Optimal rate estimation of the mixing distribution in Poisson mixture models via Laplace inversion

Klaassen, C.A.J.; Mnatsakanov, R.M.

Publication date
2020

Document Version
Final published version

Published in
Transactions of A. Razmadze Mathematical Institute

Transactions of A. Razmadze
Mathematical Institute
Vol. **174** (2020), issue 2, 175–188

OPTIMAL RATE ESTIMATION OF THE MIXING DISTRIBUTION IN POISSON MIXTURE MODELS VIA LAPLACE INVERSION

CHRIS A. J. KLAASSEN[1] AND ROBERT M. MNATSAKANOV[2]

Dedicated to Estate Khmaladze on the occasion of his 75th birthday

Abstract. Consistent estimators of the mixing distribution in Poisson mixture models are constructed for both the right censored and the uncensored case. The estimators are based on a kind of Laplace inversion via factorial moments. The rate of convergence of the mean integrated squared error of these estimators is $(\log n/\log\log n)^2$. It is also shown that there do not exist estimators for which this rate is better.

1. Introduction

Consider independent and identically distributed random variables $X, X_1, X_2, \ldots, X_n$ with discrete distribution

$$p(x) = P(X = x) = \int_0^\infty e^{-\lambda} \frac{\lambda^x}{x!} dG(\lambda), \quad x = 0, 1, \ldots . \tag{1.1}$$

In this Poisson mixture model we shall study nonparametric estimation of the unknown mixing distribution G. This estimation problem is discussed by H. Robbins in [12] (pages 162–163) who suggests to estimate the distribution p of the observations and to solve (1.1) for G with p replaced by its estimate. We apply this approach and solve (1.1) via a kind of Laplace inversion as in Section 4 of [11]. We will investigate the rate of convergence of our estimators as measured by their mean integrated squared error, both in the censored and in the uncensored case. We will also prove this rate to be optimal.

Papers [6] and [8] define multinomial models with a large number of rare events, introduce the concept of a structural function, and discuss its estimation. For polynomial distributions and occupancy problems with a large number of rare events, asymptotic results for the relevant statistics are obtained in [9] and [7], respectively. In [14], the kernel type estimators of the structural distribution function in the multinomial scheme of [6] and [8] are studied via Poissonization.

Approximating the binomial marginals of such multinomial models by Poisson distributions, one arrives at the Poisson mixture model with the distribution function G as a structural function.

2. Construction of the Inverse Transformation

Consider the inhomogeneous Fredholm equation of the first kind (cf. [3])

$$\mathcal{K}G = p, \tag{2.1}$$

where the probability mass function $p(x)$, $x = 0, 1, \ldots$, denotes the Poisson mixture distribution from (1.1). Our construction of estimators of the unknown mixing distribution G is based on a particular type of Laplace inversion as in [2] (Section VII.6, formulae (6.1)–(6.4)). For (2.1) it can be written as follows:

$$(\mathcal{K}_\alpha^{-1}\mathcal{K}G)(z) = (\mathcal{K}_\alpha^{-1}p)(z) = \sum_{k=0}^{\lfloor \alpha z \rfloor} \frac{\alpha^k}{k!} \sum_{j=k}^{\infty} \frac{(-\alpha)^{j-k}}{(j-k)!} \sum_{x=j}^{\infty} \frac{x!}{(x-j)!}\, p(x), \tag{2.2}$$

where $\lfloor y \rfloor$ denotes the integer part of y.

2020 *Mathematics Subject Classification.* Primary: 62G05, Secondary: 62G20.
Key words and phrases. Demixing; Poisson mixture; Asymptotics; Mean integrated squared error.

Lemma 2.1. *Assume* $\int_0^\infty e^{\beta\lambda} dG(\lambda) < \infty$ *for all* $\beta \in \mathbb{R}$. *As* α *tends to infinity,* $\mathcal{K}_\alpha^{-1}$ *from* (2.2) *represents inversion of* $\mathcal{K}$ *from* (2.1) *in the weak sense, i.e.,*

$$\mathcal{K}_\alpha^{-1}\mathcal{K}G \xrightarrow{w} G, \quad \text{as} \quad \alpha \to \infty.$$

Moreover, the transformation $\mathcal{K}_\alpha^{-1}$ *can be written as*

$$(\mathcal{K}_\alpha^{-1}\mathcal{K}G)(z) = (\mathcal{K}_\alpha^{-1}p)(z) = \sum_{x=0}^{\infty} \sum_{k=0}^{\lfloor \alpha z \rfloor \wedge x} \binom{x}{k} \alpha^k (1-\alpha)^{x-k}\ p(x). \tag{2.3}$$

Proof. The factorial moments of the Poisson distribution are powers of its parameter. Consequently, Fubini's theorem (actually Tonelli's theorem) implies

$$\sum_{x=j}^{\infty} \frac{x!}{(x-j)!} p(x) = \int_0^\infty \sum_{x=j}^{\infty} \frac{x!}{(x-j)!} e^{-\lambda} \frac{\lambda^x}{x!} dG(\lambda) = \int_0^\infty \lambda^j dG(\lambda), \ j = 0, 1, \dots . \tag{2.4}$$

Note that retrieving G from these moments is called the moment problem. Subsequently, as the Laplace transform $\int_0^\infty e^{\beta\lambda} dG(\lambda)$ of G is finite for all $\beta \in \mathbb{R}$, Fubini's theorem yields

$$\sum_{j=k}^{\infty} \frac{(-\alpha)^{j-k}}{(j-k)!} \int_0^\infty \lambda^j dG(\lambda) = \int_0^\infty \sum_{j=k}^{\infty} \frac{(-\alpha\lambda)^{j-k}}{(j-k)!} \lambda^k dG(\lambda) = \int_0^\infty e^{-\alpha\lambda} \lambda^k dG(\lambda). \tag{2.5}$$

From (2.2), (2.4) and (2.5) it follows, again by Tonelli's theorem, that

$$(\mathcal{K}_\alpha^{-1}\mathcal{K}G)(z) = \sum_{0 \le k \le \alpha z} \frac{\alpha^k}{k!} \int_0^\infty e^{-\alpha\lambda} \lambda^k dG(\lambda) = \int_0^\infty \sum_{0 \le k \le \alpha z} e^{-\alpha\lambda} \frac{(\alpha\lambda)^k}{k!} dG(\lambda).$$

Chebyshev's inequality for a Poisson random variable Z with parameter $\alpha\lambda$ yields
$P(|Z - \alpha\lambda| \ge \alpha|z - \lambda|) \le \lambda/(\alpha(z-\lambda)^2)$. Hence we obtain

$$\left| \sum_{0 \le k \le \alpha z} e^{-\alpha\lambda} \frac{(\alpha\lambda)^k}{k!} - \mathbf{1}_{[\alpha\lambda \le \alpha z]} \right| \le \left(\frac{\lambda}{\alpha(z-\lambda)^2} \right) \wedge 1.$$

Consequently, at any point of continuity z of G we have

$$(\mathcal{K}_\alpha^{-1}\mathcal{K}G)(z) \to \int_0^\infty \mathbf{1}_{[\lambda \le z]} dG(\lambda) = G(z),$$

as $\alpha \to \infty$; cf. formula (6.1) from Section VII.6 of [2].

Furthermore, from (2.4) and (2.5) with $-\alpha$ replaced by α it follows that Fubini's theorem may be applied to

$$\begin{aligned}\sum_{j=k}^{\infty} \frac{(-\alpha)^{j-k}}{(j-k)!} \sum_{x=j}^{\infty} \frac{x!}{(x-j)!}\ p(x) &= \sum_{x=k}^{\infty} \sum_{j=k}^{x} \binom{x-k}{j-k} (-\alpha)^{j-k} \frac{x!}{(x-k)!}\ p(x) \\ &= \sum_{x=k}^{\infty} (1-\alpha)^{x-k} \frac{x!}{(x-k)!}\ p(x),\end{aligned}$$

which implies

$$(\mathcal{K}_\alpha^{-1}p)(z) = \sum_{k=0}^{\lfloor \alpha z \rfloor} \sum_{x=k}^{\infty} \binom{x}{k} \alpha^k (1-\alpha)^{x-k}\ p(x). \tag{2.6}$$

For $\alpha > 1$, Tonelli's theorem yields

$$\begin{aligned}\sum_{k=0}^{\lfloor \alpha z\rfloor}\sum_{x=k}^{\infty}\binom{x}{k}\alpha^k(\alpha-1)^{x-k}\ p(x) &= \sum_{x=0}^{\infty}\sum_{k=0}^{\lfloor \alpha z\rfloor\wedge x}\binom{x}{k}\alpha^k(\alpha-1)^{x-k}\ p(x)\\ &\le \sum_{x=0}^{\infty}(2\alpha-1)^x\ p(x) = \sum_{x=0}^{\infty}(2\alpha-1)^x\int_0^\infty e^{-\lambda}\frac{(\lambda)^x}{x!}dG(\lambda)\\ &= \int_0^\infty e^{-\lambda}\sum_{x=0}^{\infty}\frac{((2\alpha-1)\lambda)^x}{x!}dG(\lambda) = \int_0^\infty e^{2(\alpha-1)\lambda}dG(\lambda).\end{aligned} \tag{2.7}$$

By the finiteness of the Laplace transform of G the right hand side of (2.7) is finite. Consequently, Fubini's theorem can be applied to (2.6), which yields (2.3). $\square$

Estimating the probability mass function $\mathcal{K}G = p$ from the observations and applying (2.3), we can see by Lemma 2.1 that we might obtain consistent estimators of the mixing distribution G. This is verified for the case of i.i.d. uncensored random variables in Section 3, and under random right censoring in Section 4.

We remark here that the estimator of the so-called structural distribution function for a multinomial random variable discussed in Section 4 of [11] is also based on inversion (2.3) with p replaced by an appropriate empirical version of p.

3. Uncensored Data

Let $X, X_1, \dots, X_n$ be i.i.d. random variables with the Poisson mixture distribution p as in (1.1); cf. (2.1). Replacing the marginal distribution $p(x) = P(X = x)$ in (2.3) by the corresponding empirical version

$$\hat{p}_n(x) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{[X_i=x]}, \tag{3.1}$$

restricting the sum over x in (2.3) to $x \le K_n$, and taking $\alpha = \alpha_n > 1$ dependent on n, we obtain the estimator $\hat{G}_n$ of G with

$$\hat{G}_n(z) = \sum_{x=0}^{K_n}\sum_{k=0}^{\lfloor \alpha_n z\rfloor\wedge x}\binom{x}{k}\alpha_n^k(1-\alpha_n)^{x-k}\ \hat{p}_n(x), \quad z \ge 0. \tag{3.2}$$

In view of Lemma 2.1 and (2.3), the estimator $\hat{G}_n$ should be consistent for appropriately chosen α_n and K_n that tend to infinity when n does. Under reasonable assumptions on the class of mixing distribution functions G it is consistent indeed.

Theorem 3.1. *Let C, D and L be positive constants. Let $G(D) = 1$ hold and let G have a density g that is bounded by C and is Lipschitz continuous with the Lipschitz constant L. Then the mean integrated squared error of $\hat{G}_n$ with $K_n \ge 2\alpha_n De^2$ and $\alpha_n \ge 1$ satisfies*

$$E\int_0^\infty\left(\hat{G}_n(z) - G(z)\right)^2 dG(z) \le \frac{1}{n}(2\alpha_n)^{2K_n} + 2\frac{(C+\frac{1}{2}L(D+2))^2}{\alpha_n^2} + 2e^{-2K_n}. \tag{3.3}$$

Furthermore,

$$E\int_0^\infty\left(\hat{G}_n(z) - G(z)\right)^2 dG(z) = \mathcal{O}\left(\left(\frac{\log\log n}{\log n}\right)^2\right) \tag{3.4}$$

holds as $n \to \infty$, when α_n and K_n are chosen as

$$\alpha_n = \frac{\log n}{\gamma\log\log n}, \quad K_n = \left\lceil\frac{\log n}{\kappa\log\log n}\right\rceil \tag{3.5}$$

with the constants γ and κ satisfying $\gamma \ge 2De^2\kappa,\ \kappa > 2$.

Our proof of this theorem is based on the representation of $\hat{G}_n$ as an average, to wit

$$\hat{G}_n(z) = \frac{1}{n}\sum_{i=1}^{n} B_n(z, X_i) \tag{3.6}$$

with

$$B_n(z,x) = \sum_{k=0}^{\lfloor \alpha_n z \rfloor \wedge x} \binom{x}{k} {\alpha_n}^k \, (1-\alpha_n)^{x-k} \, \mathbf{1}_{[x \le K_n]}. \tag{3.7}$$

Subsequently, both the variance and the bias part of the mean integrated squared error are studied in Appendix A.

4. Randomly Right Censored Data

Suppose now that $X, X_1, \ldots, X_n$ are i.i.d. random variables with distribution $p(x) = P(X = x)$ given by (1.1) and that $Y, Y_1, \ldots, Y_n$ are i.i.d. nonnegative random variables distributed according to some distribution function H. Assume that the X's and Y's are independent and that one observes $Z_i = \min(X_i, Y_i)$ and $\Delta_i = \mathbf{1}_{[X_i \le Y_i]}$ only. We are interested in estimation of the unknown mixing distribution function G in this random censoring model.

It is known that the distribution of the X_i's can be estimated at the same $\sqrt{n}$ rate as in the uncensored case, provided the right censoring is not too strict (cf. [4]). Therefore, it should be possible to estimate the mixing distribution under right censoring at the same rate as without censoring. Our results here confirm this heuristic.

First consider the case where the censoring distribution function H is known. Observe

$$\begin{aligned} P(Z_i = x, \Delta_i = 1) &= P(X_i = x, X_i \le Y_i) \\ &= P(X_i = x)(1 - H(x-)) = p(x)(1-H(x-)), \quad x = 0, 1, \ldots . \end{aligned} \tag{4.1}$$

Consequently, using the observations Z_i and Δ_i, we can estimate $p(x)$ by the following empirical expression:

$$\tilde{p}_n(x) = \frac{1}{1 - H(x-)} \, \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{[Z_i = x, \Delta_i = 1]} \tag{4.2}$$

for those x for which $1 - H(x-)$ is positive. In analogy to (3.1), (3.2) and (3.6) we construct our estimator of the unknown mixing distribution function G as follows. For $\alpha_n > 1$ and K_n a positive integer, we define

$$\begin{aligned} \tilde{G}_n(z) &= \sum_{x=0}^{K_n} \sum_{k=0}^{\lfloor \alpha_n z \rfloor \wedge x} \binom{x}{k} \alpha_n^k (1-\alpha_n)^{x-k} \; \tilde{p}_n(x) \\ &= \frac{1}{n} \sum_{i=1}^{n} \frac{\Delta_i}{1 - H(Z_i -)} \sum_{x=0}^{K_n} \sum_{k=0}^{\lfloor \alpha_n z \rfloor \wedge x} \binom{x}{k} \alpha_n^k (1-\alpha_n)^{x-k} \; \mathbf{1}_{[Z_i = x]} \\ &= \frac{1}{n} \sum_{i=1}^{n} \frac{\Delta_i}{1 - H(Z_i -)} \sum_{k=0}^{\lfloor \alpha_n z \rfloor \wedge Z_i} \binom{Z_i}{k} \alpha_n^k (1-\alpha_n)^{Z_i - k} \; \mathbf{1}_{[Z_i \le K_n]}. \end{aligned} \tag{4.3}$$

Note that this estimator has the form

$$\tilde{G}_n(z) = \frac{1}{n} \sum_{i=1}^{n} \frac{\Delta_i}{1 - H(Z_i -)} B_n(z, Z_i) = \frac{1}{n} \sum_{i=1}^{n} \frac{\Delta_i}{1 - H(X_i -)} B_n(z, X_i), \tag{4.4}$$

where again $B_n(z,x)$ is defined by (3.7).

Studying (4.3), we see that if the censoring random variables Y_i have bounded support, then for $\alpha_n z$ and K_n large our estimator reduces to

$$\tilde{G}_n(z) = \frac{1}{n} \sum_{i=1}^{n} \frac{\Delta_i}{(1 - H(X_i -))} \, ,$$

which by the Law of Large Numbers converges to 1. Consequently, it is crucial for the consistency of our estimator that the right-hand tail of H be not too thin. In fact, no estimator can behave properly if H has bounded support $[0,\tau]$, say, as in this case it is possible to estimate $p(x)$ consistently only for $x = 0, 1, \dots, \lfloor\tau\rfloor$. However, the mixing distribution G is **not** identifiable from the $\lfloor\tau\rfloor + 1$ equations

$$p(x) = \int_0^\infty e^{-\lambda} \frac{\lambda^x}{x!}\, dG(\lambda), \quad x = 0, 1, \dots, \lfloor\tau\rfloor. \tag{4.5}$$

Actually, in Section 5, we construct mixing densities g and g_n such that they differ, but yield the same values of $p(x)$ in (4.5) for $x = 0, 1, \dots, m-5$, where m may be chosen arbitrarily large (see (5.5) up to but not including (5.9)). Because of this unidentifiability phenomenon we will assume that the right-hand tail of the censoring distribution H does not decrease too fast. More precisely, we will assume condition (1.1) from [4].

Assumption 4.1. There exists a finite constant M with

$$\sum_{x=0}^{\infty} \frac{1}{P(Y \geq x)}\, p(x) = \sum_{x=0}^{\infty} \frac{1}{1 - H(x-)}\, p(x) \leq M. \tag{4.6}$$

As $\tilde{G}_n$ and $\hat{G}_n$ are similar averages (cf. (3.6), (4.3) and (4.4)), we can establish the consistency of $\tilde{G}_n$ along the lines of the proof of Theorem 3.1 as given in Appendix A. In the censored case, $\tilde{G}_n$ attains the same rate as $\hat{G}_n$ in the uncensored case.

Theorem 4.1. *Let the conditions of Theorem* 3.1 *be satisfied and let α_n and K_n be chosen as in* (3.5). *If the censoring distribution H is known and fulfills Assumption* 4.1, *then the mean integrated squared error of $\tilde{G}_n$ is of the order* $(\log\log n/\log n)^2$ *as $n \to \infty$. More precisely,*

$$E\int_0^\infty \Big(\tilde{G}_n(z) - G(z)\Big)^2 dG(z) \leq \frac{1}{n}(2\alpha_n)^{2K_n} M + 2\frac{(C + \frac{1}{2}L(D+2))^2}{\alpha_n^2} + 2e^{-2K_n} \tag{4.7}$$

holds.

Proof. First, we estimate the variance of $\tilde{G}_n(z)$ under Assumption 4.1 (see (A.4)) as follows:

$$\begin{aligned}\operatorname{var}\tilde{G}_n(z) &= \frac{1}{n}\operatorname{var}\left(\frac{\Delta_1}{1 - H(Z_1-)} B_n(z, Z_1)\right) \leq \frac{1}{n} E\left(\frac{B_n^2(z,X)}{(1 - H(X-))^2} E(\mathbf{1}_{[X \leq Y]} \mid X)\right)\\ &\leq \frac{1}{n} E\left(\frac{(2\alpha_n)^{2K_n}}{(1 - H(X-))}\right) \leq \frac{1}{n}(2\alpha_n)^{2K_n} M. \end{aligned}\tag{4.8}$$

Furthermore, the bias of $\tilde{G}_n$ equals

$$E\left(\frac{1}{1 - H(X-)} B_n(z,X) E(\mathbf{1}_{[X \leq Y]} \mid X)\right) - G(z) = E\big[B_n(z,X)\big] - G(z),$$

which in view of (A.6) is the same expression as in (A.7). Together with (A.1), (A.2), (4.8) and (A.11) this yields (4.7) and hence the Theorem. □

Next, we consider the case where the survival function $S = 1 - H$ of the censoring variable Y is unknown, but is known to be continuous. Observe that the estimator $\hat{G}_n$ for the non-censored case (cf. (3.6)) can de written as $\hat{G}_n(z) = \int B_n(z,x) d\hat{F}_n(x)$ with $\hat{F}_n$ the empirical distribution function of X. So it is natural in the censored case to consider $\tilde{G}_n^{\mathrm{KM}}(z) = \int B_n(z,x) d\tilde{F}_n(x)$ with $\tilde{F}_n$ the Kaplan-Meier estimator of the distribution function of X. However, we have not been able to study the asymptotic performance of the mean integrated squared error of this estimator of G.

Therefore, we construct another estimator. It is based on the technique of sample splitting as in [10]. To explain the idea, we assume for the time being that we have an extra sample $(\tilde{\Delta}, \tilde{Z}) = ((\tilde{\Delta}_1, \tilde{Z}_1), \dots, (\tilde{\Delta}_n, \tilde{Z}_n))$ available of size n, that is, independent of and identically distributed to

$((\Delta_1, Z_1), \ldots, (\Delta_n, Z_n))$. The product-limit or Kaplan-Meier estimator $\tilde{S}_n$ of the survival function S based on this extra sample is defined as

$$\tilde{S}_n(x) = \begin{cases} 1, & 0 \leq x \leq \tilde{Z}_{(1)}, \\ \prod_{i=1}^{k-1} (\frac{n-i}{n-i+1})^{1-\tilde{\Delta}_{(i)}}, & \tilde{Z}_{(k-1)} < x \leq \tilde{Z}_{(k)}, \quad k = 2, \ldots, n, \\ 0, & \tilde{Z}_{(n)} < x, \end{cases}$$

where $\tilde{Z}_{(i)}$ and $\tilde{\Delta}_{(i)}$ denote the ordered $\tilde{Z}_i$'s and corresponding $\tilde{\Delta}_i$'s. Note that $\tilde{S}_n$ is well defined, as there are no ties among the $\tilde{Z}_{(i)}$'s for which the $\tilde{\Delta}_{(i)}$'s vanish in view of the continuity of H.

We define δ_n and redefine K_n as follows:

$$\delta_n = n^{\frac{1}{2\kappa_0} - \frac{1}{2}} \sqrt{\log n}, \quad K_n = \left\lceil \frac{\log n}{\kappa \log \log n} \right\rceil, \quad \kappa > \kappa_0 > 0. \tag{4.9}$$

Replacing in (4.2), (4.3) and (4.4) the survival function $S = 1 - H$ by its estimator $\tilde{S}_n$ with δ_n added to it, we obtain our estimator

$$\tilde{G}_n^*(z) = \frac{1}{n} \sum_{i=1}^{n} \frac{\Delta_i}{\tilde{S}_n(Z_i-) + \delta_n} B_n(z, Z_i) = \frac{1}{n} \sum_{i=1}^{n} \frac{\Delta_i}{\tilde{S}_n(X_i-) + \delta_n} B_n(z, X_i),$$

which is based on the original sample together with the extra one.

Following the proof of Theorem 4.1, we estimate the variance of $\tilde{G}_n^*(z)$ as follows:

$$\begin{aligned}
\operatorname{var}\left(\tilde{G}_n^*(z) \,\middle|\, \tilde{\Delta}, \tilde{Z}\right) &= \frac{1}{n} \operatorname{var}\left(\frac{\Delta_1}{\tilde{S}_n(X_1-) + \delta_n} B_n(z, X_1) \,\middle|\, \tilde{\Delta}, \tilde{Z}\right) \\
&\leq \frac{1}{n} E\left(\frac{B_n^2(z, X)}{\left(\tilde{S}_n(X-) + \delta_n\right)^2} E(\mathbf{1}_{[X \leq Y]} \mid X) \,\middle|\, \tilde{\Delta}, \tilde{Z}\right) \\
&\leq \frac{1}{n} E\left(\frac{S(X-)(2\alpha_n)^{2K_n}}{\left(\tilde{S}_n(X-) + \delta_n\right)^2} \,\middle|\, \tilde{\Delta}, \tilde{Z}\right) \leq \frac{1}{n} E\left(\frac{S(X-)(2\alpha_n)^{2K_n}}{\left(S(X-)[1 - \tilde{D}_n] + \delta_n\right)^2} \,\middle|\, \tilde{\Delta}, \tilde{Z}\right),
\end{aligned} \tag{4.10}$$

where $\tilde{D}_n$ is defined as

$$\tilde{D}_n = \sup_{0 \leq y \leq K_n} \left| \frac{\tilde{S}_n(y) - S(y)}{S(y)} \right|.$$

From (B.3) in Appendix B we know that for large n

$$-S(X-)\tilde{D}_n + \delta_n \geq 0 \quad \text{almost surely}$$

holds, which combined with (4.10) and Assumption 4.1 yields

$$\operatorname{var}\left(\tilde{G}_n^*(z) \,\middle|\, \tilde{\Delta}, \tilde{Z}\right) \leq \frac{1}{n} E\left(\frac{(2\alpha_n)^{2K_n}}{S(X-)}\right) \leq \frac{1}{n} (2\alpha_n)^{2K_n} M, \quad \text{a.s.} \tag{4.11}$$

Note that for $0 \leq x \leq K_n$ and large n formula (B.3) implies $\tilde{S}_n(x-) + \delta_n \in [S(x-), S(x-) + 2\delta_n]$ a.s. and hence

$$1 - \frac{S(x-)}{\tilde{S}_n(x-) + \delta_n} \in \left[0, \frac{2\delta_n}{S(x-)}\right] \quad \text{a.s.}$$

By (A.4) and Assumption 4.1, this implies that conditionally on the extra sample the bias of our estimator satisfies

$$\begin{aligned}
\left| E\, B_n(z, X) - E\left(\tilde{G}_n^*(z) \mid \tilde{\Delta}, \tilde{Z}\right) \right| &\leq E\left(\left| \left[1 - \frac{S(X-)}{\tilde{S}_n(X-) + \delta_n}\right] \ B_n(z, X) \right| \mid \tilde{\Delta}, \tilde{Z}\right) \\
&\leq 2\delta_n (2\alpha_n)^{K_n} E\left(\frac{1}{S(X-)}\right) \leq 2\delta_n (2\alpha_n)^{K_n} M \quad \text{a.s.}
\end{aligned}$$

Together with (A.11) this means (see also (4.9))

$$\int_0^\infty \left[E\left(\tilde{G}_n^*(z) \mid \tilde{\Delta}, \tilde{Z}\right) - G(z)\right]^2 dG(z)$$

$$\leq 4\frac{\log n}{n^{1-1/\kappa_0}}(2\alpha_n)^{2K_n}M^2 + 4\frac{(C+\frac{1}{2}L(D+2))^2}{\alpha_n^2} + 4e^{-2K_n}, \quad \text{a.s.},$$

which in combination with (4.11) results in

$$E\int_0^\infty \left(\tilde{G}_n^*(z) - G(z)\right)^2 dG(z)$$

$$\leq \frac{1}{n}(2\alpha_n)^{2K_n}M + 4\frac{\log n}{n^{1-1/\kappa_0}}(2\alpha_n)^{2K_n}M^2 + 4\frac{(C+\frac{1}{2}L(D+2))^2}{\alpha_n^2} + 4e^{-2K_n}.$$

With α_n and K_n defined as in (3.5) and (4.9), where the constants γ and κ satisfy $\gamma \geq 2De^2\kappa$, $\kappa > \kappa_0 = 3$, we obtain

$$\left(\frac{\log n}{\log\log n}\right)^2 E\int_0^\infty \left(\tilde{G}_n^*(z) - G(z)\right)^2 dG(z) = \mathcal{O}(1),$$

as n tends to infinity.

However, $\tilde{G}_n^*$ is based on $2n$ observations, as it is based on the original and the extra samples. But, if one has a sample of n observations available, the rate $\log n/(\log\log n)$ can still be obtained by splitting the sample into two subsamples of about the same size and applying the natural modification of our construction in order to get an estimate of G. Here, the first subsample plays the role of the original sample in our construction and the second subsample the role of the extra sample. Interchanging the roles of the two subsamples one gets another estimate of G and it makes sense to average these two estimates to obtain $\tilde{G}_n^{**}$ (cf. [13], Section 2 of [10], or page 396 of [1]). In summary,

Theorem 4.2. *Fix positive constants C, D, L, and β with $CD \geq 1$. Consider the class of mixing distributions G that have support contained in $[0, D]$ and have a density bounded by C that is Lipschitz continuous with Lipschitz constant L. For the class of censoring distributions H that are continuous and fulfill Assumption 4.1, there exists an estimator $\tilde{G}_n^{**}$ of G based on $(\Delta_1, Z_1), \ldots, (\Delta_n, Z_n)$, for which the mean integrated squared error is of the order $(\log\log n/\log n)^2$.*

5. Lower Bound to the Mean Integrated Squared Error

Information in the data about the mixing distribution G in the right censored case equals at most the information in the uncensored case. Therefore, the optimal lower bound to the mean integrated squared error for estimators of G in the censored case should have a convergence rate at most as large as the rate in the uncensored case. As we have seen in the preceding Section, our estimators for these two cases attain the same convergence rate $(\log n/\log\log n)^2$. In this Section we shall prove that the convergence rate equals at most this rate $(\log n/\log\log n)^2$ in the uncensored case, and hence in the censored case. Thus we have shown that our estimators attain the optimal rate and that our bound on the rate is also optimal, in both the uncensored and censored cases.

We study the minimax risk and note that it is bounded from below by a Bayes risk. Namely, we have

$$\inf_{\hat{G}_n}\sup_{G} E_G \int_0^\infty \left(\hat{G}_n(\lambda; X) - G(\lambda)\right)^2 dG(\lambda)$$

$$= \inf_{\hat{G}_n}\sup_{\alpha, G_0, G_n} \left\{\alpha E_{G_0}\int_0^\infty \left(\hat{G}_n(\lambda; X) - G_0(\lambda)\right)^2 dG_0(\lambda)\right.$$

$$+ (1-\alpha) E_{G_n} \int_0^\infty \left(\hat{G}_n(\lambda; X) - G_n(\lambda) \right)^2 dG_n(\lambda) \Big\}$$

$$\geq \sup_{\alpha, G_0, G_n} \inf_{\hat{G}_n} \Big\{ \alpha E_{G_0} \int_0^\infty \left(\hat{G}_n(\lambda; X) - G_0(\lambda) \right)^2 g_0(\lambda) d\lambda$$

$$+ (1-\alpha) E_{G_n} \int_0^\infty \left(\hat{G}_n(\lambda; X) - G_n(\lambda) \right)^2 g_n(\lambda) d\lambda \Big\}, \tag{5.1}$$

where the G's are supposed to have densities g with respect to the Lebesgue measure. We introduce the notation

$$p_{n0}(x) = p_{n0}(x_1, \ldots, x_n) = \prod_{i=1}^n \left(\int_0^\infty e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} g_0(\lambda) d\lambda \right) \tag{5.2}$$

and similarly for $p_{nn}(x)$. Now the right-hand side of (5.1) can be written as

$$\sup_{\alpha, g_0, g_n} \inf_{\hat{G}_n} \sum_x \int_0^\infty \Big\{ \alpha \left(\hat{G}_n(\lambda; x) - G_0(\lambda) \right)^2 p_{n0}(x) g_0(\lambda)$$

$$+ (1-\alpha) \left(\hat{G}_n(\lambda; x) - G_n(\lambda) \right)^2 p_{nn}(x) g_n(\lambda) \Big\} d\lambda$$

and this infimum is attained by

$$\hat{G}_n(\lambda; x) = \frac{G_0(\lambda) \alpha p_{n0}(x) g_0(\lambda) + G_n(\lambda)(1-\alpha) p_{nn}(x) g_n(\lambda)}{\alpha p_{n0}(x) g_0(\lambda) + (1-\alpha) p_{nn}(x) g_n(\lambda)},$$

which results into

$$\sup_{\alpha, g_0, g_n} \int_0^\infty (G_0(\lambda) - G_n(\lambda))^2 \sum_x \frac{\alpha(1-\alpha) p_{n0}(x) p_{nn}(x) g_0(\lambda) g_n(\lambda)}{\alpha p_{n0}(x) g_0(\lambda) + (1-\alpha) p_{nn}(x) g_n(\lambda)} d\lambda. \tag{5.3}$$

For positive reals s and t we have $st/(s+t) \geq \frac{1}{2}(s \wedge t)$. Consequently, the right-hand side of (5.3) is bounded from below by

$$\sup_{\alpha, g_0, g_n} \frac{1}{2} \int_0^\infty (G_0(\lambda) - G_n(\lambda))^2 \sum_x \left\{ (\alpha p_{n0}(x) g_0(\lambda)) \wedge ((1-\alpha) p_{nn}(x) g_n(\lambda)) \right\} d\lambda,$$

which for $\alpha = \frac{1}{2}$ and combined with (5.1) through (5.3) results in

$$\inf_{\hat{G}_n} \sup_G E_G \int_0^\infty \left(\hat{G}_n(\lambda; X) - G(\lambda) \right)^2 dG(\lambda)$$

$$\geq \sup_{g_0, g_n} \frac{1}{4} \int_0^\infty (G_0(\lambda) - G_n(\lambda))^2 (g_0(\lambda) \wedge g_n(\lambda)) \, d\lambda \sum_x p_{n0}(x) \wedge p_{nn}(x). \tag{5.4}$$

In order to come close to this supremum one has to choose g_0 and g_n in such a way that p_{n0} and p_{nn} are close together and that simultaneously G_0 and G_n are as different as possible. We shall choose g_0 and g_n with the help of the orthogonal system of Chebyshev polynomials C_m on $[-1,1]$, $m = 0, 1, \ldots$. They are defined as

$$C_m(z) = \cos(m \arccos z), \quad -1 \leq z \leq 1,$$

and are orthogonal with respect to the weight function $1/\sqrt{1-z^2}$, $-1 < z < 1$. Now we choose

$$g_0(\lambda) = \mathbf{1}_{(0,1)}(\lambda), \quad \lambda > 0,$$

$$H_n(\lambda) = \int_0^\lambda h_n(\mu) g_0(\mu) d\mu = [\lambda(1-\lambda)]^{3/2} e^\lambda C_m(2\lambda - 1), \tag{5.5}$$
$$g_n(\lambda) = g_0(\lambda)(1 + a_n h_n(\lambda)),$$

where $m = m_n$ depends on n in an appropriate way to be determined below. By differentiation, for $0 < \lambda < 1$, we obtain

$$h_n(\lambda) = \tfrac{1}{2}\sqrt{\lambda(1-\lambda)}(3 - 4\lambda - 2\lambda^2) e^\lambda C_m(2\lambda - 1) \\ - m\lambda(1-\lambda) e^\lambda \sin(m \arccos(2\lambda - 1)),$$

which we may bound by

$$|h_n(\lambda)| < \tfrac{1}{4} e(m+3). \tag{5.6}$$

In view of

$$\int_0^1 h_n(\lambda) g_0(\lambda) d\lambda = H_n(1) = 0,$$

equations (5.5) and (5.6) imply that g_n is a proper density provided

$$|a_n| \le \frac{4}{e(m+3)} \tag{5.7}$$

holds. With (5.5) in mind, by partial integration, we compute

$$\int_0^\infty e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} h_n(\lambda) g_0(\lambda) d\lambda = \left[e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} H_n(\lambda) \right]_0^1 - \int_0^1 e^{-\lambda} \frac{\lambda^{x_i - 1}}{x_i!} (x_i - \lambda) H_n(\lambda) d\lambda$$
$$= \int_0^1 \frac{\lambda^{x_i - 1}}{x_i!} (\lambda - x_i) \frac{2\lambda^2 (1-\lambda)^2}{\sqrt{1 - (2\lambda - 1)^2}} C_m(2\lambda - 1) d\lambda$$
$$= \int_{-1}^1 \frac{(1+z)^{x_i - 1}}{2^{x_i + 4} x_i!} (z + 1 - 2x_i) \frac{(1 - z^2)^2}{\sqrt{1 - z^2}} C_m(z) dz. \tag{5.8}$$

As the Chebyshev polynomial of degree m is orthogonal with respect to the weight function $1/\sqrt{1 - z^2}$, $-1 < z < 1$, to all polynomials of degree at most $m - 1$, the integrals in (5.8) vanish for $x_i \le m - 5$. Hence,

$$p_{nn}(x) = p_{n0}(x) \tag{5.9}$$

holds (cf. (5.2)), unless at least one of the x_i's equals $m - 4$ or more. Actually, the probability q_n that X_i equals at least $m - 4$, may be bounded both under g_0 and g_n via

$$q_n = P(X_i \ge m - 4) = \int_0^1 \sum_{k=m-4}^\infty e^{-\lambda} \frac{\lambda^k}{k!} d\lambda \le \int_0^1 \frac{\lambda^{m-4}}{(m-4)!} d\lambda = \frac{1}{(m-3)!}. \tag{5.10}$$

Let Z_n be the random variable denoting the number of X_i that equal at least $m - 4$. Note that Z_n has a binomial distribution with parameters n and q_n.

Combining (5.6) and (5.10), we arrive at

$$\sum_x p_{n0}(x) \wedge p_{nn}(x) \ge E_{g_0} \left(\left(1 - a_n \sup_{0<\lambda<1} |h_n(\lambda)| \right)^{Z_n} \right)$$
$$= \left(1 - q_n a_n \sup_{0<\lambda<1} |h_n(\lambda)| \right)^n \ge \left(1 - \frac{e a_n (m+3)}{4(m-3)!} \right)^n, \tag{5.11}$$

which converges to $1/\sqrt{e}$ by the choice

$$a_n = \frac{2(m-3)!}{en(m+3)}. \tag{5.12}$$

For the time being we assume that

$$n = (m-3)! \tag{5.13}$$

holds. By Stirling's formula this means

$$\lim_{m\to\infty} \frac{(m-3)\log\log n}{\log n} = 1. \tag{5.14}$$

Note that these choices of n, m and a_n satisfy (5.7). Some computation shows that all together the above choices imply

$$\begin{aligned}
\int_0^\infty (G_0(\lambda) - G_n(\lambda))^2 (g_0(\lambda) \wedge g_n(\lambda))\, d\lambda &= \int_0^\infty a_n^2 H_n^2(\lambda) (g_0(\lambda) \wedge g_n(\lambda))\, d\lambda \\
&\geq \tfrac{1}{2} a_n^2 \int_0^1 \lambda^3 (1-\lambda)^3 e^{2\lambda} \cos^2(m \arccos(2\lambda - 1)) d\lambda \\
&= 2^{-8} a_n^2 \int_0^\pi [(1+\cos\alpha)(1-\cos\alpha)]^3 e^{1+\cos\alpha} \cos^2(m\alpha) \sin\alpha\, d\alpha \\
&\geq 2^{-8} a_n^2 \int_0^\pi \sin^7 \alpha \cos^2(m\alpha)\, d\alpha \geq 2^{-23/2} a_n^2 \int_{\pi/4}^{3\pi/4} \cos^2(m\alpha)\, d\alpha.
\end{aligned} \tag{5.15}$$

Because this last integral converges to $\pi/4$ as m tends to infinity, the relations (5.15), (5.12), (5.13), and (5.14) imply

$$\liminf_{n=(m-3)!, m\to\infty} \left(\frac{\log n}{\log\log n}\right)^2 \int_0^\infty (G_0(\lambda) - G_n(\lambda))^2 (g_0(\lambda) \wedge g_n(\lambda))\, d\lambda \geq 2^{-15/2} \pi e^2 > 0.$$

Together with (5.11), this yields

$$\begin{aligned}
\liminf_{n=(m-3)!, m\to\infty} \left(\frac{\log n}{\log\log n}\right)^2 &\int_0^\infty (G_0(\lambda) - G_n(\lambda))^2 (g_0(\lambda) \wedge g_n(\lambda))\, d\lambda \\
&\sum_x p_{n0}(x) \wedge p_{nn}(x) \geq 2^{-15/2} \pi e^{3/2} > 0.
\end{aligned} \tag{5.16}$$

If $\tilde{n}$ satisfies $n = (m-3)! \leq \tilde{n} < (m-2)!$, then $1 \leq \tilde{n}/n < m-2$ holds and hence

$$1 \leq \frac{\log \tilde{n}}{\log n} < 1 + \frac{\log(m-2)}{\log n} \to 1, \quad \text{as } m \to \infty. \tag{5.17}$$

Combining (5.4), (5.16) and (5.17), we obtain the following lower bound.

Theorem 5.1. *Let $X_1, \ldots, X_n$ be i.i.d. random variables with the Poisson mixture distribution* (1.1) *from the class of mixing distributions G that have density bounded by $C \in [2, \infty)$ and have $G(D) = 1$ for some $D \in [1, \infty)$. With $\hat{G}_n$ an estimator of G based on $X_1, \ldots, X_n$ the minimax value of the mean integrated squared error of $\hat{G}_n$ in estimating G does not tend to 0 faster than* $(\log\log n/\log n)^2$ *as n tends to infinity, more precisely,*

$$\liminf_{n\to\infty} \left(\frac{\log n}{\log\log n}\right)^2 \inf_{\hat{G}_n} \sup_G E_G \int_0^\infty \left(\hat{G}_n(\lambda; X) - G(\lambda)\right)^2 dG(\lambda) > 0.$$

Appendix A. Proof of Theorem 3.1

The proof in this Appendix of Theorem 3.1 will be based on (3.6) with (3.7). In view of

$$E\int_0^\infty \left(\hat{G}_n(z) - G(z)\right)^2 dG(z) = \int_0^\infty E\left(\hat{G}_n(z) - G(z)\right)^2 dG(z), \tag{A.1}$$

we first fix z, study

$$E\left(\hat{G}_n(z) - G(z)\right)^2 = \text{var}\ \hat{G}_n(z) + \left(E\hat{G}_n(z) - G(z)\right)^2 \tag{A.2}$$

and note that

$$\text{var}\ \hat{G}_n(z) = \frac{1}{n}\ \text{var}\ B_n(z, X). \tag{A.3}$$

As $B_n(z, x)$ from (3.7) satisfies

$$|B_n(z,x)| \le \sum_{k=0}^{x} \binom{x}{k} \alpha_n^k \left|1 - \alpha_n\right|^{x-k} \mathbf{1}_{[x \le K_n]} \le (2\alpha_n)^{K_n}, \tag{A.4}$$

we have

$$\text{var}\ B_n(z, X) \le E\left(B_n^2(z, X)\right) \le (2\alpha_n)^{2K_n}. \tag{A.5}$$

The study of the bias is more involved. We choose the random variable Λ with a distribution function G in such a way that the conditional distribution of X given Λ is Poisson (Λ); so

$$P(X = x \mid \Lambda = \lambda) = e^{-\lambda}\,\frac{\lambda^x}{x!}, \quad x = 0, 1, \dots .$$

By Taylor's theorem (or partial integration),

$$e^x = \sum_{k=0}^{K} \frac{x^k}{k!} + \int_0^x \frac{(x-y)^K}{K!} e^y dy$$

holds. Consequently, we have

$$\begin{aligned}
E\left(B_n(z, X) \,|\, \Lambda = \lambda\right) &= \sum_{x=0}^{K_n} \sum_{k=0}^{\lfloor \alpha_n z \rfloor \wedge x} \binom{x}{k} {\alpha_n}^k \left(1 - \alpha_n\right)^{x-k} e^{-\lambda}\,\frac{\lambda^x}{x!} \\
&= \sum_{k=0}^{\lfloor \alpha_n z \rfloor \wedge K_n} e^{-\lambda}\,\frac{(\alpha_n\lambda)^k}{k!} \sum_{x=k}^{K_n} \frac{1}{(x-k)!} \left((1-\alpha_n)\lambda\right)^{x-k} \\
&= \sum_{k=0}^{\lfloor \alpha_n z \rfloor \wedge K_n} e^{-\alpha_n\lambda}\,\frac{(\alpha_n\lambda)^k}{k!} \\
&\quad - \sum_{k=0}^{\lfloor \alpha_n z \rfloor \wedge K_n} e^{-\lambda}\,\frac{(\alpha_n\lambda)^k}{k!} \int_0^{(1-\alpha_n)\lambda} \frac{((1-\alpha_n)\lambda - y)^{K_n - k}}{(K_n - k)!}\, e^y\, dy \\
&= P\left(U_{\alpha_n\lambda} \le \alpha_n z \wedge K_n\right) - R_n(z, \lambda)
\end{aligned} \tag{A.6}$$

with U_μ, distributed as Poisson(μ). In view of $G(D) = 1$, only z that are at most D, are relevant and for such z we have $\alpha_n z < K_n$. Hence, the bias of $\hat{G}_n$ equals

$$E\hat{G}_n(z) - G(z) = \int_0^\infty \left(P(U_{\alpha_n\lambda} \le \alpha_n z) - \mathbf{1}_{[\lambda \le z]}\right) dG(\lambda) - \int_0^\infty R_n(z,\lambda) dG(\lambda). \tag{A.7}$$

First, we note

$$|R_n(z,\lambda)| \le \sum_{k=0}^{\lfloor \alpha_n z \rfloor} e^{-\lambda}\,\frac{(\alpha_n\lambda)^k}{k!} \frac{((\alpha_n - 1)\lambda)^{K_n - k}}{(K_n - k)!} \le e^{-\lambda}\,\frac{(\alpha_n\lambda)^{K_n}}{K_n!} \sum_{k=0}^{\lfloor \alpha_n z \rfloor} \binom{K_n}{k}$$

$$\leq e^{-\lambda}\,\frac{(2\alpha_n\lambda)^{K_n}}{K_n!} \leq \frac{e^{-\lambda}}{\sqrt{2\pi K_n}}\left(\frac{2\alpha_n\lambda e}{K_n}\right)^{K_n} \leq e^{-K_n}, \tag{A.8}$$

where the second to last inequality stems from Stirling's formula.

Furthermore, we note

$$G(0)=0, \quad \lim_{\lambda\to\infty} P(U_{\alpha_n\lambda} \leq \alpha_n z) = 0$$

and

$$\frac{\partial}{\partial\lambda}P(U_{\alpha_n\lambda} \leq \alpha_n z) = -\alpha_n e^{-\alpha_n\lambda}\,\frac{(\alpha_n\lambda)^{\lfloor\alpha_n z\rfloor}}{\lfloor\alpha_n z\rfloor!}.$$

Consequently, partial integration yields

$$\int_0^\infty P(U_{\alpha_n\lambda} \leq \alpha_n z)\,dG(\lambda) = \int_0^\infty G(\lambda)\alpha_n e^{-\alpha_n\lambda}\,\frac{(\alpha_n\lambda)^{\lfloor\alpha_n z\rfloor}}{\lfloor\alpha_n z\rfloor!}\,d\lambda = EG(\Lambda_n), \tag{A.9}$$

where Λ_n has a gamma distribution with shape parameter $\lfloor\alpha_n z\rfloor+1$ and rate parameter α_n. Hence we have

$$E\Lambda_n - z = \frac{1+\lfloor\alpha_n z\rfloor-\alpha_n z}{\alpha_n} \in (0, 1/\alpha_n], \quad \mathrm{var}(\Lambda_n) = \frac{\lfloor\alpha_n z\rfloor+1}{\alpha_n^2},$$

$$E(\Lambda_n-z)^2 = \frac{\lfloor\alpha_n z\rfloor+1}{\alpha_n^2} + \left(\frac{1+\lfloor\alpha_n z\rfloor-\alpha_n z}{\alpha_n}\right)^2 \leq \frac{z+2}{\alpha_n}.$$

As G has a density g that is Lipschitz continuous with Lipschitz constant L, we have

$$|G(\lambda)-G(z)-(\lambda-z)g(z)| = \left|\int_z^\lambda (g(y)-g(z))\,dy\right| \leq \tfrac{1}{2}L(\lambda-z)^2. \tag{A.10}$$

Equations (A.9) through (A.10) yield

$$\left|\int_0^\infty \left(P(U_{\alpha_n\lambda} \leq \alpha_n z) - \mathbf{1}_{[\lambda\leq z]}\right)dG(\lambda)\right| \leq \frac{g(z)+\frac{1}{2}L(z+2)}{\alpha_n},$$

which together with (A.7) and (A.8) shows that the bias of $\hat{G}_n(z)$ satisfies

$$\left(E\hat{G}_n(z)-G(z)\right)^2 = \left(EB_n(z,X)-G(z)\right)^2$$

$$\leq 2\frac{(g(z)+\frac{1}{2}L(z+2))^2}{\alpha_n^2} + 2e^{-2K_n} \leq 2\frac{(C+\frac{1}{2}L(D+2))^2}{\alpha_n^2} + 2e^{-2K_n}, \quad z\leq D. \tag{A.11}$$

Together with (A.1)–(A.3) and (A.5) this yields (3.3) and consequently (3.4) when α_n is chosen as in (3.5). We have chosen $\kappa>2$ because the first term at the right hand side of (3.3) is of the order $n^{-1+2/\kappa}$.

Appendix B. Kaplan–Meier

In Section 4, we have used the Kaplan-Meier estimator $\tilde{S}_n$ of the survival function S of the censoring distribution H. In this appendix we study the consistency of $\tilde{S}_n$ by applying Theorem 7 of [5]. We choose their d_n to be equal to $\log n/(\delta\log\log n)$ and their T_n to our K_n. Their ε_n is related to K_n via

$$\varepsilon_n = 8P(X\wedge Y > K_n). \tag{B.1}$$

As $\sum_{x\geq K_n+1} e^{-\lambda}\lambda^x/x!$ is decreasing in λ for $\lambda\leq K_n+1$, the support of G is contained in $[0,D]$, and $K_n+1>D$ holds for n large, we have for such n

$$P(X>K_n) = \int_0^\infty \sum_{x=K_n+1}^\infty e^{-\lambda}\frac{\lambda^x}{x!}dG(\lambda) \geq C\int_0^{1/C} \sum_{x=K_n+1}^\infty e^{-\lambda}\frac{\lambda^x}{x!}\,d\lambda$$

$$\geq \sum_{x=K_n+1}^{\infty} e^{-1/C} \frac{(1/C)^x}{x!} \geq e^{-1/C} \frac{(1/C)^{K_n+1}}{(K_n+1)!}. \tag{B.2}$$

From (B.1), (B.2) and Assumption 4.1, we derive

$$\varepsilon_n \geq 8e^{-1/C} \frac{(1/C)^{K_n+1}}{(K_n+1)!} \beta^{-K_n}.$$

With K_n as in (4.9), some computation with the help of Stirling's formula shows

$$\varepsilon_n \geq \exp\left(-(\log n)/\kappa(1+o(1))\right),$$

which in view of $\kappa > \kappa_0$ implies that for sufficiently large n

$$\varepsilon_{2n} \geq n^{-1/\kappa_0}$$

holds. Now, formula (4.16) from Theorem 7 of [5] shows that almost surely

$$\sqrt{\frac{n^{1-1/\kappa_0}}{\log n}}\ \tilde{D}_n = \sqrt{\frac{n^{1-1/\kappa_0}}{\log n}} \sup_{0 \leq y \leq K_n} \left| \frac{\tilde{S}_n(y) - S(y)}{S(y)} \right| \to 0 \tag{B.3}$$

holds as n tends to infinity.

Acknowledgement

This paper has been prepared under NSF grant DMS-0906639, partly during a visit of the first author to the Department of Statistics, West Virginia University, in August 2010. The excellent suggestions by the referee are gratefully acknowledged.

References

1. P. J. Bickel, C. A. J. Klaassen, Y. Ritov, J. A. Wellner, *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins Series in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, MD, 1993.
2. W. Feller, *An Introduction to Probability Theory and Its Applications*. vol. II. Second edition John Wiley & Sons, Inc., New York-London-Sydney, 1971.
3. E. I. Fredholm, Sur une classe d'équations fonctionnelles. (French) *Acta Math.* **27** (1903), no. 1, 365–390. doi:10.1007/bf02421317
4. R. D. Gill, Large sample behaviour of the product-limit estimator on the whole line. *Ann. Statist.* **11** (1983), no. 1, 49–58.
5. E.Giné, A. Guillou, Laws of the iterated logarithm for censored data. *Ann. Probab.* **27** (1999), no. 4, 2042–2067.
6. E. V. Khmaladze, *The statistical Analysis of a Large Number of Rare Events*. CWI Report MS-R8804, 1988.
7. E. V. Khmaladze, Convergence properties in certain occupancy problems including the Karlin–Rouault law. *J. Appl. Probab.* **48** (2011), no. 4, 1095–1113.
8. E. V. Khmaladze, R. Ya. Chitashvili, Statistical analysis of a large number of rare events and related problems. Probability theory and mathematical statistics. (Russian) *Trudy Tbiliss. Mat. Inst. Razmadze Akad. Nauk Gruzin. SSR* **92** (1989), 196–245.
9. E. V. Khmaladze, Z. P. Tsigroshvili, On polynomial distributions with a large number of rare events. *Math. Methods Statist.* **2** (1993), no. 3, 240–247.
10. C. A. J. Klaassen, Consistent estimation of the influence function of locally asymptotically linear estimators. *Ann. Statist.* **15** (1987), no. 4, 1548–1562.
11. C. A. J. Klaassen, R. M. Mnatsakanov, Consistent estimation of the structural distribution function. *Scand. J. Statist.* **27** (2000), no. 4, 733–746.
12. H. Robbins, An empirical Bayes approach to statistics. In: *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 1954-1955, vol. I, pp. 157–163. University of California Press, Berkeley and Los Angeles, 1956.
13. A. Schick, On asymptotically efficient estimation in semiparametric models. *Ann. Statist.* **14** (1986), no. 3, 1139–1151.
14. B. Van Es, C. A. J. Klaassen, R. M. Mnatsakanov, Estimating the structural distribution function of cell probabilities. *Austrian Journal of Statistics* **32** (2003), 85–98.

(Received 30.10.2019)

[1]Korteweg-de Vries Institute for Mathematics, University of Amsterdam, P.O. Box 94248 1090 GE, The Netherlands
E-mail address: C.A.J.Klaassen@uva.nl

[2]Department of Mathematics, West Virginia University, P.O. Box 6310, Morgantown, WV 26506, USA
E-mail address: Robert.Mnatsakanov@mail.wvu.edu