INSTITUT NATIONAL DE LA STATISTIQUE ET DES ETUDES ECONOMIQUES
Série des Documents de Travail du CREST
(Centre de Recherche en Economie et Statistique)

n° 2004-06

Bootstrap Confidence Intervals in Mixtures of Discrete Distributions

D. KARLIS[1]
V. PATILEA[2]

---

[1] Department of Statistics, Athens University of Economics, 76 Patission Str., Athens 10434, Greece. Email : karlis@hermes.aueb.gr
[2] CREST-ENSAI, Campus de Ker Lann, Rue Blaise Pascal, BP 37203, 35172 BRUZ Cedex, France. Email : valentin.patilea@ensai.fr.
Part of the work for this paper was done while this author was at LEO, Université d'Orléans, France.

Bootstrap confidence intervals in mixtures of discrete distributions

Dimitris Karlis* and Valentin Patilea†

* Department of Statistics, Athens University of Economics, 76 Patission Str, Athens 10434, Greece. Email: karlis@hermes.aueb.gr

† CREST-ENSAI, Campus de Ker Lann, Rue Blaise Pascal BP 37203, 35172 Bruz Cedex, France. E-mail: valentin.patilea@ensai.fr. Part of the work for this paper was done while this author was at LEO, Université d'Orléans, France.

Abstract

The problem of building bootstrap confidence intervals for small probabilities with count data is considered. The true probability distribution generating the independent observations is supposed to be a mixture of a given family of power series distributions. The mixing distribution is estimated by nonparametric maximum likelihood and the corresponding mixture is used for resampling. We build percentile$-t$ and Efron percentile bootstrap confidence intervals for the probabilities and we prove their consistency in probability. The theoretical results are supported by simulation experiments for Poisson and Geometric mixtures. We compare percentile$-t$ and Efron percentile bootstrap intervals with other eight bootstrap or asymptotic theory based intervals. It appears that Efron percentile bootstrap interval outperforms the competitors in terms of coverage probability and length.

Key words: percentile$-t$ confidence intervals, Efron percentile confidence intervals, mixture models, power series distributions, nonparametric maximum likelihood, asymptotic normality

MSC 2000: 62G15, 62G09, 62G20

Résumé

Nous voulons construire des intervalles de confiance pour les probabilités d'une variable aléatoire discrète. La vraie loi marginale des observations iid est un mélange infini de lois discrètes appartenant à une famille définie par une série de puissances. La mesure de probabilité mélangeante est estimée par le maximum de vraisemblance nonparamétrique. Le mélange donné par cet estimateur est utilisé comme distribution de rééchantillonnage. Nous utilisons les échantillons bootstrap pour construire des intervalles de confiance par les méthodes $t-$percentile et percentile d'Efron. Nous prouvons la validité asymptotique de ces intervalles. Nos résultats sont illustrés par des simulations.

Mots clefs: Intervalles de confiance, bootstrap, maximum de vraisemblance nonparametrique, mélanges de lois discrètes.

1 Introduction

In certain circumstances, researchers face the problem of reliably estimating and/or building confidence intervals for small probabilities using count data. In many cases the available data are either too few to allow for asymptotic arguments or they need to be smoothed in order to be practical. For example, when examining the seismicity of an area, reliable data exist only for the last years. If the interest focuses on seismic events above a certain scale, only very few observations are available. Volcanic eruptions are also rare events with few data. In insurance, interest lies in identifying customers with high number of accidents but usually they represent very few observations. Moreover, insuring rare events like ship salvages is very difficult due to the small number of previous events.

The empirical proportion as an estimate of the event probability is satisfactory only if the event is observed sufficiently many times and it completely useless if the event does not occurred yet. Therefore, in the examples we mentioned above, smoother estimates are needed in order to produce reliable results. Such estimates can be obtained from mixture models for count data. Here, we consider a broad class of mixtures of discrete distributions which allows a flexible modelling of the count distribution. (See, e.g., Lindsay (1995) for a list of applications of mixture models to count data.)

In this paper we focus on confidence intervals for small probabilities in count data. Such intervals can be derived from the asymptotic behavior of the estimates. When the asymptotic approximation of the law of the statistics used for building confidence intervals is poor, the bootstrap represents a remedy. If the model is trustful, bootstrap procedures that take into account the model considered represent appealing alternatives to the usual nonparametric bootstrap based on the empirical distribution [see Efron (1987)]. Such model based bootstrap, usually called *parametric bootstrap*, is even more valuable if the underlying model impose only mild restrictions on the distribution of the observations. Here, we propose resampling from the estimated count distribution in a nonparametric mixture model.

Let us introduce the problem in a more strict mathematical formulation. Consider that the observations are independent copies of a discrete random variable X distributed according to $\pi_{Q_0} = \int_\Theta \pi_\theta \, Q_0(d\theta)$, a mixture of a given family $\{\pi_\theta;\, \theta \in \Theta\}$ of power series distributions (also called linear exponential distributions) with unknown mixing distribution Q_0. We are interested in building confidence intervals for the probabilities

$$\pi_{Q_0,J} = \sum_{k \in J} \pi_{Q_0,k} = P[X \in J],$$

where J is any finite subset of the support of the observations and $\pi_{Q_0,k} = P(X = k)$. In particular, we are interested in intervals with good performances for quantities involving small probabilities $\pi_{Q_0,J}$ or $1 - \pi_{Q_0,J}$.

As mentioned above, the individual probabilities $P(X = k)$ or tail probabilities like $P(X \geq k)$ appear in many statistical applications in seismology, biology, insurance, marketing, *etc.* Sometimes, the interest focuses on certain transformations of the individual probabilities. For instance, in non-life insurance the bonus malus rating involves the quantity $P(X = k)/P(X = k + 1)$, where X is the number of claims of a policy holder [see

Walhin and Paris (1999)]. In engineering, one may be interested to estimate the hazard function $\lambda(k) = P(X = k)/P(X \geq k)$ for an equipment that operates on demand, where X is the number of demands successfully completed.

For estimating Q_0 we consider the nonparametric maximum likelihood approach which is appropriate for situations where only little information about the true mixing distribution is available. Moreover, this adds flexibility in the resulting mixture model. For example, Poisson mixtures are very popular models to capture heterogeneity with respect to the simplistic Poisson model. Using the nonparametric estimate of the mixing distribution we allow for a rich family of count distributions instead of restricting our attention to particular families (such as the negative binomial distribution, perhaps the most famous member of the family of Poisson mixtures).

Let $\widehat{Q}$ and $\widehat{\pi} = \pi_{\widehat{Q}}$ denote the nonparametric maximum likelihood estimator (NPMLE) of Q_0 and the corresponding mixture, respectively. Lambert and Tierney (1984) and Patilea (2002) showed that for power series distribution mixture models,

$$\sqrt{n}\frac{\widehat{\pi}_J - \pi_{Q_0,J}}{\sqrt{\widehat{\pi}_J(1-\widehat{\pi}_J)}} = \sqrt{n}\frac{p_{n,J} - \pi_{Q_0,J}}{\sqrt{p_{n,J}(1-p_{n,J})}} + o_P(1), \tag{1.1}$$

where n is the sample size, $p_{n,J}$ denotes the empirical proportion of observations in J and $\widehat{\pi}_J$ is the estimate of $\pi_{Q_0,J}$ yielded by the NPMLE of Q_0. From this result and a classical central limit theorem one can derive a confidence interval for $\pi_{Q_0,J}$.

As an alternative to the confidence interval based on the asymptotic behavior of $\widehat{\pi}_J$ we consider bootstrap confidence intervals. The following resampling scheme has been used several times in the literature: given the observed data, draw samples of the same size from the mixture $\widehat{\pi}$ estimated by NPMLE and use these samples for inference [e.g., Laird and Louis (1987), Böhning (1999) and Mao and Lindsay (2002)]. Herein, we compute bootstrap estimates $\widehat{\pi}_J^*$ and we use them to obtain an approximation for the law of $\sqrt{n}(\widehat{\pi}_J - \pi_{Q_0,J})/\sigma\left(\widehat{\pi}_J\right)$ with $\sigma^2\left(\widehat{\pi}_J\right) = \widehat{\pi}_J(1-\widehat{\pi}_J)$. Next, we deduce *percentile*$-t$ confidence intervals for $\pi_{Q_0,J}$. This approach is equivalent to a type of Empirical Bayes confidence intervals proposed by Laird and Louis (1987) [for a refinement see also Carlin and Gelfand (1991)]. A second idea we analyze is to build bootstrap percentile confidence intervals defined as $[\widehat{\zeta}_{\alpha/2}, \widehat{\zeta}_{1-\alpha/2}]$, where $\widehat{\zeta}_\alpha$ is the $\alpha-$quantile of $\widehat{\pi}_J^*$ [Efron (1982)]. Let us call these intervals *Efron percentile confidence intervals.*

In this paper we prove the consistency in probability of such infinite-dimensional parametric bootstrap confidence intervals in power series distributions mixture models. When the parameter of the model is of infinite dimension, as it is the case for our models where Q_0 is estimated nonparametrically, proving consistency for the parametric bootstrap intervals become a delicate task. (See, e.g., Putter and van Zwet (1996), Beran (1997) for some consistency results.) Our strategy is to obtain an identity similar to (1.1) in the bootstrap world.

Here, we focus on mixtures of distributions with infinite support. When the observations have a finite support (e.g., binomial mixture models) the identity (1.1) holds for any subset J only if the true mixture is in the interior of the set of all mixtures [see Patilea (2002)]. In this case, almost surely, the NPMLE based mixture $\widehat{\pi}$ is equal to the empirical distribution provided that the sample size is sufficiently large. Therefore, results on consistency of the nonparametric bootstrap of the mean can be directly applied.

Definitions and results for mixtures of power series distributions are recalled in section 2. Moreover, we prove the almost sure convergence of the NPMLE in the case of triangular arrays. The consistency of the percentile$-t$ and Efron percentile bootstrap confidence intervals is obtained in section 3. In section 4 we present empirical results on the performances (expressed in terms of coverage probability and length) of percentile$-t$, Efron percentile and other eight types of intervals in Poisson and Geometric mixtures. In view of simulation output we recommend the Efron percentile method. The empirical results indicate that, at least for some probabilities $P(X \in J)$, consistency of bootstrap intervals holds also for discrete mixing distributions, a case that is not covered by the theoretical results. Finally, we build confidence intervals for probabilities and tail probabilities using earthquakes data. Concluding remarks can be found in section 5.

2 Mixtures of power series distributions

Consider $a_k \geq 0$, $k = 0, 1, ...$, and define the power series $a(\theta) = \sum_{k\geq 0} a_k \theta^k$. Denote by R the radius of convergence of the series and set $\Theta = [0, R)$. Moreover, define the set $\mathbb{K} = \{k \geq 0;\ a_k > 0\}$. Let $\{\pi_{\theta,k}\}_{k\in\mathbb{K}}$, $\theta \in \Theta$, with

$$\pi_{\theta,k} = a_k \theta^k a(\theta)^{-1},$$

be a family of power series distributions (PSD) with support $\mathbb{K}$. Here, we consider the case where the support $\mathbb{K}$ is an infinite set of nonnegative integers. Some common examples of PSD with infinite support are Poisson $[a(\theta) = \exp(\theta),\ R = \infty]$, zero-truncated Poisson $[a(\theta) = \exp(\theta) - 1,\ R = \infty]$, logarithmic series $[a(\theta) = -\log(1-\theta),\ R = 1]$, negative binomial $[a(\theta) = (1-\theta)^{-v}$ with $v > 0$ fixed, $R = 1]$ and geometric $[a(\theta) = \theta(1-\theta)^{-1}$, $R = 1]$. See Johnson *et al.* (1992) for many other examples.

A mixture of a given family $\{\pi_\theta;\ \theta \in \Theta\}$ of PSD is a probability measure $\pi_Q = \{\pi_{Q,k}\}_{k\in\mathbb{K}}$ with the individual probabilities defined as

$$\pi_{Q,k} = \int_\Theta \pi_{\theta,k} Q(d\theta) = \int_\Theta \frac{a_k \theta^k}{a(\theta)} Q(d\theta), \quad k \in \mathbb{K},$$

where Q is the mixing distribution, that is a probability measure on Θ endowed with the Borel $\sigma-$field. Consider the independent observations distributed according to a mixture $q_0 = \pi_{Q_0}$ with individual probabilities $q_{0,k}$, $k \in \mathbb{K}$. The true mixing distribution Q_0 is unknown but its support is included in a known compact interval $[0, M] \subset \Theta$.

By definition, Q_0 is identifiable if $\pi_{Q_0} = \pi_Q$ implies $Q_0 = Q$. The following result proved by Patilea (2002) [see also Milhaud and Mounime (1995)] states the identifiability of Q_0 in PSD mixture models under a mild additional condition on $\mathbb{K}$. For other identifiability results see, e.g., Sapatinas (1995).

Proposition 2.1 *Assume that the support of the true mixing distribution Q_0 is contained in a compact interval $[0, M] \subset \Theta$. If $\sum_{k\in\mathbb{K},\, k>0} k^{-1} = \infty$, then Q_0 is identifiable among all the mixing distributions with the support in Θ.*

The mixing distribution Q_0 is estimated by nonparametric maximum likelihood. Let $X_1, ..., X_n \in \mathbb{K}$ be an i.i.d. sample with distribution q_0. The log-likelihood function is

$$l_n(Q) = n \sum_{\mathbb{K}} p_{n,k} \log \pi_{Q,k},$$

where $p_n = \{p_{n,k}\}_{k \in \mathbb{K}}$ is the vector of observed proportions. Let $\widehat{Q}$ be the NPMLE of Q_0, that is

$$l_n(\widehat{Q}) = \sup_{Q} l_n(Q),$$

where the maximum is taken over all probability measures on $[0, M]$. The mixture $\widehat{\pi} = \pi_{\widehat{Q}}$ is the NPMLE of q_0. Existence, support size, uniqueness and other finite sample properties of $\widehat{Q}$ can be deduced using the same arguments as Simar (1976) and Lindsay (1995, ch. 5).

Concerning the consistency of the NPMLE, it was proved that, almost surely, $\widehat{\pi}_k \to q_{0,k}$, $k \in \mathbb{K}$ (e.g., van de Geer (2000), section 4.2). If Q_0 is identifiable it follows that $\widehat{Q} \to Q_0$ weakly, almost surely. To obtain bootstrap theoretical results we need to extend the consistency result to triangular arrays. Recall that if π, π' are two probability distributions with support $\mathbb{K}$, the Hellinger distance between π and π' is

$$h(\pi, \pi') = \left(\frac{1}{2} \sum_{k \in \mathbb{K}} \left(\sqrt{\pi_k} - \sqrt{\pi'_k} \right)^2 \right)^{1/2}.$$

Clearly, $h(\pi_n, \pi) \to 0$ implies $\pi_{n,k} \to \pi_k$, for all $k \in \mathbb{K}$. Consider a family of PSD and fix $0 < M < R$. Let $\mathcal{P}_M$ be the set of all mixtures π_Q with $Q([0, M]) = 1$. The proof of the following lemma is given in the appendix.

Lemma 2.2 *Consider $\{\pi_n\}$ a sequence of mixtures in $\mathcal{P}_M$. Let $\widehat{\pi}_n$ denote the NPMLE obtained from an i.i.d. sample of size n distributed according to π_n, $n \geq 1$. Then, $h(\widehat{\pi}_n, \pi_n) \to 0$ almost surely. In particular, if $\{\pi_n\}$ is such that $h(\pi_n, \pi) \to 0$ for some $\pi \in \mathcal{P}_M$, then $h(\widehat{\pi}_n, \pi) \to 0$ almost surely.*

For the asymptotic law of $\widehat{\pi}$ let us recall the χ^2-type norms considered by Lambert and Tierney (1984). If $x \in \mathbb{R}^{\mathbb{K}}$ and π is a probability measure with support $\mathbb{K}$, define $\|x\|_\pi = (\sum_{\mathbb{K}} x_k^2 / \pi_k)^{1/2}$. Moreover, the inner product between x and y is defined by $\langle x, y \rangle_\pi = \sum_{\mathbb{K}} x_k y_k / \pi_k$ if $\|x\|_\pi$, $\|y\|_\pi < \infty$.

In the case of Poisson mixtures Lambert and Tierney (1984) proved that if $q_0 \in \mathcal{P}_M$ for some given $M > 0$, then

$$\sqrt{n} \langle \widehat{\pi}, x \rangle_{q_0} = \sqrt{n} \langle p_n, x \rangle_{q_0} + R_n(q_0, x), \tag{2.1}$$

with $R_n(q_0, x) = o_P(1)$, provided that the vector x satisfies certain conditions. (See also Patilea (2002) for an extension of this result to more general PSD families as defined above.) In view of this identity, the asymptotic normality

$$\sqrt{n} \langle \widehat{\pi} - q_0,\, x \rangle_{q_0} \rightsquigarrow N(0, \sigma^2(x)) \qquad \text{with} \qquad \sigma^2(x) = \|x\|_{q_0}^2 - \langle q_0, x \rangle_{q_0}^2, \tag{2.2}$$

is a simple consequence of a classical central limit theorem; $\rightsquigarrow$ denotes weak convergence.

Lambert and Tierney (1984), for the case of Poisson mixtures, and Patilea (2002), for the case of more general PSD mixtures as considered herein, indicated a class $\mathcal{C}_1$ of vectors x for which identity (2.1) holds. This class is defined as the set of $x \in \mathbb{R}^{\mathbb{K}}$ such that there exists a sequence $\{g_j\}$ of real-valued measurable bounded functions defined on Θ with i) $\sup_\Theta |g_j(\theta)| \leq Cj^{\xi}$ for some $C, \xi > 0$; and ii) $\|x - x(g_j)\|_{q_0} = O\left(j^{-\beta}\right)$ for some $\beta > 0$,

$$x_k(g_j) = \int_\Theta \frac{a_k\,\theta^k}{a(\theta)}\, g_j\,(\theta)\; Q_0(d\theta) = \int_\Theta \pi_{\theta,\,k}\, g_j\,(\theta)\; Q_0(d\theta).$$

Under certain conditions on the behavior of Q_0 near the origin, the unit vectors $e_i = \{e_{i,k}\}_{k\in\mathbb{K}}$ where $e_{i,k} = 1$ if $k = i$ and 0 otherwise, belong to $\mathcal{C}_1$ and thus satisfy (2.1) and (2.2). The same property follows for any linear combination of q_0 and the unit vectors. The next lemma was proved by Lambert and Tierney (1984) and Patilea (2002).

Lemma 2.3 *Assume that there exist positive constants d, γ, ϵ such that*

$$Q_0((\theta, \theta + \tau]) \geq d\tau^{\gamma}, \qquad for \quad all \quad \theta, \tau \in (0, \epsilon). \tag{2.3}$$

Then, for any $i \in \mathbb{K}$ the unit vector e_i belongs to $\mathcal{C}_1$.

3 Bootstrap confidence intervals

Consider a PSD mixture model as above and define a bootstrap procedure where the bootstrap samples $X_1^*, ..., X_n^*$ are generated according to $\widehat{\pi} = \pi_{\widehat{Q}}$. This is a parametric bootstrap procedure where the unknown parameter is the mixing distribution and the parameter space is of infinite dimension. The unknown parameter is estimated by nonparametric maximum likelihood.

Let $p_n^* = \{p_{n,k}^*\}_{k\in\mathbb{K}}$ and $\widehat{\pi}^* = \{\widehat{\pi}_k^*\}_{k\in\mathbb{K}}$ be the empirical proportions and the NPMLE mixture, respectively, obtained from a bootstrap sample. Like for computing $\widehat{\pi}$, the NPMLE $\widehat{\pi}^*$ is obtained from nonparametric maximum likelihood over the mixing distributions with the support in $[0, M]$.

For $\alpha \in (0,1)$, let $\widehat{\kappa}_\alpha$ (resp. $\widehat{\zeta}_\alpha$) denote the smallest value z that satisfies the inequality

$$P\left(\sqrt{n}\frac{\widehat{\pi}_J^* - \widehat{\pi}_J}{\sigma(\widehat{\pi}_J^*)} \leq z \mid \widehat{\pi}\right) \geq \alpha \qquad \text{(resp.} \quad P\left(\widehat{\pi}_J^* \leq z \mid \widehat{\pi}\right) \geq \alpha). \tag{3.1}$$

The notation $P(\cdot \mid \widehat{\pi})$ indicates that the distribution of $\widehat{\pi}_J^*$ must be evaluated assuming that the bootstrap observations are sampled according to $\widehat{\pi}$ given the original data $X_1, ..., X_n$ (in particular, $\widehat{\pi}_J$ is considered nonrandom). The percentile$-t$ (resp. Efron percentile) confidence interval for $q_{0,J}$ is defined as

$$\left[\widehat{\pi}_J - \frac{\widehat{\kappa}_{1-\alpha/2}}{\sqrt{n}}\,\sigma(\widehat{\pi}_J)\,, \quad \widehat{\pi}_J - \frac{\widehat{\kappa}_{\alpha/2}}{\sqrt{n}}\,\sigma(\widehat{\pi}_J)\right] \qquad \text{(resp.} \quad [\widehat{\zeta}_{\alpha/2},\, \widehat{\zeta}_{1-\alpha/2}]). \tag{3.2}$$

The asymptotic normality of $\sqrt{n}(\widehat{\pi}_J - q_{0,J})/\sigma(\widehat{\pi}_J)$ is a consequence of (2.2). Therefore, to prove the asymptotic consistency of the percentile$-t$ confidence intervals it suffices to prove that, for every z

$$P\left(\sqrt{n}\frac{\widehat{\pi}_J^* - \widehat{\pi}_J}{\sigma(\widehat{\pi}_J^*)} \leq z \mid \widehat{\pi}\right) \to \Phi(z), \tag{3.3}$$

in probability, where $\widehat{\pi}_J^* = \sum_{k\in J} \widehat{\pi}_k^*, \ \ \sigma^2(\widehat{\pi}_J^*) = \widehat{\pi}_J^*(1-\widehat{\pi}_J^*)$ and $\Phi(\cdot)$ denotes the standard normal distribution function [e.g., van der Vaart (1998), page 329].

The asymptotic consistency of the Efron percentile confidence intervals will be obtained from the asymptotic normality of $\sqrt{n}(\widehat{\pi}_J - q_{0,J})$ and the fact that, for any z

$$P\left(\sqrt{n}\,(\widehat{\pi}_J^* - \widehat{\pi}_J) \leq z \mid \widehat{\pi}\right) \to F(z), \tag{3.4}$$

in probability, where F denotes the normal $N(0, q_{0,J}(1-q_{0,J}))$ distribution function.

To prove the asymptotic consistency of the two bootstrap confidence intervals above we have to distinguish two cases depending on whether the radius of convergence R of the PSD family is finite or not. First, we consider the case $R = \infty$ which includes, for instance, the mixtures of Poisson distributions. Conditions (3.3) and (3.4) are then consequences of the following proposition showing that, in some sense, the difference between the empirical distribution p_n^* and the NPMLE mixture $\widehat{\pi}^*$ is negligible in the bootstrap world. This extends the result of Lambert and Tierney (1984) on the asymptotic equivalence between the NPMLE mixture $\widehat{\pi}$ and the observed proportion p_n [see also equation (2.1) above].

Proposition 3.1 *Assume that the support of Q_0 is contained in some known $[0, M]$. Moreover, Q_0 is identifiable. Then, for any $x \in \mathcal{C}_1$,*

$$\sqrt{n}\langle\widehat{\pi}^*, x\rangle_{\widehat{\pi}} = \sqrt{n}\langle p_n^*, x\rangle_{\widehat{\pi}} + R_n^*(\widehat{\pi}, q_0, x), \tag{3.5}$$

where, for any $\delta > 0$, we have $P\left(|R_n^*(\widehat{\pi}, q_0, x)| > \delta \mid \widehat{\pi}\right) \to 0$ in probability.

The proof is given in the appendix. As a consequence of this result we have the consistency of the two bootstrap confidence intervals considered above.

Proposition 3.2 *Consider $\{\pi_\theta;\, \theta \in [0,\infty)\}$ a family of PSD with infinite radius of convergence. Assume that the i.i.d. observation are distributed according to a mixture with mixing distribution Q_0 having the support contained in some known $[0, M]$. Moreover, Q_0 is identifiable and satisfies condition (2.3). Let J be a finite subset of the support of the observations. Then the percentile$-t$ and Efron percentile confidence intervals defined in (3.2) are asymptotically consistent at level $1-\alpha$.*

Proof. It remains to prove (3.3) and (3.4) [see, e.g., van der Vaart (1998), Lemma 23.3]. For (3.3), use Proposition 3.1 for $x = e_i$, $i \in J$ and write

$$\sqrt{n}\,\frac{\widehat{\pi}_J^* - \widehat{\pi}_J}{\sigma(\widehat{\pi}_J^*)} = \frac{\sigma\left(\widehat{\pi}_J\right)}{\sigma\left(\widehat{\pi}_J^*\right)}\sqrt{n}\,\frac{p_{n,J}^* - \widehat{\pi}_J}{\sigma\left(\widehat{\pi}_J\right)} + \frac{1}{\sigma\left(\widehat{\pi}_J^*\right)}\sum_{i\in J}\widehat{\pi}_i R_n^*(\widehat{\pi}, q_0, e_i).$$

Moreover, in view of Lemma 2.2, for any $\delta > 0$, $P\left(\left|\sigma\left(\widehat{\pi}_J\right)/\sigma\left(\widehat{\pi}_J^*\right) - 1\right| > \delta \mid \widehat{\pi}\right) \to 0$, in probability. Finally, for every z,

$$P\left(\sqrt{n}\,\frac{p_{n,J}^* - \widehat{\pi}_J}{\sigma\left(\widehat{\pi}_J\right)} \leq z \mid \widehat{\pi}\right) \to \Phi(z),$$

in probability. This last convergence can be obtained from a central limit theorem for a triangular array [e.g., Van der Vaart (1998), pages 20, 330-1; recall that $\sqrt{n}\left(\widehat{\pi}_J - p_{n,J}\right) \to 0$, in probability, where p_n denotes the empirical proportions of the original sample]. Deduce that (3.3) holds. Similar arguments apply for (3.4). ∎

Let us complete this theoretical section with some remarks. Lambert and Tierney (1984) proved the asymptotic equivalence between the NPMLE mixture $\widehat{\pi}$ and the observed proportion p_n when Q_0 satisfies condition (2.3) which is a kind of continuity assumption in the neighborhood of the origin. In particular, this condition implies that the support of Q_0 is an infinite set. Moreover, Lambert and Tierney showed that, in general, the asymptotic equivalence between the NPMLE and the empirical proportions no longer holds when the true mixing distribution Q_0 is discrete. In our bootstrap procedure the samples are drawn from $\widehat{\pi} = \pi_{\widehat{Q}}$ which is a mixture with a discrete mixing distribution $\widehat{Q}$. This may raise questions about the asymptotic validity of our bootstrap. In view of Propositions 3.1 and 3.2, the fact that $\widehat{Q}$ converges to Q_0 with Q_0 satisfying the continuity condition (2.3) suffices to prove a kind of asymptotic equivalence between NPMLE and the empirical proportions in the bootstrap world and to derive the asymptotic consistency of our bootstrap intervals.

The assumption that the support of the true mixing distribution of contained in some known compact interval $[0, M]$ is a convenient technical condition. It can be relaxed at the expense of more complicated arguments. In practice, this assumption has no real impact because M can be fixed arbitrarily large and thus the probability that the largest point in the support of the unrestricted NPMLE lies outside $[0, M]$ is practically null.

The steps followed for proving Propositions 3.1 and 3.2 can be also used when $R < \infty$. However, when R is finite knowing M such that $Q_0([0, M]) = 1$ is no longer enough for the proofs, except when M is the smallest value having this property, that is if M is the upper limit of the support of Q_0. (In fact, Lemma 6.1 in the Appendix is no longer valid for any interval $[0, M]$ including the support of Q_0 and for any $a \geq 1$. This failure becomes clear, for instance, in view of equation (6.4).) Consequently, for proving the asymptotic consistency of the bootstrap intervals in the case $R < \infty$, one may either suppose that the upper limit of the support of Q_0 is known, or estimate this upper limit in a suitable way. Milhaud and Mounime (1995) proposed an almost surely convergent estimator of the upper limit of the support of Q_0. Given such an estimator, say $\widehat{M}$, the case R finite can be treated exactly as the case $R = \infty$ provided that everywhere M is replaced by $\widehat{M}$. The construction of the estimator $\widehat{M}$ allows the probability that $\widehat{M} = R$ up to very large sample sizes to be arbitrarily close to one. In other words, in applications the unrestricted NPMLE can be used.

The delta method states that if the bootstrap is consistent for estimating the distribution of $\sqrt{n}(\widehat{\pi}_J - q_{0,J})$, then it also consistent for estimating the distribution of $\sqrt{n}(\varphi(\widehat{\pi}_J) - \varphi(q_{0,J}))$ if φ is differentiable at $q_{0,J}$ (e.g., van der Vaart (1998), ch. 23). This allows us to extend Proposition 3.2 in two ways. First, we can derive the asymptotic consistency of the bootstrap confidence intervals for certain transformations of the individual probabilities $q_{0,k}$, $k \in \mathbb{K}$. For instance, the asymptotic validity of percentile$-t$ and Efron percentile intervals for the hazard function $\lambda(k) = P(X = k)/P(X \geq k)$ can be obtained. On the other hand, the delta method allows to prove the asymptotic consistency of percentile$-t$ intervals based on monotone transformations. (Efron percentile interval is invariant under monotone transformations.) In particular, this ensures the asymptotic validity of percentile$-t$ intervals for $q_{0,J}$ based on the logit transform or the variance-stabilizing arcsine transform. These intervals are analyzed in the simulation study reported in section 4.

4 Empirical evidence

4.1 Simulation experiments

We conduct a simulation experiment in order to assess the performances of various confidence intervals (CI) for $q_{0,J} = P(X \in J)$ with $J = \{k_0\}$ and $J = \{k \in \mathbb{K}; k \geq k_1\}$ for some k_0 and k_1. Two kind of PSD families are considered: Poisson ($R = \infty$) and geometric ($R = 1$). Let z_α denote the $\alpha-$quantile of the standard normal distribution.

Three types of intervals are considered: a) intervals based on the empirical proportions $p_{n,J}$; b) intervals based on the asymptotic behavior of $\widehat{\pi}_J$ and some transformations of $\widehat{\pi}_J$; c) bootstrap intervals. The intervals we study are the following.

1) Agresti-Coull CI (abbreviated AC) based on the sample proportions $p_{n,J}$:

$$[\ \widetilde{p_{n,J}} \pm \frac{z_{\alpha/2}}{\sqrt{n + z^2_{\alpha/2}}} \sqrt{\widetilde{p_{n,J}}(1 - \widetilde{p_{n,J}})}\ ],$$

where $\widetilde{p_{n,J}} = (np_{n,J} + z^2_{\alpha/2}/2)(n + z^2_{\alpha/2})^{-1}$;

2) Wilson CI based on the sample proportions $p_{n,J}$:

$$[\ \widetilde{p_{n,J}} \pm \frac{z_{\alpha/2}\sqrt{n}}{n + z^2_{\alpha/2}} \sqrt{p_{n,J}\,(1 - p_{n,J}) + z^2_{\alpha/2}/4n}\ ];$$

3) Wald CI: $[\widehat{\pi}_J \pm z_{\alpha/2}\sqrt{\widehat{\pi}_J(1 - \widehat{\pi}_J)/n}]$ based on *unrestricted* NPMLE;

4) Arcsine transformation based CI: $[\sin^2 \widehat{A},\ \sin^2 \widehat{B}]$, where

$$\widehat{A} = \arcsin \sqrt{\widehat{\pi}_J} \ - \ z_{1-\alpha/2}\, \frac{1}{2\sqrt{n}} \qquad \text{and} \qquad \widehat{B} = \arcsin \sqrt{\widehat{\pi}_J} \ - \ z_{\alpha/2}\, \frac{1}{2\sqrt{n}}$$

5) Logit transformation based CI: $[\exp(\widehat{C})/(1 + \exp \widehat{C}\,),\ \exp(\widehat{D})/(1 + \exp \widehat{D}\,)]$ where

$$\widehat{C} = \ln \frac{\widehat{\pi}_J}{1 - \widehat{\pi}_J} \ - \ \frac{z_{1-\alpha/2}}{\sqrt{n}\sqrt{\widehat{\pi}_J(1 - \widehat{\pi}_J)}} \qquad \text{and} \qquad \widehat{D} = \ \ln \frac{\widehat{\pi}_J}{1 - \widehat{\pi}_J} \ - \ \frac{z_{\alpha/2}}{\sqrt{n}\sqrt{\widehat{\pi}_J(1 - \widehat{\pi}_J)}}.$$

6) Percentile$-t$ CI (abbreviated Per-t) defined in equation (3.2).

7) Percentile$-t$ CI based on the arcsine transformation: $[\sin^2 \widetilde{A},\ \sin^2 \widetilde{B}]$, where

$$\widetilde{A} = \arcsin\sqrt{\widehat{\pi}_J} - \hat{z}^{arcsine}_{1-\alpha/2}\frac{1}{2\sqrt{n}}, \qquad \widetilde{B} = \arcsin\sqrt{\widehat{\pi}_J} - \hat{z}^{arcsine}_{\alpha/2}\frac{1}{2\sqrt{n}}$$

and $\hat{z}^{arcsine}_{\alpha}$ denotes the $\alpha-$quantile of $2\sqrt{n}(\arcsin\sqrt{\widehat{\pi}^*_J} - \arcsin\sqrt{\widehat{\pi}_J})$.

8) Percentile$-t$ CI based on the logit transformation: $[\frac{\exp\widetilde{C}}{1+\exp\widetilde{C}},\ \frac{\exp\widetilde{D}}{1+\exp\widetilde{D}}]$, where

$$\widetilde{C} = \ln\frac{\widehat{\pi}_J}{1-\widehat{\pi}_J} - \frac{\hat{z}^{logit}_{1-\alpha/2}}{\sqrt{n}\sqrt{\widehat{\pi}_J(1-\widehat{\pi}_J)}}, \qquad \widetilde{D} = \ln\frac{\widehat{\pi}_J}{1-\widehat{\pi}_J} - \frac{\hat{z}^{logit}_{\alpha/2}}{\sqrt{n}\sqrt{\widehat{\pi}_J(1-\widehat{\pi}_J)}}$$

and $\hat{z}^{logit}_{\alpha}$ is the $\alpha-$quantile of $\sqrt{n}\sqrt{\widehat{\pi}^*_J\left(1-\widehat{\pi}^*_J\right)}\left\{\ln\left[\widehat{\pi}^*_J/\left(1-\widehat{\pi}^*_J\right)\right] - \ln\left[\widehat{\pi}_J/\left(1-\widehat{\pi}_J\right)\right]\right\}$.

9) Efron (or bootstrap) percentile CI defined in equation (3.2).

10) The bootstrap bias-corrected percentile interval (BC): $[F^{-1}_{Boot}(p_1),\ F^{-1}_{Boot}(p_2)]$, where $p_1 = \Phi(z_{\alpha/2} + 2b_0)$ and $p_2 = \Phi(z_{1-\alpha/2} + 2b_0)$ with $\Phi(\cdot)$ the standard normal distribution function, $F^{-1}_{Boot}(p)$ is the $p-$quantile of the distribution of $\widehat{\pi}^*_J$ and $b_0 = \Phi^{-1}[F_{Boot}(\widehat{\pi}_J)]$.

Brown *et al.* (2001, 2002) recommend Wilson and Agresti-Coull intervals for binomial proportions. Wald CI is justified by the asymptotic normality of $\widehat{\pi}_J$ [see Patilea (2002)]. Arcsine and logit transformations based intervals are obtained by inverting a Wald type interval for the $\arcsin\sqrt{q_{0,J}}$ and $\ln[q_{0,J}/(1-q_{0,J})]$, respectively (see Brown *et al.* (2001) for a presentation of this type of intervals). Note that arcsine is a variance stabilizing transformation for $\widehat{\pi}_J$. The confidence intervals 7) and 8) are the percentile$-t$ versions of the arcsine and logit transformations based intervals. The BC interval was proposed by Efron (1982) as an improvement of the bootstrap percentile CI.

As an alternative, let us mention the bootstrap accelerated bias-corrected percentile (BC$_a$) interval introduced by Efron (1987). The BC$_a$ interval involves the so-called acceleration constant. How to determine or estimate this extra parameter in the case of the NPMLE $\widehat{\pi}_J$ remains an open question. Finally, one may use iterative bootstrap methods to build intervals. However, iterating the bootstrap principle with the NPMLE is very computationally demanding and therefore this approach is not considered herein. (See Shao and Tu (1995, ch. 4) for a review of BC$_a$ and iterative bootstrap method based intervals.)

Several algorithms for computing NPMLE has been proposed [see Böhning (1999) for a recent review]. The algorithm we used for all computations is a variant of the EM algorithm adjusted for jumping between different number of components. Namely, the algorithm starts with maximum possible components [e.g., Lindsay (1995)]. We keep iterating the algorithm until either the convergence criterion (measured by a change of the relative likelihood) is satisfied, or redundant support points (components) are found. A support point is declared redundant if either it is closer than 10^{-6} to another support point, or its mixing proportion is smaller than 10^{-6}. In the former case, the close components are merged if this improves the likelihood. Otherwise, we continue with the same number of components. If a mixing proportion is too small, we remove the support point and rescale the other mixing proportions to sum to one. Note that since we are not interested in reporting the number of support points, redundant support points do not cause problems for the quantities we study herein. The interest when looking for redundant points is to improve the speed of the algorithm.

First, we study mixtures of Poisson distributions. Two mixing distributions Q_0 are considered: a continuous uniform $U[0,3]$ and a discrete distribution $0.5\delta_1 + 0.5\delta_3$, where δ_a denotes a degenerate at a distribution. We build confidence intervals for $P(X=0)$ and $P(X \geq 6)$. If Q_0 is the uniform distribution (resp. the discrete distribution), the true values are $P(X=0) = 0.317$ (resp. 0.208) and $P(X \geq 6) = 0.017$ (resp. 0.042).

Given the mixture $q_0 = \pi_{Q_0}$, we generated $D = 1000$ samples of size n with $n = 25$, 50 and 200, respectively. For each sample we computed $\widehat{\pi}$ and we generated $B = 1000$ bootstrap samples of the same size from $\widehat{\pi}$. For $J = \{0\}$ and $J = \{6, 7, ...\}$, we computed $\widehat{\pi}_J^*$ in each bootstrap sample and we used these values to approximate the quantiles $\widehat{\kappa}_\alpha$, $\widehat{\zeta}_\alpha$, $\hat{z}_\alpha^{arcsine}$ and $\hat{z}_\alpha^{logit}$ and to estimate b_0. All NPMLEs were computed without restriction on the support. This because, on one hand, in Poisson mixtures the largest point in the support of an unrestricted NPMLE cannot exceed the largest observation [e.g., Lindsay (1995), Proposition 25]. On the other hand, in our theoretical results the interval $[0, M]$ could be fixed arbitrarily large.

In order to compare the performances of the intervals considered we report the distribution of their lengths and the estimated levels. The true level (coverage probability) of an interval was estimated by the proportion of times among the $D = 1000$ replications when the true value belonged to the interval. In Figure 1 we provide the box-plots for the length of the CIs obtained from 1000 replications for each of ten types of intervals in the case of uniform mixing distribution. In cases where the lower limit of the derived CI had a negative value we set the lower limit to zero. This results in shortening the length of some of the computed AC, Wilson, Wald and percentile$-t$ intervals.

Insert Figure 1 here

The length distributions are quite similar for the five simple (no bootstrap) intervals. For most of the D generated samples, the five bootstrap CI have slightly smaller lengths than those of the simple CI. We also remark more variability in the lengths of the bootstrap CI. This may be explained by the error induced when approximating the quantiles $\widehat{\kappa}_\alpha$, $\widehat{\zeta}_\alpha$, $\hat{z}_\alpha^{arcsine}$ and $\hat{z}_\alpha^{logit}$ and the constant b_0. The estimated levels of the three percentile$-t$ type intervals are quite poor (see Table 1). All the estimated levels of the simple CI are greater than 0.95.

As expected, the differences between the performances of the competing intervals become significant in the case of the tail probability $P(X \geq 6)$. With one exception (the logit transformation based CI) the lengths of the two intervals based on the empirical proportions $p_{n,J}$ are much larger than the lengths of the intervals based on the mixture model. Moreover, the five bootstrap CI are often shorter then Wald and arcsine transformation based CI. Table 1 reveals a severe failure of the three percentile$-t$ intervals in terms of coverage. Notice also the poor level of the arcsine CI when $n = 25$. (This behavior of the arcsine CI is also noticed by Brown *et al.* (2001) in the case of binomial proportions.) Efron percentile CI appears as the best in terms of length and coverage. The BC intervals fails to improve the Efron percentile. Recall that BC interval relies upon the existence of a transformation ϕ_n such that, given $\widehat{\pi}$, $P(\phi_n(\widehat{\pi}_J) - \phi_n(q_{0,J}) + b_0 \leq z) = \Phi(z) = P(\phi_n(\widehat{\pi}_J^*) - \phi_n(\widehat{\pi}_J) + b_0 \leq z \mid \widehat{\pi})$.

It seems that such a transformation does not exists in our framework. This may be explained by the discrete nature of the NPMLE $\widehat{Q}$.

Table 1. Estimated level for the CIs: Poisson -Uniform

n		Simple CI Wald	AC	Wilson	Arcsin	Logit	Bootstrap CI Per-t	Per-t (arcsin)	Per-t (logit)	Efron	BC
25	$P(X=0)$	0.950	0.969	0.967	0.965	0.970	0.911	0.902	0.889	0.945	0.929
	$P(X\geq 6)$	0.963	0.996	0.964	0.879	0.976	0.868	0.791	0.702	0.980	0.921
50	$P(X=0)$	0.957	0.970	0.968	0.966	0.970	0.923	0.918	0.911	0.956	0.938
	$P(X\geq 6)$	0.977	0.995	0.984	0.998	0.987	0.852	0.818	0.768	0.983	0.913
200	$P(X=0)$	0.954	0.965	0.960	0.957	0.960	0.942	0.934	0.931	0.958	0.942
	$P(X\geq 6)$	0.986	0.993	0.981	0.997	0.986	0.892	0.880	0.861	0.972	0.929

The box-plots for the lengths in the case of discrete mixing distribution are depicted in Figure 2, while Table 2 contains the corresponding estimated levels. The case of a discrete mixing distribution is not covered by our theoretical results. However, the simulations indicate that consistency of the bootstrap intervals may also hold in this case. The performances of the competing CIs are quite similar to those in the previous experiment. The simulations indicate the Efron percentile CI as the best of the ten intervals considered.

Insert Figure 2 here

Table 2. Estimated level for the CIs: Poisson -Discrete

n		Simple CI Wald	AC	Wilson	Arcsin	Logit	Bootstrap CI Per-t	Per-t (arcsin)	Per-t (logit)	Efron	BC
25	$P(X=0)$	0.967	0.971	0.970	0.969	0.971	0.914	0.887	0.875	0.953	0.932
	$P(X\geq 6)$	0.957	0.995	0.984	0.997	0.985	0.874	0.835	0.773	0.962	0.903
50	$P(X=0)$	0.973	0.974	0.965	0.976	0.974	0.910	0.887	0.858	0.960	0.928
	$P(X\geq 6)$	0.967	0.983	0.979	0.981	0.981	0.901	0.889	0.867	0.961	0.921
200	$P(X=0)$	0.961	0.964	0.961	0.967	0.962	0.928	0.921	0.918	0.943	0.932
	$P(X\geq 6)$	0.960	0.966	0.963	0.968	0.965	0.918	0.914	0.910	0.942	0.929

In our last experiment we studied mixtures of geometric laws. If π_θ is a geometric law of parameter $\theta \in [0,1)$, then $\pi_{\theta,k} = \theta^{k-1}(1-\theta)$, $k = 1, 2, ...$ The true mixing distribution is taken to be an uniform $U[0, 0.5]$. Note that the mixture of a geometric with a Uniform has slowly decreasing tail probabilities.

Insert Figure 3 here

Table 3. Estimated level for the CIs: Geometric -Uniform

n		Simple CI Wald	AC	Wilson	Arcsin	Logit	Bootstrap CI Per-t	Per-t (arcsin)	Per-t (logit)	Efron	BC
25	$P(X\geq 5)$	0.925	0.995	0.980	0.600	0.984	0.896	0.726	0.603	0.952	0.916
50	$P(X\geq 5)$	0.949	0.995	0.977	0.984	0.983	0.861	0.792	0.704	0.965	0.896
200	$P(X\geq 5)$	0.963	0.991	0.985	0.991	0.986	0.867	0.850	0.827	0.946	0.892

We restrict our interest in a tail probability, namely the quantity of interest is $P(X \geq 5)$ and its true value is 0.0125. The sample sizes we consider are $n = 25, 50$ and 200. We proceed as in the first experiment with the same values for D and B. Moreover, unrestricted NPMLE is considered. In Figure 3 we provide box-plots for the lengths of the competing CIs while in Table 3 we report estimated levels. Again, the performances of percentile$-t$ type intervals are poor. Meanwhile, Efron percentile interval appears to be the best CI in terms of length and coverage.

In view of the simulation results, one may remark the good performances of the Wald CI which may be considered the second best interval in terms of length and coverage. The estimated level of Wald CI is quite close to that of Efron percentile CI. Since Wald CI requires a single computation of the NPMLE, in Figure 4 we present the box-plots of the ratio between the lengths of the Wald and Efron percentile intervals for the Poisson mixtures considered above. We notice significantly shorter lengths for Efron percentile CI.

Insert Figure 4 here

4.2 Application to real data

We end this empirical section with an application with real data. The data used were downloaded from the Northern California Earthquake Data Center and they refer to the catalog of earthquakes in North Carolina contributed by Northern California Seismic Network. Only earthquakes with magnitude larger than 6.0 in Richter's scale were considered for the period from 1976 till 1995. The counts are the number of earthquakes of the required magnitude for each year of this period, that is there are 20 observations. The total number of events were 32 and thus the mean per year was 1.6 events while the variance much higher, that is 3.30. This implies a great overdispersion making mixtures of Poisson plausible models. The actual observed proportions were $(p_{20,0}, \ldots, p_{20,6}) = (0.35, 0.30, 0.05, 0.15, 0.05, 0.05, 0.05)$. We used this data set to construct CIs for several quantities, including simple probabilities and tail probabilities (see Figure 5). The sample size is quite small and, thus, bootstrap confidence intervals as considered above can be quite helpful, especially for small probabilities.

Insert Figure 5 here

Let us point out that according to the standard Agresti-Coull and Wilson methods and due to the same observed frequency, both $P(X = 2)$ and $P(X = 6)$ have the same CIs. The confidence intervals that use the properties of the mixture model compensate for the small sample size that led to the same frequencies. Similarly, since there is no observation larger than 6, estimation of the tail probability $P(X \geq 7)$ would be useless with standard CIs for binomial proportions, while the model based CIs (bootstrap or no bootstrap) overcome this problem. The intervals depicted in Figure 5 have not been adjusted for negative values. For the case $P(X = 0)$ where the probability is quite large all the CIs are quite similar. This is not the case for small probabilities like $P(X = 2)$, $P(X = 6)$ or $P(X \geq 7)$ where the intervals are very different.

5 Conclusion

We have studied the properties of bootstrap confidence intervals for probabilities $P(X \in J)$ when the law of X is a mixture of power series distributions and J is a finite set. The bootstrap samples are drawn from $\widehat{\pi}$, the nonparametric maximum likelihood estimate of the true mixture. We considered percentile$-t$ and Efron percentile confidence intervals for which we proved consistency in probability. For empirical comparisons, we also considered two types of intervals for binomial proportions, Wald confidence intervals based on the asymptotics of $\widehat{\pi}$ and some transformations of $\widehat{\pi}$ and several alternative percentile$-t$ type confidence intervals. Simulation results recommend Efron percentile interval, especially for small probabilities $P(X \in J)$. The Efron percentile method performs stably and leads to short intervals with good coverage.

Our theoretical results do not cover the case where the true law of the observations is a discrete mixture. This limitation is due to our approach based on the fact that, in some sense and under certain conditions, $\widehat{\pi}$ and the empirical distribution are asymptotically equivalent. This asymptotic equivalence no longer holds when the true mixture is discrete. In fact, in this case there exist finite sets J such that $\{\sqrt{n}(\widehat{\pi}_k - \pi_{0,k})\}_{k\in J}$ cannot have a nondegenerate normal limit [see Patilea (2002)]. However, we conjecture that for certain finite sets J of which the singletons $J = \{k\}$, $k \in \mathbb{K}$, the consistency of the two bootstrap intervals for $P(X \in J)$ holds also when the true law is a discrete mixture.

Finally, it should be possible to extend the results in the paper to mixtures of an exponential family of continuous distributions following the guidelines of Lambert and Tierney (1984, page 1398). However, quantities for which identities like (2.1) and (3.5) could be easily derived do not have much practical interest.

6 Appendix

Proof of Lemma 2.2. Let $\mathcal{R}$ be the space of all probability distributions on $[0, M]$ endowed with the topology of weak convergence. Herein p_n denotes the empirical distribution corresponding to an i.i.d. sample of size n drawn from π_n, $n \geq 1$. Recall that

$$h^2(\widehat{\pi}_n, \pi_n) \leq \sum_{k\in\mathbb{K}} \frac{2\widehat{\pi}_{n,k}}{\widehat{\pi}_{n,k} + \pi_{n,k}} (p_{n,k} - \pi_{n,k}) \tag{6.1}$$

[e.g., Lemma 4.5 of van de Geer (2000)]. In view of this inequality it suffices to show that the family $\mathcal{G} = \{g_\psi;\ \psi = (Q', Q'') \in \mathcal{R} \times \mathcal{R}\}$ with

$$g_{\psi,k} = \pi_{Q',k}/(\pi_{Q',k} + \pi_{Q'',k}), \qquad k \in \mathbb{K},$$

is Glivenko-Cantelli uniformly in $\pi \in \mathcal{P}_M$ [see van der Vaart and Wellner (1996), page 167]. Consider $\mathcal{R} \times \mathcal{R}$ with a suitable metric τ inducing the product topology. Then, $(\mathcal{R} \times \mathcal{R}, \tau)$ is a compact metric space. For any $k \in \mathbb{K}$, the map $\psi \to g_{\psi,k}$ is continuous and that $\mathcal{G}$ is uniformly bounded. Define the vector $\omega(\psi, \rho)$ by pointwise supremum:

$$\omega(\psi, \rho) = \sup_{\tau(\psi,\widetilde{\psi})<\rho} \left| g_\psi - g_{\widetilde{\psi}} \right| \in \mathbb{R}^{\mathbb{K}}, \qquad \psi \in \mathcal{R} \times \mathcal{R}, \quad \rho > 0,$$

[see also van de Geer (1993), Lemma 5.1]. By dominated convergence

$$\lim_{\rho \to 0} \sum_{k \in \mathbb{K}} \omega_k \left(\psi, \rho\right) \, a_k M^k = 0.$$

Fix $\delta > 0$ arbitrary and take ρ_ψ such that

$$\sum_{k \in \mathbb{K}} \omega_k \left(\psi, \rho_\psi\right) \, a_k M^k \leq \delta, \qquad \psi \in \mathcal{R} \times \mathcal{R}.$$

Define $B_\psi = \left\{ \widetilde{\psi} \in \mathcal{R} \times \mathcal{R}; \, \tau(\psi, \widetilde{\psi}) < \rho_\psi \right\}$ and let $B_{\psi_1}, ..., B_{\psi_r}$ be a finite cover of $\mathcal{R} \times \mathcal{R}$. Consider the pairs of functions

$$g_i^L = g_{\psi_i} - \omega\left(\psi_i, \rho_{\psi_i}\right), \qquad g_i^U = g_{\psi_i} + \omega\left(\psi_i, \rho_{\psi_i}\right), \qquad i = 1, ..., r,$$

and remark that for any $\pi \in \mathcal{P}_M$,

$$\sum_{k \in \mathbb{K}} \left(g_{i,k}^U - g_{i,k}^L\right) \pi_k \leq C \sum_{k \in \mathbb{K}} \omega_k \left(\psi_i, \rho_{\psi_i}\right) a_k M^k \leq C\delta, \qquad i = 1, ..., r,$$

for some constant $C > 0$. From the definition of these functions we have that for any $g_\psi \in \mathcal{G}$, there exists $1 \leq i \leq r$ such that $g_i^L \leq g_\psi \leq g_i^U$. Then, for any $n \geq 1$,

$$\begin{aligned} \sum_{k \in \mathbb{K}} g_{\psi,k} \left(p_{n,k} - \pi_{n,k}\right) &\leq \sum_{k \in \mathbb{K}} g_{i,k}^U \left(p_{n,k} - \pi_{n,k}\right) + \sum_{k \in \mathbb{K}} \left(g_{i,k}^U - g_{\psi,k}\right) \pi_{n,k} \\ &\leq \sum_{k \in \mathbb{K}} g_{i,k}^U \left(p_{n,k} - \pi_{n,k}\right) + C\delta. \end{aligned}$$

Similarly,

$$\begin{aligned} \sum_{k \in \mathbb{K}} g_{\psi,k} \left(p_{n,k} - \pi_{n,k}\right) &\geq \sum_{k \in \mathbb{K}} g_{i,k}^L \left(p_{n,k} - \pi_{n,k}\right) - \sum_{k \in \mathbb{K}} \left(g_{\psi,k} - g_{i,k}^L\right) \pi_{n,k} \\ &\geq \sum_{k \in \mathbb{K}} g_{i,k}^L \left(p_{n,k} - \pi_{n,k}\right) - C\delta. \end{aligned}$$

Using a strong law of large numbers that holds uniformly in the underlying distribution [e.g., van der Vaart and Wellner (1996), page 456] deduce that almost surely

$$\max_{1 \leq i \leq r} \left| \sum_{k \in \mathbb{K}} g_{i,k}^U \left(p_{n,k} - \pi_{n,k}\right) \right| \leq \delta, \qquad \max_{1 \leq i \leq r} \left| \sum_{k \in \mathbb{K}} g_{i,k}^L \left(p_{n,k} - \pi_{n,k}\right) \right| \leq \delta,$$

for n sufficiently large. So eventually,

$$\left| \sum_{k \in \mathbb{K}} \frac{2\widehat{\pi}_{n,k}}{\widehat{\pi}_{n,k} + \pi_{n,k}} \left(p_{n,k} - \pi_{n,k}\right) \right| \leq (C+1)\delta,$$

almost surely. This proves that $h(\widehat{\pi}_n, \pi_n) \to 0$, almost surely. For the last part of the statement use the triangle inequality. ■

For the proof of our Proposition 3.1 we use the following lemma extending Proposition 3.1 of Lambert and Tierney (1984) to a triangular array. Below, p_n and $\widehat{\pi}_n$ denote the empirical proportions and the NPMLE mixture, respectively, corresponding to an i.i.d. sample of size n generated according to π_n, $n \geq 1$. The proof of our lemma is a simple adaptation of the arguments of Lambert and Tierney. We provide this proof for the sake of completeness.

Lemma 6.1 *Let $\{\pi_n\} \subset \mathcal{P}_M$ such that $Q_n \to Q_0$ weakly. For any $a \geq 1$ and $\varepsilon > 0$,*

$$i) \sum_k \frac{(p_{n,k} - \pi_{n,k})^2}{\pi_{n,k}} a^k, \qquad ii) \sum_k \frac{(p_{n,k} - \pi_{n,k})^2}{\widehat{\pi}_{n,k}} a^k,$$

$$iii) \sum_k \frac{(\widehat{\pi}_{n,k} - \pi_{n,k})^2}{\widehat{\pi}_{n,k}} a^k \qquad \text{and} \qquad iv) \sum_k \frac{(\widehat{\pi}_{n,k} - \pi_{n,k})^2}{\pi_{n,k}} a^k,$$

are of order $o_P\left(n^{-(1-\varepsilon)}\right)$.

Proof. There exists $0 < m < M$ and $C_1, C_2 > 0$ such that $C_1 a_k m^k \leq \pi_{n,k} \leq C_2 a_k M^k$, $k \in \mathbb{K}$ if n is sufficiently large. Moreover, from the last part of Lemma 2.2 we have $\widehat{Q}_n \to Q_0$ weakly, almost surely, with $\widehat{Q}_n$ the mixing distributions of $\widehat{\pi}_n$. Deduce that eventually, $\pi_{n,k}/\widehat{\pi}_{n,k}$, $\widehat{\pi}_{n,k}/\pi_{n,k} \leq C_2 C_1^{-1} (M/m)^k$, $k \in \mathbb{K}$, almost surely.

Fix $\varepsilon > 0$ and choose $\gamma \in (1-\varepsilon, 1)$. Take $a > 1$. For the first quantity use Holder's inequality and bound it as follows:

$$\begin{aligned}
n^{1-\varepsilon} \sum_k (p_{n,k} - \pi_{n,k})^2 \pi_{n,k}^{-1} a^k &= n^{1-\varepsilon} \sum_k (p_{n,k} - \pi_{n,k})^2 \pi_{n,k}^{-1} a^{-k/2} a^{3k/2} \\
&\leq n^{1-\varepsilon} \left[\sum_k (p_{n,k} - \pi_{n,k})^2 \pi_{n,k}^{-1} a^{-k/2\gamma} \right]^{\gamma} \\
&\quad \times \left[\sum_k (p_{n,k} - \pi_{n,k})^2 \pi_{n,k}^{-1} a^{3k/2(1-\gamma)} \right]^{1-\gamma} =: n^{1-\varepsilon} A_n^{\gamma} B_n^{1-\gamma}.
\end{aligned}$$

Moreover, $nE\left[\sum (p_{n,k} - \pi_{n,k})^2 \pi_{n,k}^{-1} a^{-k/2\gamma}\right] \leq \sum a^{-k/2\gamma} < \infty$ and thus, for any $C > 0$, $P(nA_n > C) \leq C^{-1} \sum a^{-k/2\gamma}$. It remains to prove that B_n is bounded in probability. Since $(u-v)^2 \leq u^2 + v^2$ if $u, v > 0$, deduce that

$$B_n \leq \sum_k p_{n,k}^2 \pi_{n,k}^{-1} a^{3k/2(1-\gamma)} + \sum_k \pi_{n,k} a^{3k/2(1-\gamma)} =: B_{1n} + B_{2n}.$$

Fix $b > 0$ and define the event $E_n = \{p_{n,k} > b\pi_{n,k} a^{3k/2(1-\gamma)} \text{ for some } k\}$. By Markov's inequality

$$\begin{aligned}
P(E_n) &\leq \sum_k P\left[p_{n,k} > b\pi_{n,k} a^{3k/2(1-\gamma)}\right] \\
&\leq b^{-1} \sum_k E(p_{n,k}) \pi_{n,k}^{-1} a^{-3k/2(1-\gamma)} = b^{-1} \sum_k a^{-3k/2(1-\gamma)}.
\end{aligned}$$

Thus, $P(E_n)$ is bounded by a quantity independent of π_n which can be made arbitrarily small by choosing b sufficiently large. On the complement of E_n, we have

$$B_{1n} \leq b^2 \sum_k \pi_{n,k} a^{9k/2(1-\gamma)} \leq b^2 C_2 \sum_k a_k \left[M a^{9/2(1-\gamma)}\right]^k < \infty, \tag{6.2}$$

for some $C_2 > 0$. Finally,

$$B_{2n} \leq C_2 \sum_k a_k \left[M a^{3/2(1-\gamma)}\right]^k < \infty. \tag{6.3}$$

Conclude that, for any sequence $\{\pi_n\} \subset \mathcal{P}_M$ with $Q_n \to Q_0$ weakly, the first quantity tend to zero in probability faster then $n^{-(1/2-\varepsilon)}$.

For the second quantity, almost surely

$$\begin{aligned} n^{1-\varepsilon} \sum_k (p_{n,k} - \pi_{n,k})^2 \widehat{\pi}_{n,k}^{-1} a^k &= n^{1-\varepsilon} \sum_k (p_{n,k} - \pi_{n,k})^2 \pi_{n,k}^{-1} (\pi_{n,k}/\widehat{\pi}_{n,k})\, a^k \\ &\leq n^{1-\varepsilon} C_2 C_1^{-1} \sum_k (p_{n,k} - \pi_{n,k})^2 \pi_{n,k}^{-1} (aM/m)^k, \end{aligned}$$

if n is sufficiently large. This proves the rate $o_P(n^{-(1/2-\varepsilon)})$ for the second quantity.

For the order of the third quantity note first that by the gradient characterization of the NPMLE of the mixing distribution we have $\|\widehat{\pi}_n - \pi_n\|_{\widehat{\pi}_n} \leq \|p_n - \pi_n\|_{\widehat{\pi}_n}$, $n \geq 1$ [see, e.g., inequality (2.2) in Lambert and Tierney (1984)]. Next, by Holder's inequality

$$n^{1-\varepsilon} \sum_k (\widehat{\pi}_{n,k} - \pi_{n,k})^2 \widehat{\pi}_{n,k}^{-1} a^k \leq n^{1-\varepsilon} \|\widehat{\pi}_n - \pi_n\|_{\widehat{\pi}_n}^{\gamma} \left[\sum_k (\widehat{\pi}_{n,k} - \pi_{n,k})^2 \widehat{\pi}_{n,k}^{-1} a^{k/(1-\gamma)}\right]^{1-\gamma}.$$

Almost surely and for n sufficiently large, the term in the square brackets is bounded by

$$2C_2^2 C_1^{-1} \sum_k a_k M^k \left[a^{1/(1-\gamma)} M/m\right]^k < \infty, \tag{6.4}$$

and thus the order of the third quantity is proved.

Finally, for the order of the fourth quantity we have eventually,

$$\begin{aligned} n^{1-\varepsilon} \sum_k (\widehat{\pi}_{n,k} - \pi_{n,k})^2 \pi_{n,k}^{-1} a^k &= n^{1-\varepsilon} \sum_k (\widehat{\pi}_{n,k} - \pi_{n,k})^2 \widehat{\pi}_{n,k}^{-1} (\widehat{\pi}_{n,k}/\pi_{n,k})\, a^k \\ &\leq n^{1-\varepsilon} C_2 C_1^{-1} \sum_k (\widehat{\pi}_{n,k} - \pi_{n,k})^2 \widehat{\pi}_{n,k}^{-1} (aM/m)^k, \end{aligned}$$

almost surely. The arguments used for the third quantity ensure that the fourth one is also of order $o_P(n^{-(1/2-\varepsilon)})$. ■

Proof of Proposition 3.1. Take a sequence $j(n) \sim n^{(1/2-\varepsilon)/\xi}$, $\varepsilon \in (0, 1/2)$. Decompose $\sqrt{n}\langle \widehat{\pi}^* - p_n^*, x\rangle_{\widehat{\pi}} = A_n + B_n$ where

$$A_n = \sqrt{n}\langle \widehat{\pi}^* - p_n^*, x - x\left(g_{j(n)}\right)\rangle_{\widehat{\pi}}, \qquad B_n = \sqrt{n}\langle \widehat{\pi}^* - p_n^*, x\left(g_{j(n)}\right)\rangle_{\widehat{\pi}}.$$

We have

$$
\begin{aligned}
|A_n| &\leq \sqrt{n}\sum_k \frac{\left|\widehat{\pi}_k^* - p_{n,k}^*\right|}{\widehat{\pi}_k/\sqrt{q_{0,k}}} \frac{\left|x_k - x_k\left(g_{j(n)}\right)\right|}{\sqrt{q_{0,k}}} \\
&\leq \sqrt{n}\left[\sum_k \frac{\left(\widehat{\pi}_k^* - p_{n,k}^*\right)^2}{\widehat{\pi}_k}\frac{q_{0,k}}{\widehat{\pi}_k}\right]^{1/2} \left\|x - x\left(g_{j(n)}\right)\right\|_{q_0} \\
&\leq \sqrt{n}\left\|x - x\left(g_{j(n)}\right)\right\|_{q_0}\left\{\left[\sum_k \frac{\left(\widehat{\pi}_k^* - \widehat{\pi}_k\right)^2}{\widehat{\pi}_k}\frac{q_{0,k}}{\widehat{\pi}_k}\right]^{1/2} + \left[\sum_k \frac{\left(p_{n,k}^* - \widehat{\pi}_k\right)^2}{\widehat{\pi}_k}\frac{q_{0,k}}{\widehat{\pi}_k}\right]^{1/2}\right\}.
\end{aligned}
$$

From the convergence of $\widehat{\pi}$, there exists $0 < m < M$ and $C > 0$ such that, almost surely, $q_{0,k}/\widehat{\pi}_k \leq C(M/m)^k$, $k \in \mathbb{K}$ if n is sufficiently large. Use Lemma 6.1 *i)* , *iv)* to deduce that conditionally on $X_1, X_2, ...$, for almost every sequence $X_1, X_2, ...$, we have $|A_n| \to 0$, in probability. That is, for almost every sequence $X_1, X_2, ...$, if $\delta > 0$, then $P\left(|A_n| > \delta \mid X_1, ..., X_n\right) \to 0$.

Next, write $B_{1n} = \sqrt{n}\langle\widehat{\pi}^* - p_n^*, x\left(g_{j(n)}\right)\rangle_{\widehat{\pi}^*}$ and $B_{2n} = B_n - B_{1n}$. For some $C > 0$,

$$
\begin{aligned}
|B_{2n}| &\leq \sqrt{n}\sum_k \frac{\left|\widehat{\pi}_k^* - p_{n,k}^*\right|\left|\widehat{\pi}_k^* - \widehat{\pi}_k\right|}{\widehat{\pi}_k^*\widehat{\pi}_k}\left|x_k\left(g_{j(n)}\right)\right| \\
&\leq C\sqrt{n}\,j(n)^\xi \sum_k \frac{\left|\widehat{\pi}_k^* - p_{n,k}^*\right|\left|\widehat{\pi}_k^* - \widehat{\pi}_k\right|}{\widehat{\pi}_k^*}\frac{q_{0,k}}{\widehat{\pi}_k} \\
&\leq Cn^{1-\varepsilon}\left\{\left[\sum_k \frac{\left(\widehat{\pi}_k^* - \widehat{\pi}_k\right)^2}{\widehat{\pi}_k^*}\frac{q_{0,k}}{\widehat{\pi}_k}\right]^{1/2} + \left[\sum_k \frac{\left(p_{n,k}^* - \widehat{\pi}_k\right)^2}{\widehat{\pi}_k^*}\frac{q_{0,k}}{\widehat{\pi}_k}\right]^{1/2}\right\} \\
&\qquad\qquad \times\left[\sum_k \frac{\left(\widehat{\pi}_k^* - \widehat{\pi}_k\right)^2}{\widehat{\pi}_k^*}\frac{q_{0,k}}{\widehat{\pi}_k}\right]^{1/2}.
\end{aligned}
$$

Again, using a bound for $q_0/\widehat{\pi}$ and Lemma 6.1 *ii)*, *iii)* deduce that conditionally on $X_1, X_2, ...$, for almost every sequence $X_1, X_2, ...$, we have $|B_{2n}| \to 0$, in probability.

Finally, we have

$$
|B_{1n}| \leq \sqrt{n}\int_{[0,M]}\left|\sum_k \frac{\widehat{\pi}_k^* - p_{n,k}^*}{\widehat{\pi}_k^*}\pi_{\theta,k}\right|\left|g_{j(n)}(\theta)\right| Q_0(d\theta) \leq C\sqrt{n}\,j(n)^\xi\sum_k \frac{\widehat{\pi}_k^* - p_{n,k}^*}{\widehat{\pi}_k^*}q_{0,k},
$$

where for the last inequality we used the gradient characterization of the NPMLE, that is

$\sum_k \left(\widehat{\pi}_k^* - p_{n,k}^*\right) \widehat{\pi}_k^{*-1} \pi_{\theta,k} \geq 0$, $\theta \in [0, M]$. We can further write

$$\begin{aligned}
|B_{1n}| &\leq C\sqrt{n}\, j(n)^\xi \sum_k \frac{\left(\widehat{\pi}_k^* - p_{n,k}^*\right)\left(q_{0,k} - \widehat{\pi}_k^*\right)}{\widehat{\pi}_k^*} \\
&\leq C\sqrt{n}\, j(n)^\xi \left[\sum_k \frac{\left|\widehat{\pi}_k^* - \widehat{\pi}_k\right| \left|\widehat{\pi}_k - q_{0,k}\right|}{\widehat{\pi}_k^*} + \sum_k \frac{\left|\widehat{\pi}_k^* - \widehat{\pi}_k\right|^2}{\widehat{\pi}_k^*} \right. \\
&\qquad \left. + \sum_k \frac{\left|p_{n,k}^* - \widehat{\pi}_k\right| \left|\widehat{\pi}_k - q_{0,k}\right|}{\widehat{\pi}_k^*} + \sum_k \frac{\left|p_{n,k}^* - \widehat{\pi}_k\right| \left|\widehat{\pi}_k - \widehat{\pi}_k^*\right|}{\widehat{\pi}_k^*}\right] \\
&\leq C\sqrt{n}\, j(n)^\xi \left[T_{1n} + T_{2n} + T_{3n} + T_{4n}\right],
\end{aligned}$$

where

$$\begin{aligned}
T_{1n} &= \left\|\widehat{\pi}^* - \widehat{\pi}\right\|_{\widehat{\pi}^*} \left(\sum_k \frac{\left(\widehat{\pi}_k - q_{0,k}\right)^2}{\widehat{\pi}_k} \frac{\widehat{\pi}_k}{\widehat{\pi}_k^*}\right)^{1/2}, \qquad T_{2n} = \left\|\widehat{\pi}^* - \widehat{\pi}\right\|_{\widehat{\pi}^*}^2, \\
T_{3n} &= \left\|p_n^* - \widehat{\pi}\right\|_{\widehat{\pi}^*} \left(\sum_k \frac{\left(\widehat{\pi}_k - q_{0,k}\right)^2}{\widehat{\pi}_k} \frac{\widehat{\pi}_k}{\widehat{\pi}_k^*}\right)^{1/2}, \qquad T_{4n} = \left\|p_n^* - \widehat{\pi}\right\|_{\widehat{\pi}^*} \left\|\widehat{\pi}^* - \widehat{\pi}\right\|_{\widehat{\pi}^*}.
\end{aligned}$$

Use the orders obtained in Lemma 6.1 *ii*), *iii*) to prove that, conditionally on $X_1, X_2, \ldots$, for almost every sequence $X_1, X_2, \ldots$, we have $\sqrt{n}\, j(n)^\xi T_{in} \to 0$ in probability, $i = 2$ and 4. The more difficult terms to be examined are T_{1n} and T_{3n} in which appear q_0, $\widehat{\pi}$ and $\widehat{\pi}^*$. Let $0 < m < M$ and $C > 0$ such that, for almost every fixed sequence $X_1, X_2, \ldots$, we have eventually $\widehat{\pi}_k/\widehat{\pi}_k^* \leq C(M/m)^k$, $k \in \mathbb{K}$, almost surely [see also the arguments at the beginning of the proof of Lemma 6.1]. In particular, m and C are such that if we define the event $E_n = \{\widehat{\pi}_k/\widehat{\pi}_k^* > C(M/m)^k$ for some $k\}$, then $P(E_n \mid X_1, \ldots, X_n) \to 0$, for almost every sequence $X_1, X_2, \ldots$ We study only T_{1n} but the same arguments apply for T_{3n}. For any $\delta > 0$, we can write

$$\begin{aligned}
P\left(n^{1-\varepsilon} T_{1n} > \delta \mid X_1, \ldots, X_n\right) &= P\left(\left\{n^{1-\varepsilon} T_{1n} > \delta\right\} \cap E_n \mid X_1, \ldots, X_n\right) \\
&\quad + P\left(\left\{n^{1-\varepsilon} T_{1n} > \delta\right\} \cap E_n^c \mid X_1, \ldots, X_n\right),
\end{aligned}$$

and therefore we only have to study $\widetilde{T}_{1n} = a_n \left\|\widehat{\pi}^* - \widehat{\pi}\right\|_{\widehat{\pi}^*}$, where

$$a_n^2 = \sum_k \frac{\left(\widehat{\pi}_k - q_{0,k}\right)^2}{\widehat{\pi}_k} (M/m)^k.$$

Let δ, ρ and $c > 0$ be arbitrary and write

$$\begin{aligned}
&P\left[P\left(n^{1-\varepsilon}\, \widetilde{T}_{1n} > \delta \mid X_1, \ldots, X_n\right) > \rho\right] \\
&\quad \leq P\left[P\left(n^{1-\varepsilon}\, \widetilde{T}_{1n} > \delta \mid X_1, \ldots, X_n\right)\ \mathbf{1}\left\{n^{1/2-\varepsilon/2}\, a_n \leq c\right\} > \rho/2\right] \\
&\qquad + P\left[P\left(n^{1-\varepsilon}\, \widetilde{T}_{1n} > \delta \mid X_1, \ldots, X_n\right)\ \mathbf{1}\left\{n^{1/2-\varepsilon/2}\, a_n > c\right\} > \rho/2\right] \\
&\quad \leq P\left[P\left(n^{1/2-\varepsilon/2}\, \left\|\widehat{\pi}^* - \widehat{\pi}\right\|_{\widehat{\pi}^*} > \delta/c \mid X_1, \ldots, X_n\right) > \rho/2\right] \\
&\qquad + P\left[n^{1/2-\varepsilon/2}\, a_n > c\right],
\end{aligned}$$

where $\mathbf{1}_A$ denotes the indicator function of the set A. Use Lemma 6.1 and deduce that $P\left(n^{1/2-\varepsilon/2}\,\|\widehat{\pi}^* - \widehat{\pi}\|_{\widehat{\pi}^*} > \delta/c \mid X_1, ..., X_n\right) \to 0$, for almost every sequence $X_1, X_2, ...$ Moreover, $P\left[n^{1/2-\varepsilon/2}a_n > c\right] \to 0$. In conclusion, if $\delta > 0$, $P\left(n^{1-\varepsilon}\,T_{in} > \delta \mid X_1, ..., X_n\right) \to 0$, in probability, $i = 1$ and 3, and thus $P\left(|B_{1n}| > \delta \mid X_1, ..., X_n\right) \to 0$, in probability. The proof is now complete. ■

References

[1] Beran, R. (1997), "Diagnosing bootstrap success", *Ann. Inst. Statist. Math.*, **49**, 1-24.

[2] Böhning, D. (1999), *Computer-Assisted Analysis of Mixtures and Applications: Meta-Analysis, Disease Mapping and Others,* Chapman & Hall/CRC.

[3] Brown, L.D., Cai, T. and DasGupta, A. (2001), "Interval Estimation for a Binomial Proportion", *Statist. Sci.*, **16**, 101-116.

[4] Brown, L.D., Cai, T. and DasGupta, A. (2002), "Confidence Intervals for a Binomial Proportion and Asymptotic Expansions", *Ann. Statist.*, **30**, 160-201.

[5] Carlin B.P. and Gelfand A.E. (1991), "A sample reuse method for accurate parametric empirical Bayes confidence intervals", *J. Royal Statist. Soc. Series B*, **53**, 189-200.

[6] Efron, B. (1982), "The Jackknife, the Bootstrap and Other Resampling Plans", CBMS 38, SIAM-NSF.

[7] Johnson, N.L., Kotz, S. and Kemp, A.W. (1992), *Univariate Discrete Distributions*, Second Edition. New-York: John Wiley & Sons, Inc.

[8] Laird, N.M. and Louis T.A. (1987), "Empirical Bayes Confidence Intervals Based on Bootstrap Samples" (with discussion), *J. Amer. Statist. Assoc.*, **82**, 739-757.

[9] Lambert, D. and Tierney, L. (1984), "Asymptotic properties of the maximum likelihood estimates in mixed Poisson model", *Ann. Statist.* ***12***, 1388-1399.

[10] Lindsay, B.G. (1995), "Mixture Models: Theory, Geometry and Applications". *NSF-CBMS Regional Conference Series in Probability and Statistics* ***5***, Institute of Mathematical Statistics and American Statistical Association.

[11] Mao, C. and Lindsay, B.G. (2002), "How many genes can be discovered if sequencing more ESTs?", Technical Report, Pennsylvania State University.

[12] Milhaud, X. and Mounime S. (1995), "A Modified Maximum Likelihood Estimator for Infinite Mixtures", Working paper, Université de Toulouse 3.

[13] Patilea, V. (2002), "Asymptotic normality in mixtures of discrete distributions", Document de Recherche 2002-07, Université d'Orléans.

[14] Putter H. and van Zwet, W.R. (1996), "Resampling: Consistency of Substitution estimators", *Ann. Statist.* **24**, 2297-2318.

[15] Sapatinas, T. (1995), "Identifiability of mixtures of power-series distributions and related characterizations", *Ann. Inst. Statist. Math.* **47**, 447-459.

[16] Shao, J. and Tu, D. (1995), *The Jacknife and Bootstrap*, Springer-Verlag, New-York.

[17] Simar, L. (1976), "Maximum likelihood estimation of compound Poisson process", *Ann. Statist.* **4**, 1200-1209.

[18] van de Geer, S. (1993), "Hellinger-consistency of certain nonparametric maximum likelihood estimators", *Ann. Statist.* **21**, 14-44.

[19] van de Geer, S. (2000), *Empirical processes in M-estimation*, Cambridge University Press.

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

[21] van der Vaart, A.W. and Wellner, J.A. (1996), *Weak Convergence and Empirical Processes*, Springer-Verlag, New-York.

[22] Walhin, J.F. and Paris, J. (1999), "Using mixed poisson processes in connection with Bonus-Malus systems", *Astin Bulletin*, **29**, 81-99.

Figure 1: Box-plots for interval lengths: Poisson mixed with a Uniform on $[0, 3]$. The target probabilities are $P(X = 0) = 0.317$ and $P(X \geq 6) = 0.017$.

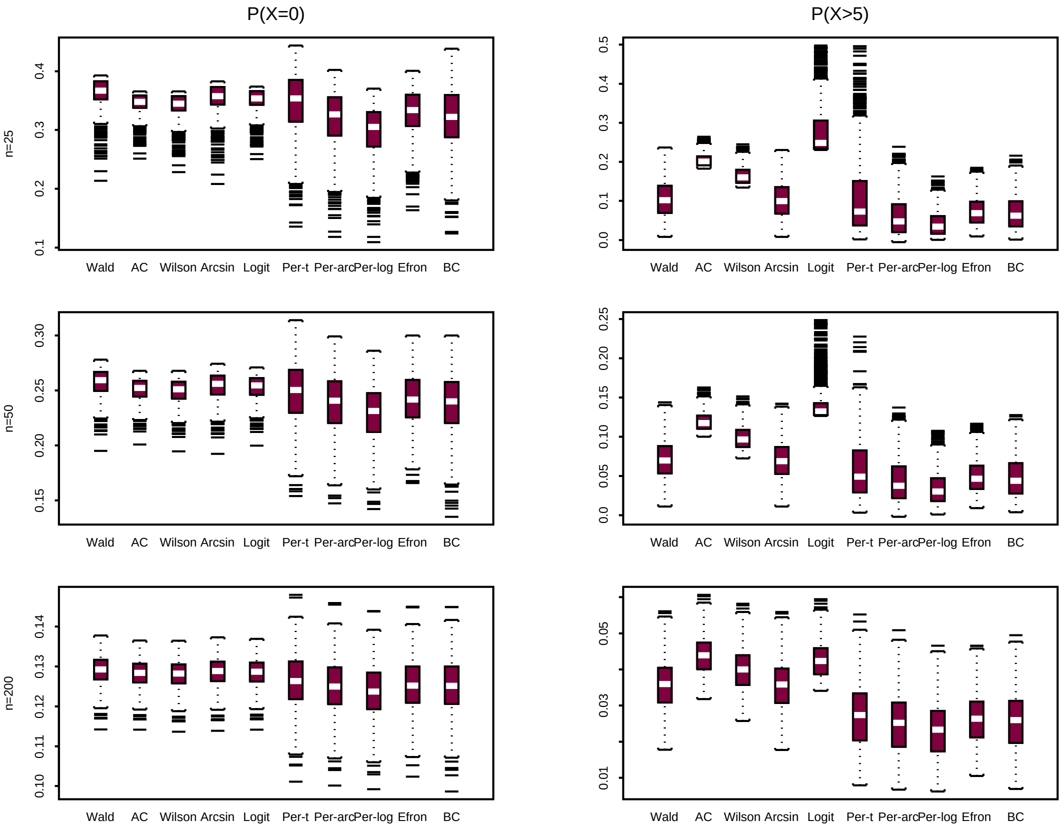

Figure 2: Box-plots for interval lengths: 0.5Poisson(1) + 0.5Poisson(3). The target probabilities are $P(X = 0) = 0.208$ and $P(X \geq 6) = 0.042$.

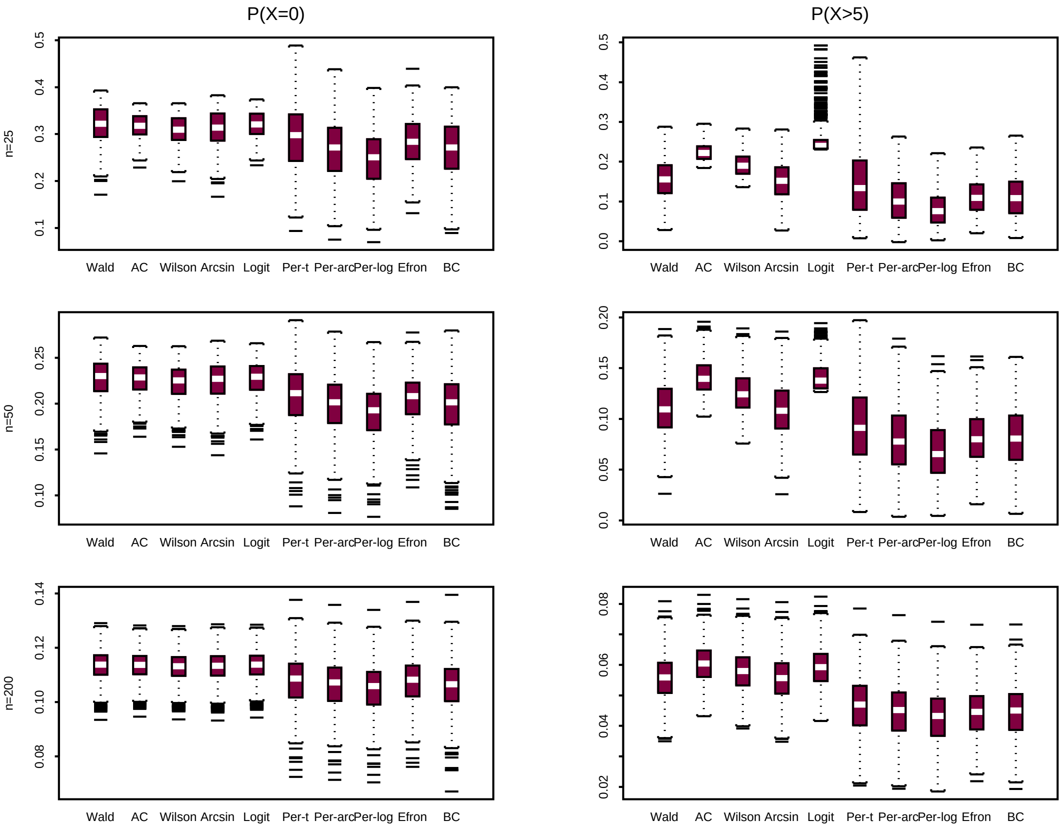

Figure 3: Box-plots for interval lengths: Geometric mixed with a Uniform on $[0, 0.5]$. The target probability is $P(X \geq 5) = 0.0125$.

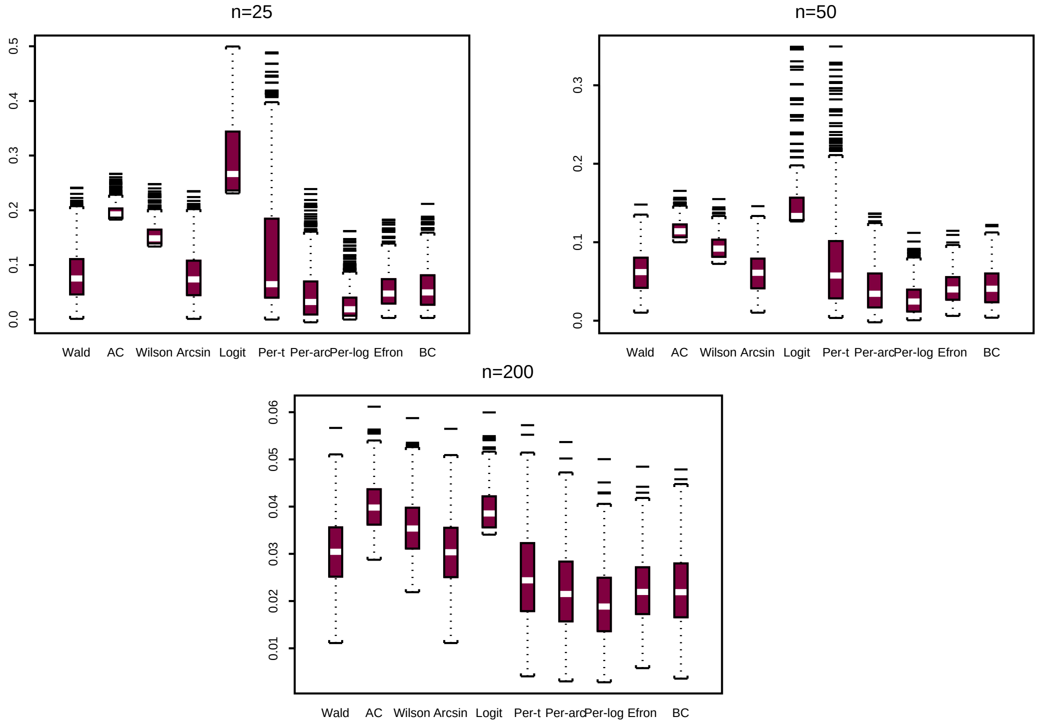

Figure 4: Box-plots of the ratio (length of Wald interval/length of Efron percentile interval).

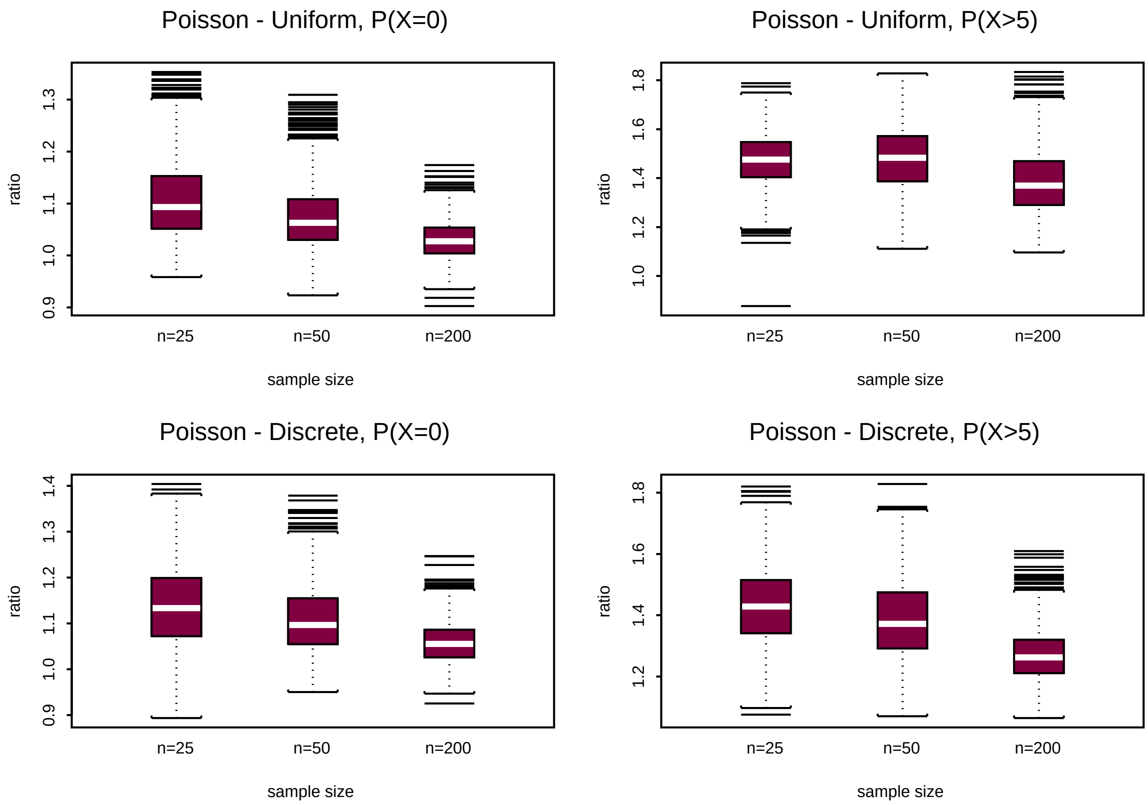

Figure 5: Confidence intervals with Earthquake Data (20 observations). The count variable X represents the number of earthquakes with magnitude larger than 6 in Richter's scale recorded in a year. The solid vertical lines indicate the observed proportions. The vertical dotted line corresponds to the NPMLE of the probabilities.