved latent components, each of which may be regarded as a sub-population, and the unconditional distribution of the number of claims in (4) can be regarded as a finite mixture of count distributions. In BMS literature research Lemaire (1995) considered the good risk/bad risk model employing a two component Poisson mixture distribution for the number of claims. Tzougas et al. (2014) focused on modelling claim frequency as a finite Poisson, Delaporte and negative binomial mixture respectively. Secondly, with respect to the nonparametric case, the interested reader can refer to Walhin and Paris (1999) and Denuit and Lambert (2001) who both resort to nonparametric estimators for the mixing distribution.

In the setup we adopt, as described in Section 2, $\hat{F}_{\Lambda}$ will be attained for a discrete distribution function $F_{\Lambda_0}$ with a maximum number $\hat{q}$ of support points that maximize the log-likelihood. Then, the NPMLE of $\pi_{F_{\Lambda_0}}(k)$ is the mixture $\hat{\pi}(k) = \pi_{\hat{F}_{\Lambda_0}}(k)$ given by

$$\hat{\pi}(k) = \sum_{z=1}^{\hat{q}} \hat{p}_z \frac{e^{-\hat{\lambda}_z} \hat{\lambda}_z^k}{k!}, \tag{5}$$

for $k \in \mathbb{N}$, where $p_q > 0$ and where $\sum\limits_{z=1}^{\hat{q}} \hat{p}_z = 1$. (5) gives the pmf of a finite Poisson mixture model with mean and variance equal to

$$E(N_j) = \sum_{z=1}^{\hat{q}} \hat{p}_z \hat{\lambda}_z \text{ and } V(N_j) = \sum_{z=1}^{\hat{q}} \hat{p}_z \hat{\lambda}_z + \sum_{z=1}^{\hat{q}} \hat{p}_z \hat{\lambda}_z^2 - \left( \sum_{z=1}^{\hat{q}} \hat{p}_z \hat{\lambda}_z \right)^2 .$$

In this respect, population heterogeneity is accounted for by choosing a finite number of $\hat{q}$ categories of policyholders classified according to their driving skills.

Let us now present the optimal BMS determined by the finite Poisson mixture. Consider a policyholder who is observed for $t$ years of their presence in the portfolio and has claim frequency history $N_1, ..., N_t$. Given $N_1 = k_1, ..., N_t = k_t$, denote by $K = \sum\limits_{j=1}^{t} k_j$ the total number of claims that the policyholder had in $t$ years. The problem is to determine, at the renewal of the policy, the premium that must be charged to the policyholder for the period $t+1$ given the observation of their reported accidents in the preceding $t$ periods, i.e. to determine the posterior mean, denoted by $\lambda_{t+1}(K)$. By means of the Bayes theorem and using the quadratic error loss function we have that (also see Walhin and Paris, 1999 and Denuit and Lambert, 2001)

$$\begin{aligned}
\lambda_{t+1}(K) &= E(\Lambda | N_1 = k_1, ..., N_t = k_t) \\
&= \frac{\int_{\lambda\in\mathbb{R}^+} \lambda \prod_{i=1}^{t} \frac{e^{-\lambda}\lambda^{k_i}}{k_i!} F_{\Lambda_0}(d\lambda)}{\int_{\theta\in\mathbb{R}^+} \prod_{i=1}^{t} \frac{e^{-\theta}\theta^{k_i}}{k_i!} F_{\Lambda_0}(d\theta)} \\
&= \frac{\int_{\lambda\in\mathbb{R}^+} e^{-\lambda t}\lambda^{K+1} F_{\Lambda_0}(d\lambda)}{\int_{\theta\in\mathbb{R}^+} e^{-\theta t}\theta^{K} F_{\Lambda_0}(d\theta)} \\
&= \frac{(K+1)}{t}\frac{\pi_{F_{\Lambda_0}}(K+1)}{\pi_{F_{\Lambda_0}}(K)}. \qquad (6)
\end{aligned}$$

It is interesting to note that $\lambda_{t+1}(K)$ only depends on the total number of claims $K$ caused during the past $t$ years that the policy was in force and not on past claim history records.

After $t$ years of coverage and given $N_1 = k_1, ..., N_t = k_t$ (5) becomes

$$\hat{\pi}(K) = \sum_{z=1}^{\hat{q}} \hat{p}_z \frac{e^{-\hat{\lambda}_z t}\left(t\hat{\lambda}_z\right)^K}{K!}. \qquad (7)$$

Based on (7) we estimate $\lambda_{t+1}(K)$ by

$$\hat{\lambda}_{t+1}(K) = \frac{(K+1)}{t}\frac{\hat{\pi}(K+1)}{\hat{\pi}(K)} = \frac{\sum_{z=1}^{\hat{q}} \hat{p}_z e^{-\hat{\lambda}_z t}\hat{\lambda}_z^{K+1}}{\sum_{z=1}^{\hat{q}} \hat{p}_z e^{-\hat{\lambda}_z t}\hat{\lambda}_z^{K}}. \qquad (8)$$

Let us call $\hat{\lambda}_{t+1}(K)$ the NPMLE of $\lambda_{t+1}(K)$. When $t = 0$, $\hat{\lambda}_{t+1}(K)$ reduces to $\hat{\lambda}_1(0) = E(\Lambda) = \sum_{z=1}^{\hat{q}} \hat{p}_z\hat{\lambda}_z$ since there is no information concerning the policyholder.

# 5 Confidence Intervals

In this Section, using the NPML estimator given by (5) and based on the framework developed by Karlis and Patilea (2008), we are going to build Wald Type confidence intervals and Efron percentile bootstrap confidence intervals for $\lambda_{t+1}$.

## 5.1 Wald Type Confidence Intervals

The asymptotic distribution of the NPMLE $\hat{\lambda}_{t+1}(K)$ will be described by the following proposition. The proof can be seen in Appendix A.

**Proposition 1** *Assume that the support of* $F_{\Lambda_0}$ *is contained in some known* $[0, M] \subset \mathbb{R}^+$, *i.e. the support of* $\Lambda$. *Moreover,* $F_{\Lambda_0}$ *is identifiable since* $\sum_{k>0,k\in\mathbb{N}} k^{-1} = \infty$ *(see Lambert and Tierney, 1984, and Böhning and Patilea, 2005). Assume also that there exist positive constants* $d, \gamma, \varepsilon$ *such that* $F_{\Lambda_0}((\lambda, \lambda + \tau])) \geq d\tau^{\gamma}$ *for all* $\lambda, \tau \in (0, \varepsilon)$.

Then for each $K \in \mathbb{N}$ we have that

$$\sqrt{n}\left(\hat{\lambda}_{t+1}(K) - \lambda_{t+1}(K)\right) \Longrightarrow N(0, V_{t+1}(K)), \tag{9}$$

where $K = \sum_{j=1}^{t} k_j$ is the total number of claims after $t$ years of coverage, $n$ is the sample size, i.e. the total number of insureds, $\hat{\lambda}_{t+1}(K)$ is the estimate of $\lambda_{t+1}(K)$ yielded by $\hat{F}_\Lambda$ the NPMLE of $F_\Lambda$ and where, denoting by $V_{t+1}(K)$ the variance of $\hat{\lambda}_{t+1}(K)$,

$$V_{t+1}(K) = \left(\frac{K+1}{t}\right)^2 \left[\frac{\pi_{F_{\Lambda_0}}(K+1)}{\pi^2_{F_{\Lambda_0}}(K)}\right]\left[1 + \frac{\pi_{F_{\Lambda_0}}(K+1)}{\pi_{F_{\Lambda_0}}(K)}\right]. \tag{10}$$

Based on the asymptotic normality of $\hat{\lambda}_{t+1}(K)$, consider the Wald type two-sided confidence interval (CI)

$$\left[\hat{\lambda}_{t+1}(K) - \frac{z_{1-\alpha/2}}{\sqrt{n}}\sqrt{\hat{V}_{t+1}(K)}, \hat{\lambda}_{t+1}(K) + \frac{z_{\alpha/2}}{\sqrt{n}}\sqrt{\hat{V}_{t+1}(K)}\right] \tag{11}$$

with $K$, some fixed value in $\mathbb{N}$, $z_\alpha$ the $\alpha$ quantile of the standard normal distribution and

$$\begin{aligned}
\hat{V}_{t+1}(K) &= \left(\frac{K+1}{t}\right)^2 \left[\frac{\hat{\pi}(K+1)}{\hat{\pi}^2(K)}\right]\left[1 + \frac{\hat{\pi}(K+1)}{\hat{\pi}(K)}\right] \\
&= \left[\frac{K!}{t^K}\frac{\sum_{z=1}^{\hat{q}} \hat{p}_z e^{-\hat{\lambda}_z t}\hat{\lambda}_z^{K+1}}{\left(\sum_{z=1}^{\hat{q}} \hat{p}_z e^{-\hat{\lambda}_z t}\hat{\lambda}_z^{K}\right)^2}\right]\left[\frac{K+1}{t} + \frac{\sum_{z=1}^{\hat{q}} \hat{p}_z e^{-\hat{\lambda}_z t}\hat{\lambda}_z^{K+1}}{\sum_{z=1}^{\hat{q}} \hat{p}_z e^{-\hat{\lambda}_z t}\hat{\lambda}_z^{K}}\right].
\end{aligned}$$

When $t=0$, $\hat{V}_{t+1}(K)$ reduces to $\hat{V}_1(0)=V(\Lambda)=\sum_{z=1}^{\hat{q}}\hat{p}_z\hat{\lambda}_z+\sum_{z=1}^{\hat{q}}\hat{p}_z\hat{\lambda}_z^2-\left(\sum_{z=1}^{\hat{q}}\hat{p}_z\hat{\lambda}_z\right)^2$ since there is no information concerning the risk. The asymptotic consistency of this Wald type CI at level 1-$\alpha$ is ensured by Proposition 1.

## 5.2 Efron Percentile Bootstrap Confidence Intervals

Consider a bootstrap procedure where the bootstrap samples $N_{j,1}^*, ..., N_{j,n}^*$ are the number of claims of a policyholder $i, i=1,..,n$, during the $j$th year of their presence in the portfolio generated according to the finite Poisson mixture given by (5). This is a parametric bootstrap procedure where the unknown parameter is the mixing distribution and the parameter space is the set of all probability measures on $[0, M]$, that is, the parameter space is of infinite dimension. The unknown parameter is estimated by nonparametric maximum likelihood. See Karlis and Patilea (2008) for the proof of its asymptotic validity.

Let $\hat{\pi}^*(K)$ and $\hat{\lambda}_{t+1}^*(K)$ be the NPML estimators of the individual probabilities and the premium at $t+1$ respectively, obtained from a bootstrap sample, where $K$ is the total number of accidents caused after $t$ years of insurance. Like for computing $\hat{\pi}$, the NPMLE $\hat{\pi}^*$ is obtained from nonparametric maximum likelihood over the mixing distributions with support in $[0, M]$. In what follows the Efron percentile bootstrap CI will be considered (see Efron, 1982). For $\alpha \in (0,1)$, we denote by $\zeta_\alpha$ the smallest value $z$ that satisfies the inequality

$$P\left(\hat{\lambda}_{t+1}^*(K) \leq z|\hat{\pi}\right) \geq \alpha. \tag{12}$$

The notation $P(\cdot|\hat{\pi})$ indicates that the distribution of $\hat{\lambda}_{t+1}^*(K)$ must be evaluated assuming that the bootstrap observations are sampled according to $\hat{\pi}(K)$ given the original data $N_{j,1}, ..., N_{j,n}$ (in particular, $\hat{\lambda}_{t+1}(K)$ is considered nonrandom). The Efron percentile is given by

$$\left[\hat{\zeta}_{\alpha/2}, \hat{\zeta}_{1-\alpha/2}\right]. \tag{13}$$

The results presented in Karlis and Patilea (2008) combined with the delta-method for bootstrap "in probability" (see, for instance, van der Vaart, 1998, Section 23.2), yield the asymptotic consistency of the Efron bootstrap percentile CI at level $1-\alpha$. The proof can be found in Appendix B.

## 5.3 Discussion about the intervals

In this section we discussed two alternative ways to construct confidence intervals. It is important to note that perhaps more sophisticated approach could be used for such intervals, like improved bootstrap based intervals, at the cost of added complexity both from computational point of view but also from practicality aspect. Some comments on the derived intervals can be useful for the practitioners.

Wald type intervals are based on the asymptotic normality and hence the intervals are based on the normal distribution. For actuarial applications, typically we have reasonable sample sizes to base asymptotic arguments, however the normality assumption in some cases need to be tested. Wald type intervals may suffer from lower limits for the mean in the non admissible range (e.g. negative values) since they are typically based on a point estimate plus/minus some quantity. Also previous simulations in a relevant problem (see, Karlis and Patilea 2007) showed that they have large length and are somewhat unstable especially where not enough data are available as it can be the case at the tails of the data.

On the other hand, Efron percentile bootstrap intervals are smoother at the cost of additional computational effort. They will never provide limits in the non admissible range and in general behave better (e.g. smaller length) than Wald type intervals. Bootstrap based intervals needs more computing and hence can be more demanding in practice.

# 6 Application

## 6.1 About the data and their NPMLE

The data were kindly provided by a Greek insurance company and concern a motor third party liability (MTPL) insurance portfolio observed during 3.5 years. The data set comprises $n = 15641$ policies. Claim counts are modelled for all 15641 policies that have been in force for the entire sampling period. The expected frequency of claims at fault is 0.4848 and the variance is 0.7308.

We assume that the claim frequency data follow a Poisson mixture distribution with pdf given by (1). The unknown mixing distribution $F_{\Lambda_0}$ was estimated by the NPMLE. Algorithmic details are provided in Section 3. For our portfolio, the NPMLE $\hat{F}_\Lambda$ was found to have $\hat{q} = 4$ support points leading to a four component Poisson mixture model with

$$\begin{bmatrix} \hat{p}_1, & \hat{p}_2, & \hat{p}_3, & \hat{p}_4 \\ \hat{\lambda}_1, & \hat{\lambda}_2, & \hat{\lambda}_3, & \hat{\lambda}_4 \end{bmatrix} = \begin{bmatrix} 0.15354, & 0.68401, & 0.16039, & 0.002040 \\ 0, & 0.369133, & 1.36139, & 6.80928 \end{bmatrix},$$

where the first and second line contain the estimated mixing proportions and mixture components respectively. The four component Poisson mixture is a generalization of the good risk/bad risk model proposed by Lemaire (1995) since it gives the maximum number of support points that maximize the log-likelihood, instead of two. The gradient function of the NPMLE can be seen in Figure 1. We have plot the plot in two in order to be able to examine the case. The right figure concentrates in a smaller interval and makes obvious the behavior of the gradient at the support points (denoted by dotted vertical lines)

Table 1 reports the observed frequency, the relative frequency and the expected probabilities based on the NPMLE. Then we report based on the methodology described the 95% confidence intervals for the probabilities. The fit as expected is quite close. Finally since the data refer to a 3.5 years period we report also annualized probabilities since these are used for derived the BMS.

Let us now present the optimal BMS resulting from the four component Poisson mixture model. The NPMLE for this model led to a heterogeneous portfolio. There is one group which has a zero rate, also there is another group with very large rate (6.809), which however is only the 0.2% of the portfolio. The premiums that must be paid for various number of claims when the age of the policy is up to t=5 years will be determined by (8) and are presented in Table 2. From Table 2 we see that this optimal BMS is fair since if the policyholder has a claim free year the premium is reduced, while if the policyholder has one or more claims the premium is increased,resulting in bonus or malus respectively. Furthermore, we notice that this system can be considered generous with good risks and strict with bad risks.

Figure1a.pdf Figure1b.pdf

Figure 1: The gradient function for the data. The left plot shows the entire range while the right plot focuses on a smaller interval so as to provide a better picture

Table 1: Observed data and fitted probabilities with the associated 95% confidence intervals derived from the NPMLE using the quantile method.The observed data refer to a period of 3.5 years, so we also report annualized probabilities at the last 3 columns

| | | | | | | Annualized | | |
|---|---|---|---|---|---|---|---|---|
| | | | | 95% conf. int. | | | 95% conf. int. | |
| $x$ | observed | rel. freq | NPMLE | LL | UL | mean | LL | UL |
| 0 | 10441 | 0.667540 | 0.667540 | 0.645164 | 0.685471 | 0.878005 | 0.872903 | 0.883322 |
| 1 | 3604 | 0.230420 | 0.230541 | 0.203722 | 0.267963 | 0.107955 | 0.102907 | 0.113005 |
| 2 | 1108 | 0.070839 | 0.070366 | 0.054538 | 0.084995 | 0.012038 | 0.010155 | 0.013724 |
| 3 | 321 | 0.020523 | 0.021372 | 0.012489 | 0.026335 | 0.001640 | 0.001142 | 0.002166 |
| 4 | 109 | 0.006969 | 0.006452 | 0.003759 | 0.009697 | 0.000281 | 0.000096 | 0.000500 |
| 5 | 34 | 0.002174 | 0.001904 | 0.000798 | 0.004446 | 0.000059 | 0.000006 | 0.000157 |
| ≥6 | 24 | 0.001534 | 0.001518 | 0.000319 | 0.004331 | 0.000022 | 0.000001 | 0.000105 |

Table 2: Optimal BMS with NPMLE Model

| Number of Claims | | | | | | | |
|---|---|---|---|---|---|---|---|
| Year | | | | $k$ | | | |
| $t$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| 0 | 0.1385 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 1 | 0.1227 | 0.2264 | 0.3798 | 0.7275 | 1.3502 | 1.7784 | 1.9092 |
| 2 | 0.1116 | 0.2004 | 0.3030 | 0.4444 | 0.7757 | 1.3786 | 1.7870 |
| 3 | 0.1026 | 0.1822 | 0.2719 | 0.3616 | 0.4783 | 0.8046 | 1.4008 |
| 4 | 0.0951 | 0.1672 | 0.2496 | 0.3337 | 0.3928 | 0.4955 | 0.8259 |
| 5 | 0.0888 | 0.1546 | 0.2292 | 0.3161 | 0.3691 | 0.4067 | 0.5057 |

## 6.2 Confidence Intervals

We are interested in building confidence intervals for the premiums that must be paid by a policyholder who is observed for 5 years and whose number of claims range from 1 to 6. Firstly, Wald type two-sided intervals based on NPMLE are constructed according to (11) and are presented in Table 3. The NPMLE based approach provides smooth estimates of the posterior mean claim frequency leading to intervals of reasonable length. Nevertheless, the lower bounds of the intervals defined in (11) might be in some cases negative as they rely on the asymptotic standard deviation estimates. On these occasions, NPMLE based CIs lie outside the admissible range since the premium rates must always be positive. In our application, negative lower CI bounds were observed for $(t = 1, k = 6)$. For this purpose, in Table 3 this value is replaced by zero. Secondly, given our data, we generated B = 1000 bootstrap samples $N_{t,1}^{*}, ..., N_{t,n}^{*}$ of size 15641 from $\hat{\pi}(k)$ in order to construct the Efron percentile bootstrap intervals based on (13). The results are presented in Table 4. It should be noted that all NPMLEs were computed without restriction to the support. This is because in Poisson mixtures the largest point in the support of an unrestricted NPMLE cannot be greater than the largest observation (see for example, Lindsay, 1995) and also because in theoretical results the interval $[0, M]$, i.e. the support of $\Lambda$, may be fixed arbitrarily large (see Proposition 1 above, and Karlis and Patilea, 2008).

Table 3: NPMLE Based CI's

| Number of Claims | | | | | | | |
|---|---|---|---|---|---|---|---|
| Year | | | | $k$ | | | |
| $t$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| 0 | [0.125066, 0.151935] | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 1 | [0.109660, 0.135809] | [0.185184, 0.267589] | [0.268289, 0.491340] | [0.282276, 0.996328] | [0.458679, 1.953238] | [0.747085, 3.274443] | [0.000000, 5.814480] |
| 2 | [0.101249, 0.121899] | [0.175168, 0.225622] | [0.250150, 0.355885] | [0.332413, 0.556384] | [0.576910, 0.974589] | [1.059058, 1.698085] | [1.267187, 2.306903] |
| 3 | [0.093517, 0.111645] | [0.161832, 0.202639] | [0.237686, 0.306021] | [0.293704, 0.429557] | [0.354031, 0.602667] | [0.617892, 0.991348] | [1.142949, 1.658676] |
| 4 | [0.086998, 0.103189] | [0.148335, 0.186124] | [0.223511, 0.275676] | [0.288551, 0.378911] | [0.307061, 0.478578] | [0.355977, 0.635075] | [0.639244, 1.012645] |
| 5 | [0.081636, 0.095892] | [0.135725, 0.173462] | [0.206673, 0.251645] | [0.282834, 0.349452] | [0.310256, 0.427989] | [0.310738, 0.512743] | [0.350130, 0.661270] |

Table 4: Bootstrap Percentile CI's

| Number of Claims | | | | | | | |
|---|---|---|---|---|---|---|---|
| Year | | | | $k$ | | | |
| $t$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| 0 | [0.131961, 0.144876] | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 1 | [0.116414, 0.129309] | [0.187658, 0.255338] | [0.275891, 0.551029] | [0.324608, 1.087990] | [0.330889, 2.233373] | [0.332413, 3.401027] | [0.332446, 3.677890] |
| 2 | [0.103606, 0.122096] | [0.147649, 0.240049] | [0.229204, 0.417394] | [0.277024, 0.797113] | [0.321135, 1.247833] | [0.332401, 2.115248] | [0.332444, 2.936941] |
| 3 | [0.091261, 0.119221] | [0.128747, 0.235845] | [0.205423, 0.349681] | [0.236691, 0.642096] | [0.274231, 0.956950] | [0.321234, 1.394413] | [0.331201, 2.095983] |
| 4 | [0.079151, 0.117867] | [0.118158, 0.233527] | [0.160054, 0.311536] | [0.214810, 0.524566] | [0.239658, 0.843697] | [0.270235, 1.062482] | [0.314374, 1.534964] |
| 5 | [0.067264, 0.117130] | [0.110900, 0.233524] | [0.134524, 0.282395] | [0.204345, 0.445821] | [0.221954, 0.719269] | [0.240163, 0.974344] | [0.269635, 1.183716] |

Overall, from Tables 3 and 4 we observe that the Wald type intervals based on NPMLE, and the Efron bootstrap percentile intervals in most cases do not differ greatly. In both cases, a policyholder who is observed for $t = 5$ years of his presence in the portfolio and has a low claim frequency has a smaller confidence interval radius than one who in the same period of observation has more expected claims. For instance, when$(t = 1, k = 2)$ the premium rates range from 0.26829 to 0.49134 and from 0.27589 to 0.55103, when $(t = 4, k = 3)$ the premium rates range from 0.28855 to 0.37891 and from 0.21481 to0.52457, and when $(t = 5, k = 5)$ the premium rates range from 0.31074 to 0.51274 and from 0.24016 to 0.97434 in the case of the Wald type intervals based on NPMLE and Efron bootstrap percentile intervals respectively. However, as we have already mentioned, for $(t = 1, k = 6)$ the NPMLE based approach provides a very large and, thus, unusable CI. This aspect is improved by the bootstrap type interval which is not unreasonably long. The construction of confidence intervals is important because it indicates the precision of the estimates of the premiums of an optimal BMS. The reliability of the resulting premium estimates is bigger if the length of the intervals is smaller.

# 7 Conclusion

The present paper addressed the issue of building confidence intervals for the premiums determined by an optimal BMS, In this respect, actuarial literature research was extended since previous designs of such systems failed to identify customers with high claim frequency as they usually represent very few observations. Specifically, NPML was used for estimating the risk distribution in a mixed Poisson model for the claim counts and this system was derived by means of the Bayes theorem, i.e. by updating the posterior mean claim frequency. As a result of the asymptotic normality of the estimator of the posterior mean claim frequency, Wald type two-sided confidence intervals were constructed. Such intervals are not degenerated and therefore are more useful than the corresponding intervals which could be derived from empirical estimation and those resulting from model based probability estimates that depend heavily on the form of the model under consideration. However, the construction of Wald type CIs relies on standard deviation estimates and thus in certain circumstances may have negative bounds, and as such prove to be larger than the nominal level. Therefore, the investigation was taken another step forward by considering the construction of Efron percentile bootstrap two-sided confidence intervals which was based on bootstrap from the NPMLE of the mixture. Efron type intervals require much more computing, but may make important improvements to the asymptotic normal approximation used by Wald intervals. In an Bonus-Malus ratemaking scheme, the use of these intervals is beneficial for the insurance company as they account for the fluctuations in the imposed premiums. Moreover, their constructions can be employed with flexibility by insurance companies which are free in a competitive market to set up their own tariff structures and rating policies.

Furthermore, we would like to emphasize that the interest is not on identifying risk groups. So using a smoother mixing distribution is not a key ingredient for our derivations, since we focus on the estimated claim distribution itself and not on the number of support points themselves. Also, the usage of covariate information for the model for a priori classification is under investigation. However, the derivation of the asymptotic normality is such cases is not straightforward and hence construction of confidence intervals needs further work.

Finally, a possible line of further research would be to employ nonparametric mixtures of a multivariate Poisson distribution in order to construct an optimal BMS with a finite number of classes that takes into account different types of claims, for example claims with or without bodily injuries, or claims with full or partial liability of the insured driver. In this case, the independence assumption between claim types can be relaxed and it would be interesting to observe how the BMS might be affected. Moreover, one can investigate the asymptotic behaviour of the maximum likelihood estimators of the probabilities of a multivariate Poisson with a nonparametric mixing distribution. Specifically, if the asymptotic normality for the estimator of individual probabilities can be established, then following and extending the framework of the present work, the NPML estimator can be used for the construction of confidence intervals for the premiums that must be paid for different types of claims.

**Acknowledgments:** We would like to thank the editor and the reviewer for their constructive comments and suggestions.

# Appendix A. Proof of Proposition 1

Fix $k_j \in \mathbb{N}$ , the number of claims that the policyholder had in year $j, j = 1, ..., t$. Then, $K = \sum_{j=1}^{t} k_j$ ,the total number of claims of a policyholder after $t$ years of insurance, will be a fixed value in $\mathbb{N}$. Also, define the interval $J = \{K, K+1\}$ and consider the probabilities $\pi_{F_{\Lambda_0}}(J) = \pi_{F_{\Lambda_0}}(K) + \pi_{F_{\Lambda_0}}(K+1)$. Furthermore, let $\hat{\pi}(K)$ and $\hat{\pi}(J)$ be the NPML estimators of $\pi_{F_{\Lambda_0}}(K)$ and $\pi_{F_{\Lambda_0}}(J)$ respectively. Based on Corollary 2 of Böhning and Patilea (2005), which is an extension of Corollary 5.1 of Lambert and Tierney (1984), one can see that

$$\sqrt{n}\left(\hat{\pi}(J) - \pi_{F_{\Lambda_0}}(J), \hat{\pi}(K) - \pi_{F_{\Lambda_0}}(K)\right) \Longrightarrow N_2\left((0,0), \Omega(K)\right),$$

where $N_2$ denotes a bivariate normal law and

$$\Omega(K) = \begin{pmatrix} \pi_{F_{\Lambda_0}}(J) - \pi^2_{F_{\Lambda_0}}(J) & \pi_{F_{\Lambda_0}}(J)\left[1 - \pi_{F_{\Lambda_0}}(K)\right] \\ \pi_{F_{\Lambda_0}}(J)\left[1 - \pi_{F_{\Lambda_0}}(K)\right] & \pi_{F_{\Lambda_0}}(K) - \pi^2_{F_{\Lambda_0}}(K) \end{pmatrix} \tag{A.1}$$

On the other hand, the premium that must be paid by this specific individual at $t+1$ will be given $\lambda_{t+1}(K) = \psi\left(\pi_{F_{\Lambda_0}}(J), \pi_{F_{\Lambda_0}}(K)\right)$, where $\psi(x,y) = \left(\frac{K+1}{t}\right)\frac{x-y}{y}$. Let $\nabla\psi(x,y)$ represent the vector of first-order partial derivatives of $\psi(.,.)$ at a point $(x,y)$ with $y \neq 0$. The delta-method (see, for example, van der Vaart, 1998, Theorem 3.1) implies that

$$\sqrt{n}\left(\hat{\lambda}_{t+1}(K) - \lambda_{t+1}(K)\right) \Longrightarrow N\left(0, V_{t+1}(K)\right), \tag{A.2}$$

where

$$\begin{aligned}
V_{t+1}(K) &= \nabla\psi\left(\pi_{F_{\Lambda_0}}(J), \pi_{F_{\Lambda_0}}(K)\right)\Omega(K)\left\{\nabla\psi\left(\pi_{F_{\Lambda_0}}(J), \pi_{F_{\Lambda_0}}(K)\right)\right\}' = \\
&= \left(\frac{1}{\pi_{F_{\Lambda_0}}(K)}, -\frac{\pi_{F_{\Lambda_0}}(J)}{\pi^2_{F_{\Lambda_0}}(K)}\right) \times \begin{pmatrix} \pi_{F_{\Lambda_0}}(J) - \pi^2_{F_{\Lambda_0}}(J) & \pi_{F_{\Lambda_0}}(J)\left[1 - \pi_{F_{\Lambda_0}}(K)\right] \\ \pi_{F_{\Lambda_0}}(J)\left[1 - \pi_{F_{\Lambda_0}}(K)\right] & \pi_{F_{\Lambda_0}}(K) - \pi^2_{F_{\Lambda_0}}(K) \end{pmatrix} \times \\
&\quad \times \begin{pmatrix} \frac{1}{\pi_{F_{\Lambda_0}}(K)} \\ -\frac{\pi_{F_{\Lambda_0}}(J)}{\pi^2_{F_{\Lambda_0}}(K)} \end{pmatrix}\left(\frac{K+1}{t}\right)^2 \\
&= \frac{\pi_{F_{\Lambda_0}}(J)\,\pi_{F_{\Lambda_0}}(K)}{\pi^3_{F_{\Lambda_0}}(K)}\left(\frac{K+1}{t}\right)^2 \\
&= \left(\frac{K+1}{t}\right)^2\left[\frac{\pi_{F_{\Lambda_0}}(K+1)}{\pi^2_{F_{\Lambda_0}}(K)}\right]\left[1 + \frac{\pi_{F_{\Lambda_0}}(K+1)}{\pi_{F_{\Lambda_0}}(K)}\right].
\end{aligned}$$

# Appendix B. Proof of asymptotic consistency of Efron percentile bootstrap confidence intervals

Let us now provide a proof of the asymptotic consistency of the Efron percentile bootstrap interval (13) provided that the assumptions of Proposition 1 are satisfied. In what follows, fix $k_j \in \mathbb{N}, j = 1, ..., t$ ,thus $K \in \mathbb{N}$ is a fixed value, and define $J$ as in the previous proof. For each $l \in \mathbb{N}$, let $p_n^*(l)$ denote the proportion of observations equal to $l$ in a bootstrap sample. Under the assumptions of Proposition 1, Karlis and Patilea (2008) showed that for any $l \in \mathbb{N}$, if

$$R_n^*(l) = \sqrt{n}\left(\hat{\pi}^*(l) - p_n^*(l)\right),$$

then for any $\delta > 0$, $P\left(\left|R_n^*(l)\right| > \delta | \hat{\pi}\right) \to 0$ in probability. From this, deduce that if $\hat{\pi}^*(J) = \sum_{l \in J} \hat{\pi}^*(l) = \hat{\pi}^*(K) + \hat{\pi}^*(K+1), p_n^*(J) = \sum_{l \in J} p_n^*(l) = p_n^*(K) + p_n^*(K+1)$ and

$$R_n^* = \sqrt{n}\left(\hat{\pi}^*(K) - p_n^*(K), \hat{\pi}^*(J) - p_n^*(J)\right),$$

then for any $\delta > 0$, $P\left(\left\|R_n^*\right\| > \delta | \hat{\pi}\right) \to 0$ in probability. Based on the last display and using a central limit theorem for a triangular array (see, for instance, van der Vaart, 1998, pp. 20, 330–331) applied for the vector $\left(p_n^*(K), p_n^*(J)\right)$, deduce that for any $t_1, t_2 \in \mathbb{R}$

$$P\left(\sqrt{n}\left(\hat{\pi}^*(K) - \hat{\pi}^*(K)\right) \le t_1, \left(\hat{\pi}^*(J) - \hat{\pi}(J)\right) \le t_2 | \hat{\pi}\right) \to F_1(t_1.t_2),$$

where $F_1(.,.)$ is the distribution function of a bivariate centered normal law with the variance matrix $\Omega(K)$ given by (A.1).Working with subsequences along which the sequence $\sqrt{n}\left(\hat{\pi}^*(K) - \hat{\pi}^*(K), \hat{\pi}^*(J) - \hat{\pi}(J)\right)$ converges weakly to the bivariate normal law, conditionally, almost surely, using the delta method for bootstrap (see, for example,van der Vaart, 1998, Theorem 23.5) we deduce that for any $u \in \mathbb{R}$

$$\left(\sqrt{n}\left(\hat{\lambda}_{t+1}^*(K) - \hat{\lambda}_{t+1}(K) \le u | \hat{\pi}\right)\right) \to F_2(u) \text{ in probability,}$$

where $F_2(.)$ is the distribution function of the centered normal law with variance $V_{t+1}(K)$. Finally, the asymptotic consistency of the Efron percentile bootstrap interval for $\lambda_{t+1}(K)$ follows from the latter, the weak convergence (A.2) and Lemma 23.3 of van der Vaart (1998).

## References


[1] Böhning, D. (2000). Computer-Assisted Analysis of Mixtures and Applications: Meta-Analysis, Disease Mapping and Others. Chapman and Hall, CRC, London.

[2] Böhning, D. and V. Patilea (2005). Asymptotic normality in mixtures of discrete distributions. Scandinavian Journal of Statistics, 32, 115–132.

[3] Boucher, J. P., M. Denuit and M. Guillen (2008). Models of Insurance Claim Counts with Time Dependence Based on Generalisation of Poisson and Negative Binomial Distributions. Variance, 2, 1, 135-162.

[4] Brouhns, N., M. Guillen, M. Denuit and J. Pinquet (2003). Bonus-malus scales in segmented tariffs with stochastic migration between segments. Journal of Risk and Insurance, 70, 577-599.

[5] Coene, G. and L.G. Doray (1996). A Financially Balanced Bonus-Malus System. ASTIN Bulletin, 26, 107-115.

[6] Denuit, M. and Ph. Lambert (2001). Smoothed NPML estimation of the risk distribution underlying Bonus-Malus systems. Proceedings of the Casualty Actuarial Society, 88, 142–174.

[7] Denuit, M., Marechal, X., Pitrebois, S. Walhin, J. F. (2007). Actuarial modelling of claim counts: risk classification, and bonus-malus systems. Hoboken, NJ: Wiley.

[8] Dersimonian, R. (1986). Algorithm AS221 Maximum Likelihood Estimation of a Mixing Distribution. Applied Statistics, 35, 302-309.

[9] Dionne, G. and C. Vanasse (1989). A generalization of actuarial automobile insurance rating models: the negative binomial distribution with a regression component. ASTIN Bulletin, 19, 199-212.

[10] Dionne, G. and C. Vanasse (1992). Automobile insurance ratemaking in the presence of asymmetrical information. Journal of Applied Econometrics, 7, 149-165.

[11] Efron, B. (1982). The jackknife, the bootstrap and other resampling plans. Philadelphia, PA: CBMS 38. SIAM-NSF.

[12] Insurance Europe (2015). European Insurance Key Facts. Link: http://www.insuranceeurope.eu/sites/default/files/attachments/European %20Insurance%20-%20Key%20Facts%20-%20August%202015.pdf.

[13] Karlis, D. and V. Patilea (2007). Confidence Intervals of the hazard function for discrete distributions using mixtures. Computational Statistics and Data Analysis, 51, 5388-5401.

[14] Karlis, D. and V. Patilea (2008). Bootstrap confidence intervals intervals in mixtures of discrete distributions. Journal of Statistical Planning and Inference, 138, 2313-2329.

[15] Kestemont, R.M., and J. Paris (1985). Sur l'ajustement du Nombre de Sinistres. Bulletin of the Swiss Association of Actuaries, 157–163.

[16] Laird, N. (1978). Nonparametric Maximum Likelihood Estimation of a Mixing Distribution. Journal of the American Statistical Association, 73, 805-811.

[17] Laird, N.M. and T.A. Louis (1987). Empirical Bayes confidence intervals based on bootstrap samples (with discussion). Journal of the American Statistical Association, 82, 739–757.

[18] Lambert, D. and L. Tierney (1984). Asymptotic properties of the maximum likelihood estimates in mixed Poisson model. The Annals of Statistics, 12, 1388–1399.

[19] Lesperance, M. and Kalbfleisch, J. (1992). An Algorithm for Computing the Nonparametric MLE of a Mixing Distribution. Journal of the American Statistical Association, 87, 120-126.

[20] Lindsay, B.G. (1983). The Geometry of Mixture Likelihood. A General Theory. The Annals of Statistics, 11, 86-94.

[21] Lindsay, B. G. (1995). Mixture models: theory, geometry and applications. NSF-CBMS regional conference series in probability and statistics, vol. 5. Hayward, CA: Institute of Mathematical Statistics and the American Statistical Association.

[22] Lindsay, B. G., and Lesperance, M. L. (1995). A review of semiparametric mixture models. Journal of Statistical Planning and Inference, 47(1), 29-39.

[23] Lemaire, J. (1995). Bonus-malus systems in automobile insurance. Norwell, MA: Kluwer Academic Publishers.

[24] Pinquet, J. (1998). Designing Optimal Bonus-Malus Systems From Different Types of Claims. ASTIN Bulletin, 28, 205-220.

[25] Pinquet, J., M. Guillen, and C. Bolance (2001). Long-range contagion in automobile insurance data: estimation and implications for experience rating. ASTIN Bulletin, 31(2), 337-348.

[26] Pitrebois, S., M. Denuit, and JF. Walhin (2005). Bonus-malus systems with varying deductibles. ASTIN Bulletin, 35, 261–274.

[27] Simar, L., (1976). Maximum likelihood estimation of compound Poisson process. The Annals of Statistics, 4, 1200–1209.

[28] Tremblay, L. (1963). Using the Poisson-Inverse Gaussian distribution in Bonus-Malus Systems. ASTIN Bulletin, 22, 97–106.

[29] Tzougas,G.Frangos,N. (2014). The design of an optimal bonus-malus system-based on the Sichel distribution.Modern problems in insurance mathematics. New York: Springer Verlag.

[30] Tzougas, G., S, Vrontos and N. Frangos (2014). Optimal Bonus-Malus Systems Using Finite Mixture Models. ASTIN Bulletin, 44, 417-444.

[31] Walhin, J.-F., and J. Paris (1999). Using Mixed Poisson Distributions in Connection with Bonus-Malus Systems. ASTIN Bulletin, 29, 81–99.

[32] Wang, Y. (2007). On the fast computation of the nonparametric maximum likelihood estimate of a mixing distribution. Journal of the Royal Statistical Society, Series B, 69, 185-198.

[33] Willmot, G. E. (1987). The Poisson-inverse Gaussian distribution as an alternative to the negative binomial. Scandinavian Actuarial Journal, 3-4, 113-127.

[34] van de Geer, S. (1993). Hellinger-consistency of certain nonparametric maximum likelihood estimators. The Annals of Statistics, 21, 14-44.

[35] van der Vaart, A.W. (1998). Asymptotic Statistics. Cambridge University Press, Cambridge.