

The International Journal of Biostatistics

Volume 6, Issue 1

2010

Article 9

A Corrected Likelihood Approach for the Nonlinear Transformation Model with Application to Fluorescence Lifetime Measurements Using Exponential Mixtures

Tabea Rebafka*

François Roueff[†]

Antoine Souloumiac[‡]

*Institut Télécom, Télécom ParisTech (CNRS LTCl) and CEA, LIST, rebafka@telecom-paristech.fr

[†]Institut Télécom, Télécom ParisTech (CNRS LTCl), roueff@telecom-paristech.fr

[‡]CEA, LIST, antoine.souloumiac@cea.fr

A Corrected Likelihood Approach for the Nonlinear Transformation Model with Application to Fluorescence Lifetime Measurements Using Exponential Mixtures*

Tabea Rebafka, François Roueff, and Antoine Souloumiac

Abstract

A fast and efficient estimation method is proposed that compensates the distortion in nonlinear transformation models. A likelihood-based estimator is developed that can be computed by an EM-type algorithm. The consistency of the estimator is shown and its limit distribution is provided. The new estimator is particularly well suited for fluorescence lifetime measurements, where only the shortest arrival time of a random number of emitted fluorescence photons can be detected and where arrival times are often modeled by a mixture of exponential distributions. The method is evaluated on real and synthetic data. Compared to currently used methods in fluorescence, the new estimator should allow a reduction of the acquisition time of an order of magnitude.

KEYWORDS: pile-up model, nonlinear transformation model, likelihood-based contrast, EM algorithm, time-resolved fluorescence, exponential mixtures

*The authors wish to thank PicoQuant GmbH, Berlin, Germany for kindly providing the TCSPC data used in Section 5.

1 Introduction

In this paper we consider nonlinear transformation models where the observed distribution is the result of a nonlinear distortion of some initial distribution (Tsodikov, 2003). A special case is the so-called pile-up model, which will be of particular interest in this paper, and where an observation is defined as the minimum of a random number of independent variables from the target distribution. The goal is to estimate the parameters of the target distribution from a sample of the distorted distribution. However, the nonlinear distortion generally makes estimation difficult even for simple target distributions.

This work is motivated by an application in time-resolved fluorescence where a specific pile-up model is encountered. Fluorescence is the emission of photons by excited molecules and one of its characteristics is the duration that a molecule stays in the excited state before emitting a fluorescence photon. This duration is called the *fluorescence lifetime* and it is well known that these lifetimes have exponential distribution whose parameter depends on the fluorescent molecule as well as on its microenvironment as pH, viscosity or polarity (Lakowicz, 1999; Valeur, 2002). Due to the high sensitivity of the exponential parameters on the microenvironment, fluorescence lifetimes are a precious source of information on molecular processes and they are used in many applications in biology, medicine or chemistry. For instance, in biochemical applications of fluorescence imaging the effect of the environment on the fluorescence lifetimes is used to map chemical or physical changes within a sample (Crissman & Steinkamp, 2000). Further examples are the measurement of molecular distances (Pin *et al.*, 2007) or the measurement of molecular rotation (Serdyuk *et al.*, 2007) based on fluorescence lifetimes.

Measurements of fluorescence lifetimes are obtained by the technique Time-Correlated Single Photon Counting (TCSPC) (O'Connor & Phillips, 1984). First molecules are excited with a short laser pulse and then a random number of fluorescence photons is emitted and hit the detector. The instrument measures the time between the laser pulse and the arrival of the *first* fluorescence photon on the detector. For technical reasons the arrival times of later arriving photons can not be measured. Indeed, the distribution of the minimum arrival time is a distortion of the distribution of the arrival times of *all* photons striking the detector referred to as the *pile-up effect*.

The extent of distortion depends on the laser intensity which determines the average number λ of photons per excitation cycle. In fact, the higher the laser intensity, the more fluorescence photons are emitted. It is standard practice to discard the pile-up effect by using a very low laser intensity.

However, the study of the Fisher information as a function of λ in Rebafka *et al.* (2008) has revealed that the information can be maximized if data are collected at an intensity that causes a significant pile-up effect. That is, this study suggests that a significant reduction of the variance of the estimator compared to the standard practice can be obtained by using pile-up affected observations. Due to the involved form of the pile-up density standard estimation procedures as the maximum likelihood estimator or moment estimators are intractable. Hence the question about an estimator which is numerically achievable and yet performs well in comparison with the information bound arises in order to reduce the variance.

The very first method dealing with pile-up observations is a ‘correction’ of a pile-up histogram that goes back to Coates (1968). This approach is refined and generalized in Walker (2002) and Souloumiac (2007). Okano *et al.* (2005) propose a least-squares fitting of the histogram. It is well known that least-squares methods are large sample methods that are unbiased when applied to small samples, as pointed out by Hall & Selinger (1981) for the particular fluorescence context. This method is hence not appropriate to reduce the variance. In Rebafka (2007) a Gibbs sampler is presented that is adapted to pile-up models with multi-exponential target distributions. This algorithm is rather time-consuming and hence it is not appropriate to analyze a large number of data sets, which is often required in practice.

The pile-up model has also been used in carcinogenesis studies. The model assumes that at the end of a cancer treatment there remains a random number of cells that will propagate into a new detectable tumor. An individual random time is associated to each cell representing the time it takes for this cell to produce a detectable tumor. Then the time to tumor recurrence is the minimum of those cell individual times. In a parametric setting Yakovlev & Tsodikov (1996) compute the maximum likelihood estimator by some random search algorithm.

The paper is organized as follows. The formal definition of the nonlinear transformation model and the pile-up model is given in Section 2. In Section 3 we develop a likelihood-based contrast, whose maximization complexity is essentially the same as the likelihood associated with the target distribution. In particular, if the EM algorithm applies to the likelihood of the target distribution, then it applies similarly to the new contrast. The asymptotic behavior of the new estimator is analyzed in Section 4. The numerical performance is evaluated in Section 5, where an application on TCSPC measurements is provided as well as a comparison to the information bounds obtained in Rebafka *et al.* (2008). Appendix A contains the technical arguments for the results presented in the previous sections, while

Appendix B provides details on the derivation of the central limit theorem for L -statistics that is used in Section 4.

2 General Setting and Notation

We first define the pile-up model, then we generalize the definition to the nonlinear transformation model. Let $\{Y_k, k \geq 1\}$ be a sequence of independent positive random variables with distribution function F and survival function $\bar{F} = 1 - F$. Denote by N a random variable that is independent of the sequence $\{Y_k, k \geq 1\}$ taking its values in $\mathbb{N}^* = \{1, 2, \dots\}$. Each pile-up observation Z_i for $i = 1, 2, \dots, n$ is distributed as the random variable Z defined by

$$Z = \min\{Y_1, \dots, Y_N\} . \tag{1}$$

By Rebafka *et al.* (2008), Lemma 1 the survival function $\bar{G} = 1 - G$ of Z is given by

$$\bar{G}(z) = \gamma(\bar{F}(z)) , \quad z \in \mathbb{R}_+ , \tag{2}$$

where γ is the probability generating function associated with N , defined as $\gamma(u) = \mathbb{E}[u^N]$ for all $u \in [0, 1]$. Moreover, if F admits a density f with respect to the Lebesgue measure \mathcal{L}_+ defined on \mathbb{R}_+ , then G admits a density g . Denoting $\dot{\gamma}(u) = \mathbb{E}[Nu^{N-1}]$ for all $u \in [0, 1]$, the pile-up density g is given by

$$g(z) = f(z)\dot{\gamma}(\bar{F}(z)) , \quad z \in \mathbb{R}_+ . \tag{3}$$

In survival analysis the class of models defined by (2), where γ is the probability generating function of any nonnegative random variable N , is the family of proportional hazard mixture models, also called *univariate frailty models* (Kosorok *et al.*, 2004; Hougaard, 1984). Note that in this case the hazard function μ_G of G given N can be written as

$$\mu_G(t|N) = N\mu_F(t) ,$$

where μ_F is the hazard function of F and N is the frailty. Univariate frailty models are an important subclass of *nonlinear transformation models* defined by (2) for any function γ such that $\gamma \circ \bar{F}$ is a survival function (Tsodikov, 2003).

We will estimate the target distribution F based on a sample of the distorted distribution G in a parametric setting. Our approach applies to models where γ is a known function that is sufficiently smooth. The most general setting that will be considered is described in the following assumption that is supposed to hold throughout the paper.

Assumption 1. The *target distribution* F belongs to an identifiable parametric family dominated by \mathcal{L}_+ and is thus described by a collection of densities $\{f_\theta, \theta \in \Theta\}$ with parameter set $\Theta \subset \mathbb{R}^d$. The function $\gamma : [0, 1] \rightarrow [0, 1]$ has range $[0, 1]$ and is continuously differentiable with positive derivative $\dot{\gamma} > 0$ on $[0, 1]$.

Note that Assumption 1 is satisfied for the probability-generating function γ of any nonnegative random variable N satisfying $\mathbb{P}(N = 1) > 0$.

As we are especially interested in the fluorescence application, we define formally the pile-up model as follows.

Definition 1. Under Assumption 1, where γ denotes the probability generating function of some random variable N with values in \mathbb{N}^* , we denote by $\{g_\theta, \theta \in \Theta\}$ the corresponding collection of densities obtained by Relation (3). This model, that is dominated by \mathcal{L}_+ , is called the *pile-up model* associated with the *target model* $\{f_\theta, \theta \in \Theta\}$ and the distribution of N .

Let us verify Assumption 1 for the pile-up model in the fluorescence application. In general the number of photons per excitation cycle has Poisson distribution. It is clear that only those cycles provide information on the unknown parameter θ where a photon is detected. Hence, we suppose that N in (1) follows a Poisson distribution restricted on \mathbb{N}^* with parameter $\lambda > 0$, namely

$$\mathbb{P}(N = k) = \frac{\lambda^k e^{-\lambda}}{k!(1 - e^{-\lambda})}, \quad k \in \mathbb{N}^*. \quad (4)$$

It follows that the probability generating function in this case is given by

$$\gamma(u) = \mathbb{E}[u^N] = \frac{e^{\lambda u} - 1}{e^\lambda - 1}, \quad (5)$$

and thus Assumption 1 is verified. Obviously, γ depends on λ . A natural estimator of λ is based on the proportion of excitation cycles where no photon is detected, more precisely

$$\hat{\lambda} = \log \left(\frac{\text{number of excitation cycles}}{\text{number of cycles where no photon is detected}} \right).$$

Hence for simplicity we consider γ as a known function.

Widely used target models in the fluorescence application include the exponential distribution or finite mixtures of exponential distributions, possibly polluted by some additive instrument noise (see Ware *et al.*, 1973; O'Connor

& Phillips, 1984). Often the exponential parameters permit a physical interpretation.

In the remainder of the paper we adopt the following notation: under the expectation sign \mathbb{E}_θ , $(Y, Y_i, i \geq 1)$ are i.i.d. with density f_θ and $(Z, Z_i, i \geq 1)$ are i.i.d. with the corresponding distorted density g_θ defined by (3). Moreover we denote by F_θ and G_θ the associated distribution functions.

3 Estimation Method

To construct a parameter estimate, we propose a modification of the maximum likelihood approach that results in an estimator that is often computable by an EM-type algorithm.

3.1 Corrected Likelihood

Consider the log-likelihood associated with an i.i.d. sample (Y_1, \dots, Y_n) from the target distribution F_{θ_0} , namely

$$L_n(\theta) = \frac{1}{n} \sum_{i=1}^n \log f_\theta(Y_i) . \tag{6}$$

Recall that the rationale in using the log-likelihood as a contrast function is that, by the strong law of large numbers, as $n \rightarrow \infty$, $L_n(\theta) - L_n(\theta_0)$ converges to the negated Kullback-Leibler divergence $\mathbb{E}_{\theta_0} [\log(f_\theta(Y)/f_{\theta_0}(Y))]$ and thus is asymptotically maximized at $\theta = \theta_0$. Now, from Equation (3), we have for any real-valued integrable function h defined on \mathbb{R}_+ ,

$$\mathbb{E}_{\theta_0} [h(Y)] = \mathbb{E}_{\theta_0} \left[\frac{h(Z)}{\dot{\gamma}(\bar{F}_{\theta_0}(Z))} \right] . \tag{7}$$

By Assumption 1, γ is bijective. We denote by γ^{-1} its inverse on $[0, 1]$. Then Equation (2) gives $\bar{F}_{\theta_0} = \gamma^{-1} \circ \bar{G}_{\theta_0}$. Define

$$w(u) = \frac{1}{\dot{\gamma} \circ \gamma^{-1}(1 - u)}, \quad u \in [0, 1] , \tag{8}$$

with the convention that $1/\infty = 0$. Under Assumption 1, w is well defined since $\dot{\gamma} > 0$. Note that (7) can be rewritten as

$$\mathbb{E}_{\theta_0} [h(Y)] = \mathbb{E}_{\theta_0} [w \circ G_{\theta_0}(Z) h(Z)] . \tag{9}$$

Taking $h = \log f_\theta$, we get that

$$\tilde{L}_n(\theta) = \frac{1}{n} \sum_{i=1}^n w \circ G_{\theta_0}(Z_i) \log f_\theta(Z_i) \quad (10)$$

has the same property as the one pointed out before for the likelihood L_n . However, (10) involves θ_0 , so it cannot be used for parameter estimation. We propose to modify (10) by replacing G_{θ_0} with the empirical distribution function $\hat{G}_n(z) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{Z_i \leq z\}$. Then a corrected log-likelihood function which reasonably estimates $L_n(\theta)$ in (6), but is obtained from the sample (Z_1, \dots, Z_n) instead of the unobserved (Y_1, \dots, Y_n) is given by

$$\hat{L}_n(\theta) = \frac{1}{n} \sum_{i=1}^n w \circ \hat{G}_n(Z_i) \log f_\theta(Z_i) = \frac{1}{n} \sum_{i=1}^n w(i/n) \log f_\theta(Z_{(i,n)}) , \quad (11)$$

where $Z_{(i,n)}$ denotes the i -th order statistic of the sample (Z_1, \dots, Z_n) satisfying $Z_{(1,n)} \leq Z_{(2,n)} \leq \dots \leq Z_{(n,n)}$. We define a new parameter estimate of θ_0 by

$$\hat{\theta}_n = \arg \max_{\theta \in \Theta} \hat{L}_n(\theta) = \arg \max_{\theta \in \Theta} \sum_{i=1}^n w(i/n) \log f_\theta(Z_{(i,n)}) , \quad (12)$$

to which we will refer as the *corrected maximum likelihood estimator* (corrected MLE). The weights $w(i/n)$ appearing in $\hat{L}_n(\theta)$ are aimed to correct the distortion, namely, the fact that (Z_1, \dots, Z_n) is a sample of g_{θ_0} instead of f_{θ_0} .

Relation (9) can be used more widely for correcting any statistical method based on moments of the target distribution that one wishes to apply with distorted observations. More precisely, denote by \hat{P}_n^c the *corrected empirical distribution* from distorted observations Z_1, \dots, Z_n such that, for every function h ,

$$\hat{P}_n^c(h) = \frac{1}{n} \sum_i w(i/n) h(Z_{(i,n)}) ,$$

where w is given by (8). It is different from the standard empirical distributions

$$\hat{P}_n(h) = \frac{1}{n} \sum_i h(Y_i) \quad \text{and} \quad \tilde{P}_n(h) = \frac{1}{n} \sum_i h(Z_i) . \quad (13)$$

The following result, which justifies the above approach, relies on the fact that $\hat{P}_n^c(h)$, $\tilde{P}_n(w \circ G \times h)$ and $\hat{P}_n(h)$ have the same limits. It is obtained by applying Lemma 4 in Appendix A.

Theorem 1. *Suppose that Assumption 1 holds. Let Y have distribution F and $(Z_i)_{i \geq 1}$ be an i.i.d. sequence with distribution G defined in (2). Then, for any real valued function h defined on \mathbb{R}_+ satisfying $\mathbb{E}[|h(Z)|] < \infty$, we have $\hat{P}_n^c(h) \rightarrow \mathbb{E}[h(Y)]$ almost surely as $n \rightarrow \infty$.*

Theorem 1 implies that any consistent moment estimator based on a sample of the target distribution remains consistent when the corrected empirical distribution \hat{P}_n^c is used. It also gives that the corrected log-likelihood $\hat{L}_n(\theta)$ is such that $\hat{L}_n(\theta) - \hat{L}_n(\theta_0)$ converges to the Kullback-Leibler divergence of f_{θ_0} with f_θ almost surely. This is the basic argument for proving the consistency of the corrected MLE, see Section 4.1.

3.2 Maximization by an EM-type Algorithm

In general it is much easier to maximize the corrected likelihood $\hat{L}_n(\theta)$ in (11) than the likelihood function associated with the nonlinear transformation model, which has the form

$$\sum_{i=1}^n \log f_\theta(Z_i) + \sum_{i=1}^n \log \dot{\gamma}(1 - F_\theta(Z_i)). \quad (14)$$

Note that the main difference with $\hat{L}_n(\theta)$ is the appearance of a second term in (14) which highly complicates the maximization. In the special case of a pile-up model with an exponential target distribution, i.e. $f_\theta(y) = \theta e^{-\theta y}$, $\theta > 0$, and where N follows the restricted Poisson distribution given in (4), it turns out that the corrected MLE is given explicitly by

$$\hat{\theta}_n = \frac{\sum_{i=1}^n w(i/n)}{\sum_{i=1}^n w(i/n) Z_{(i,n)}}, \quad \text{with} \quad w(i/n) = \frac{1 - e^{-\lambda}}{\lambda \left[\frac{i}{n} (e^{-\lambda} - 1) + 1 \right]}. \quad (15)$$

This is in sharp contrast with the classical maximum likelihood estimator associated with the pile-up model, where the maximization of

$$\arg \max_{\theta} \sum_{i=1}^n \log \left(\sum_{k=1}^{\infty} p(k) k \theta e^{-k\theta Z_i} \right) \quad (16)$$

$$= \arg \max_{\theta} \left\{ n \log \theta - \theta \sum_{i=1}^n Z_i + \lambda \sum_{i=1}^n e^{-\theta Z_i} \right\} \quad (17)$$

has no explicit solution. Note that the log-argument in (16) is an infinite discrete mixture of exponential densities $\mathcal{E}(k\theta)$. The numerical cost for computing the maximizer of the corrected likelihood is in this case thus obviously

lower than for the pile-up likelihood. This feature remains true in a large variety of situations. Indeed, the only difference of the maximization problem given by (12) compared to maximizing the target likelihood (6) are the nonnegative weights $w(i/n)$. This makes the two maximization problems of equivalent cost, while the maximization of the likelihood in the nonlinear transformation model is in general much more costly, if tractable at all. Namely, the EM algorithm cannot be used to maximize the log-likelihood (14), since to our knowledge there is no structure of missing variables that gives rise to a sufficiently simple complete likelihood. Furthermore, we remark that the Gibbs sampler for the pile-up model proposed in Rebafka *et al.* (2008) which approximates the MLE has a long computation time.

A situation of particular interest is when the EM algorithm applies to the target model $\{f_\theta, \theta \in \Theta\}$, as this is the case for a finite mixture of exponential distributions polluted by additive noise with known distribution, which is a broadly used model in fluorescence. Then an EM-type algorithm resolves the maximization (12), which has the particularity to use the same missing data as in the target model, and not as the structure of the pile-up observations may suggest the number of emitted photons N and the component labels of all emitted photons. Indeed, this EM-algorithm treats pile-up observations as observations from the target model, while the pile-up effect is taken into account by appropriate weighting of the terms. To be more precise, we make the assumption that the target model is an incomplete data model.

Assumption 2. Suppose that Assumption 1 holds. Let μ be a measure on a state space \mathcal{X} . Suppose that $\{\pi_\theta, \theta \in \Theta\}$ is a collection of densities with respect to $\mathcal{L}_+ \otimes \mu$, such that $f_\theta = \int \pi_\theta(\cdot, s) \mu(ds)$ for all $\theta \in \Theta$.

The second component of the distribution π_θ corresponds to the missing or latent variable, say X . When f_θ is an exponential mixture, then the missing variable X is the label of the mixture component that generated an observation Y from the mixture f_θ . Let $q_{\theta, \theta'}(y, x)$ denote the conditional expectation of the log-density of (Y, X) at parameter θ given that $Y = y$ at parameter θ' , namely

$$q_{\theta, \theta'}(y) = \mathbb{E}_{\theta'}[\log \pi_\theta(Y, X) \mid Y = y] = \int_{x \in \mathcal{X}} \log \pi_\theta(y, x) \frac{\pi_{\theta'}(y, x)}{f_{\theta'}(y)} \mu(dx).$$

In analogy to the standard EM algorithm but using the approach with the

corrected likelihood, we define, for all $\theta, \theta' \in \Theta$,

$$Q(\theta, \theta'; Z_1, \dots, Z_n) = \sum_{i=1}^n w(i/n) q_{\theta, \theta'}(Z_{(i,n)}), \quad \text{and} \quad (18)$$

$$H(\theta, \theta'; Z_1, \dots, Z_n) = \sum_{i=1}^n w(i/n) \mathbb{E}_{\theta'} \left[\log \frac{\pi_{\theta}(Y, X)}{f_{\theta}(Y)} \mid Y = Z_{(i,n)} \right].$$

Note that the corrected likelihood verifies, for all $\theta, \theta' \in \Theta$,

$$\hat{L}_n(\theta) = Q(\theta, \theta'; Z_1, \dots, Z_n) - H(\theta, \theta'; Z_1, \dots, Z_n).$$

Now, as for the standard EM algorithm, we define the sequence $(\theta^{(t)})_{t \geq 0}$ for any starting value $\theta^{(0)} \in \Theta$ in the following recurrent way, for all $t \in \mathbb{N}$,

$$\theta^{(t+1)} = \arg \max_{\theta \in \Theta} Q(\theta, \theta^{(t)}; Z_1, \dots, Z_n). \quad (19)$$

The sequence $(\theta^{(t)})_{t \geq 0}$ has the same properties as a sequence obtained by the standard EM algorithm. Namely, at each iteration the value of the log-likelihood $\hat{L}_n(\theta^{(t)})$ increases and the sequence $(\theta^{(t)})_{t \geq 0}$ tends to a critical point of the corrected log-likelihood function $\hat{L}_n(\theta)$.

Theorem 2. *The sequence $(\theta^{(t)})_{t \geq 0}$ defined by (19) satisfies $\hat{L}_n(\theta^{(t+1)}) \geq \hat{L}_n(\theta^{(t)})$, for all $t \geq 0$. Moreover, if $(\theta^{(t)})_{t \geq 0}$ converges to some θ^* in the closure of Θ and if*

$$\nabla_{\theta} Q(\theta, \theta^{(t)}; Z_1, \dots, Z_n) |_{\theta = \theta^{(t+1)}} = 0, \quad \text{for all } t,$$

then θ^* is a critical point of $\hat{L}_n(\theta)$, i.e. $\nabla_{\theta} \hat{L}_n(\theta) |_{\theta = \theta^*} = 0$.

Theorem 1 and 4 in Dempster *et al.* (1977) correspond to the case where $w(i/n) \equiv 1$. Using that $w(i/n)$ are non negative numbers, the arguments provided in Dempster *et al.* (1977) continue to hold.

4 Asymptotic Behavior

In this section we study the asymptotic behavior of the corrected MLE $\hat{\theta}_n$ defined in (12). We show its consistency and determine its limit distribution. To this end we make use of the fact that the corrected likelihood function $\hat{L}_n(\theta)$ is an L -statistic, that is a linear combination of order statistics.

4.1 Consistency

Although the process \tilde{L}_n defined by (10) cannot be used in practice for estimation because it depends on the unknown parameter θ_0 , it is a standard empirical mean of the i.i.d. observations Z_1, \dots, Z_n and thus the standard theory for proving the consistency of M -estimators (see e.g. van der Vaart (1998)) can be applied. This theory says that, under standard conditions, the following assumption holds.

Assumption 3. Any random sequence $(t_n)_n$ valued in Θ satisfies $t_n \xrightarrow{P} \theta_0$, if

$$\tilde{L}_n(t_n) \geq \sup_t \tilde{L}_n(t) + o_P(1). \quad (20)$$

For instance, by van der Vaart (1998), Theorem 5.7, this is true, if both

- (3-i) the collection of functions $\mathcal{F} = \{w \circ G_\theta \log(f_\theta/f_{\theta_0}), \theta \in \Theta\}$ is a G_{θ_0} -Glivenko-Cantelli class;
- (3-ii) for all $\varepsilon > 0$, $\inf_{|\theta - \theta_0| > \varepsilon} \text{KL}(f_{\theta_0} \| f_\theta) > 0$, where $\text{KL}(f \| g) = \int \log \frac{f(x)}{g(x)} f(x) dx$ denotes the Kullback-Leibler divergence of two densities f and g .

The Kullback-Leibler divergence $\text{KL}(f_{\theta_0} \| f_\theta)$ appears here as a consequence of (9).

Theorem 3. Suppose that Assumptions 1 and 3 hold and that

$$\sup_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^n \left| \log \frac{f_\theta(Z_i)}{f_{\theta_0}(Z_i)} \right| = O_P(1). \quad (21)$$

Then any random sequence $(\hat{\theta}_n)$ is consistent for θ_0 , i.e. $\hat{\theta}_n \xrightarrow{P} \theta_0$, if it satisfies

$$\hat{L}_n(\hat{\theta}_n) \geq \sup_{\theta \in \Theta} \hat{L}_n(\theta) + o_P(1).$$

The following corollary provides a simple condition implying the general ones of Theorem 3.

Corollary 1. Suppose that Assumption 1 holds. Let Θ be a compact subset of \mathbb{R}^d . If there exists a function $K : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ such that $\mathbb{E}_{\theta_0}[K(Z)] < \infty$ and

$$|\log f_\theta(z) - \log f_{\theta'}(z)| \leq K(z) \|\theta - \theta'\|, \quad \theta, \theta' \in \Theta, z \in \mathbb{R}_+, \quad (22)$$

then the conclusion of Theorem 3 is true.

4.1.1 Example

In all specific cases that are relevant in the fluorescence application the corrected MLE is consistent. Consider a *multi-exponential* target density defined as

$$f_{\theta}(y) = \sum_{k=1}^K \alpha_k \nu_k e^{-\nu_k y}, \quad y \in \mathbb{R}_+, \quad (23)$$

and let γ be the probability generating function of the restricted Poisson distribution given in (5). For $K = 1$ we have the *single-exponential case*. We can use a parsimonious parametrization where α_K is given by $1 - \sum_{k=1}^{K-1} \alpha_k$. Let the parameter space Θ be defined as

$$\left\{ \theta = (\alpha_1, \dots, \alpha_{K-1}, \nu_1, \dots, \nu_K)^T : \nu_k \in (a, b), \alpha_k \in (\delta, 1), \sum_{k=1}^{K-1} \alpha_k < 1 - \delta \right\}$$

where $0 < a < b < \infty$ and $\delta > 0$. The identifiability of exponential mixtures is assured by the results of Teicher (1961). Now for all $\theta_1 < \theta_2 \in \Theta$

$$|\log f_{\theta_1}(z) - \log f_{\theta_2}(z)| = \frac{|\dot{f}_{\tilde{\theta}}(z)|}{f_{\tilde{\theta}}(z)} |\theta_2 - \theta_1|,$$

for some $\tilde{\theta} \in [\theta_1, \theta_2]$ and where $\dot{f}_{\tilde{\theta}}(z)$ denotes the gradient of $\theta \mapsto f_{\theta}(z)$ at $\tilde{\theta}$. Furthermore, since $\alpha_k > \delta$ and $\nu_k \in (a, b)$ for all k , we obtain

$$\begin{aligned} \left| \frac{\partial f_{\theta}(z)}{\partial \alpha_j}(z) \right| &= \frac{|\nu_j e^{-\nu_j z} - \nu_K e^{-\nu_K z}|}{f_{\theta}(z)} = \frac{|\nu_j e^{-\nu_j z} - \nu_K e^{-\nu_K z}|}{\sum_{k=1}^K \alpha_k \nu_k e^{-\nu_k z}} \leq \frac{1}{\alpha_j} + \frac{1}{\alpha_K} \leq \frac{2}{\delta}, \\ \left| \frac{\partial f_{\theta}(z)}{\partial \nu_j}(z) \right| &= \frac{|\alpha_j (1 - \nu_j z) e^{-\nu_j z}|}{f_{\theta}(z)} \leq \frac{\alpha_j |1 - \nu_j z| e^{-\nu_j z}}{\sum_{k=1}^K \alpha_k \nu_k e^{-\nu_k z}} \leq \frac{|1 - \nu_j z|}{\nu_j} \leq \frac{1 + bz}{a}. \end{aligned}$$

Hence, (22) holds with $K(z) = \frac{2}{\delta} \wedge \frac{1+bz}{a}$ and the corrected MLE is consistent.

Similar arguments apply to the case of a *multi-exponential target distribution polluted by additive noise*, where we denote by η the probability density of the instrument noise. Consider the convolution of η and the multi-exponential density given by (23). The target density is then given by

$$f_{\theta}(z) = \eta \star \sum_{k=1}^K \alpha_k \nu_k e^{-\nu_k \cdot}(z) = \sum_{k=1}^K \alpha_k \nu_k e^{-\nu_k z} \int_0^z \eta(u) e^{\nu_k u} du. \quad (24)$$

Corollary 1 can be used to derive the consistency of the corrected MLE in this case.

4.2 Limit Distribution

We show that a theorem, similar to the central limit theorem for M -estimators of Pollard (1985), holds for M -estimators maximizing an L -statistic. The theorem applies to the corrected MLE $\hat{\theta}_n$.

For an i.i.d. sample (Z_1, \dots, Z_n) from the distorted distribution G defined by (2) and weight function w defined by (8), we consider a contrast process defined as the L -statistic

$$M_n(t) = \hat{P}_n^c h(\cdot, t) = \frac{1}{n} \sum_{i=1}^n w(i/n) h(Z_{(i,n)}, t),$$

with $h(z, t)$ defined on $\mathbb{R}_+ \times \Theta$. If $h(\cdot, t) = \log f_t$ (or, equivalently $h(\cdot, t) = \log(f_t/f_{\theta_0})$), then $M_n(t)$ is the corrected likelihood (11). Denote

$$M(t) = \mathbb{E}_{\theta_0}[h(Y, t)],$$

where θ_0 is the true parameter. Furthermore, the *corrected* empirical process corresponding to \hat{P}_n^c is defined, for a function k on \mathbb{R}_+ , as

$$\nu_n^c k = \sqrt{n} \left(\hat{P}_n^c(k) - \mathbb{E}_{\theta_0}[k(Y)] \right).$$

We have $\nu_n^c h(\cdot, t) = \sqrt{n}(M_n(t) - M(t))$. Adapting Pollard (1985), the following result holds.

Theorem 4. *Suppose that the following assertions are true.*

- (i) θ_0 is an interior point of $\Theta \subset \mathbb{R}^d$;
- (ii) $t \mapsto M(t)$ has a nonsingular second derivative $-J \equiv \nabla_t^2 M(\theta_0)$ at its maximizing value θ_0 ;
- (iii) Let θ_n be a maximizer of M_n for every $n \geq 1$ and suppose that $(\theta_n)_n$ converges in probability to θ_0 ;

Furthermore, suppose that the following expansion holds for all $z \in \mathbb{R}_+$ and $t \in \Theta$,

$$h(z, t) = h(z, \theta_0) + (t - \theta_0)^T \Delta(z) + |t - \theta_0| r(z, t), \quad (25)$$

with functions $r : \mathbb{R}_+ \times \Theta \rightarrow \mathbb{R}$ and $\Delta : \mathbb{R}_+ \rightarrow \mathbb{R}^d$ satisfying

- (iv) there exists a covariance matrix Σ such that $\nu_n^c \Delta \xrightarrow{d} \mathcal{N}(0, \Sigma)$;
- (v) for any sequence of balls U_n that shrinks to θ_0 as $n \rightarrow \infty$, we have

$$\sup_{t \in U_n} \frac{|\nu_n^c r(\cdot, t)|}{1 + \sqrt{n}|t - \theta_0|} \xrightarrow{P} 0.$$

Then, it follows that

$$\sqrt{n}(\theta_n - \theta_0) \xrightarrow{d} \mathcal{N}(0, J^{-1}\Sigma J^{-1}) .$$

The proof of this result is omitted as it is easily obtained by adapting the proof of the theorem in Pollard (1985) in the following way.

1. The standard empirical process ν_n defined by $\nu_n k = \sqrt{n}(\sum k(Y_i) - \mathbb{E}_{\theta_0}[k(Y)])$ is replaced by the corrected empirical process ν_n^c defined above.
2. The conditions $\mathbb{E}_{\theta_0}[|\Delta(Y)|^2] < \infty$ and $\mathbb{E}_{\theta_0}[\Delta(Y)] = 0$ are replaced by (iv).

These adaptations yield the two unusual conditions (iv) and (v). Condition (v) is similar to the so called *stochastic differentiability* condition in Pollard (1985) except that the empirical process ν_n is replaced by ν_n^c which is based on an L -statistic. However the following lemma shows that one can use a standard empirical process to verify this condition. Recall that \tilde{P}_n is defined in (13). Then denote by $\tilde{\nu}_n$ the standard empirical process associated with the i.i.d. sequence (Z_i) ,

$$\tilde{\nu}_n k = \sqrt{n} \left(\tilde{P}_n(k) - \mathbb{E}_{\theta_0}[k(Z)] \right) .$$

Lemma 1. *Suppose that w defined in (8) is Lipschitz on $[0, 1]$, that is, there exists $L > 0$ such that $|w(u) - w(v)| \leq L |u - v|$, for all $u, v \in [0, 1]$. Then Condition (v) holds, if for any sequence of balls U_n that shrinks to θ_0 as $n \rightarrow \infty$, we have*

$$\sup_{t \in U_n} \tilde{P}_n |r(\cdot, t)| \xrightarrow{P} 0 \quad \text{and} \quad \sup_{t \in U_n} \frac{|\tilde{\nu}_n(w \circ G \times r(\cdot, t))|}{1 + \sqrt{n}|t - \theta_0|} \xrightarrow{P} 0 . \quad (26)$$

In most cases, a smoothness condition on $t \mapsto h(z, t)$ holds with some uniformity in z , which implies Condition (v), as described by the following lemma.

Lemma 2. *Suppose that w is Lipschitz on $[0, 1]$ and that for all $z \in \mathbb{R}_+$, $h(z, \cdot)$ is continuously differentiable in Θ with gradient denoted by $\dot{h}(z, \cdot)$. Assume moreover that there exists K defined on \mathbb{R}_+ and a neighborhood U of θ_0 such that*

$$\left| \dot{h}(z, t) - \dot{h}(z, \theta_0) \right| \leq K(z) |t - \theta_0|, \text{ for all } t \in U, \text{ and } \mathbb{E} [K^2(Z)] < \infty. \quad (27)$$

Then Condition (v) holds for r given by (25) with $\Delta(z) = \dot{h}(z, \theta_0)$.

To verify Condition (iv) the following result for L -statistics can be used. It is an immediate application of Theorem 5 and Proposition 1 in Appendix B to the nonlinear transformation model.

Lemma 3. *Suppose that Assumption 1 holds. Let k be a continuous function of bounded variation on bounded intervals and assume that the weight function w is continuously differentiable. If furthermore $\mathbb{E}_{\theta_0}[|k(Z)|] < \infty$ and $\mathbb{E}_{\theta_0}[(w \circ G(Z)k(Z))^2] < \infty$, then*

$$\nu_n^c k \xrightarrow{d} \mathcal{N}(0, \sigma^2(k)) ,$$

with limit variance $\sigma^2(k)$ given by

$$\begin{aligned} \sigma^2(k) &= \mathbb{E}_{\theta_0}[(w \circ G_{\theta_0}(Z))^2 k^2(Z)] \\ &\quad + 2\mathbb{E}_{\theta_0}[\dot{w}_1 \circ G_{\theta_0}(Z_1)\dot{w}_2 \circ G_{\theta_0}(Z_2)k(Z_1)k(Z_2)\mathbf{1}\{Z_1 > Z_2\}] , \end{aligned}$$

with $w_1(x) = (1-x)w(x)$ and $w_2(x) = xw(x)$ and \dot{w}_1 and \dot{w}_2 are the derivatives of w_1 and w_2 .

Using the Cramér-Wold device and the linearity of $\nu_n^c k$, a multidimensional version of the central limit theorem holds. Namely, if the above conditions hold for every function k_1, \dots, k_m and denoting $\mathbf{k} = [k_1, \dots, k_m]^T$, then $\nu_n^c(\mathbf{k}) \xrightarrow{d} \mathcal{N}(0, \Sigma(\mathbf{k}))$ where $\Sigma(\mathbf{k}) = (\sigma^2(k_i, k_j))_{i,j}$ and

$$\begin{aligned} \sigma^2(k_i, k_j) &= \mathbb{E}_{\theta_0}[(w \circ G_{\theta_0}(Z))^2 k_i(Z)k_j(Z)] \\ &\quad + 2\mathbb{E}_{\theta_0}[\dot{w}_1 \circ G_{\theta_0}(Z_1)\dot{w}_2' \circ G_{\theta_0}(Z_2)k_i(Z_1)k_j(Z_2)\mathbf{1}\{Z_1 > Z_2\}] . \end{aligned}$$

4.2.1 Example

Theorem 4 applies to the corrected MLE defined in (12). In this case $M(\theta) = \mathbb{E}_{\theta_0}[\log f_{\theta}(Y)]$. Hence J in Condition (ii) is the Fisher information matrix of the target model, but Σ is not. This is why the asymptotic variance $J^{-1}\Sigma J^{-1}$ does not equal to the Fisher information, see further discussion in Subsection 6.2.

In the *exponential case*, when $\Theta = (a, b)$ with $0 < a < b < \infty$, the conditions of Theorem 4 are easily verified using Lemma 2 and 3. As $M(\theta) = \log \theta - \frac{\theta}{\theta_0}$, we have $J = \theta_0^{-2}$. According to Theorem 4 and some straightforward computation, $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \mathcal{N}(0, \theta_0^2 \tau)$, where

$$\tau = \frac{2}{\lambda^2} \sum_{n=1}^{\infty} \frac{(-\lambda)^n}{n!n^2} - e^{\lambda} \sum_{n=1}^{\infty} \frac{(-\lambda)^n}{n!n} \sum_{m=0}^{\infty} \frac{(-\lambda)^m}{m!(m+n)} . \quad (28)$$

See Subsection 6.2 for a visualization of the limit variance and a comparison with the Cramér-Rao bound computed in Rebafka *et al.* (2008).

In the multi-exponential case of Subsection 4.1.1, the same conditions hold, yielding $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \mathcal{N}(0, J^{-1}\Sigma J^{-1})$, where J is the Fisher information matrix of the multi-exponential distribution and

$$\begin{aligned} \Sigma &= \frac{1 - e^{-\lambda}}{\lambda} \int_0^\infty e^{\lambda F_{\theta_0}(y)} \frac{\dot{f}_{\theta_0}^T(y) \dot{f}_{\theta_0}(y)}{f_{\theta_0}(y)} dy \\ &\quad - 2e^{-\lambda} \int_0^\infty \int_{y_2}^\infty e^{\lambda F_{\theta_0}(y_1) + \lambda F_{\theta_0}(y_2)} \dot{f}_{\theta_0}^T(y_1) \dot{f}_{\theta_0}(y_2) dy_1 dy_2 . \end{aligned}$$

4.2.2 Confidence Intervals

Based on the corrected MLE $\hat{\theta}_n$ defined by (12), confidence intervals are easily constructed. For simplicity consider the one-dimensional case, i.e. $\theta \in \mathbb{R}$. Theorem 4 suggests the following asymptotic confidence interval of θ with confidence level $1 - \alpha$

$$\text{IC}_n = \left[\hat{\theta}_n + q_{\alpha/2} \sqrt{\hat{V}_n/n}, \hat{\theta}_n - q_{\alpha/2} \sqrt{\hat{V}_n/n} \right] ,$$

where $q_{\alpha/2}$ is the $\alpha/2$ -quantile of the standard gaussian distribution $\mathcal{N}(0, 1)$ and \hat{V}_n is an estimator of the limit variance $V = J^{-1}\Sigma J^{-1}$ of Theorem 4. Note that here J is the Fisher information of the target model $\{f_\theta, \theta \in \Theta\}$ evaluated at θ_0 , that is

$$J = -\mathbb{E} \left[\frac{\partial^2 \log f_{\theta_0}(Y)}{\partial^2 \theta_0} \right] .$$

A natural estimator of J is hence given by

$$\hat{J}_n = -\frac{1}{n} \sum_{i=1}^n w(i/n) \frac{\partial^2 \log f_{\hat{\theta}_n}(Z_{(i,n)})}{\partial^2 \hat{\theta}_n} .$$

Then, with the estimator $\hat{\sigma}_n^2$ of Σ given in Appendix B by (42) with X replaced by Z , we obtain an estimator of the limit variance V by

$$\hat{V}_n = \hat{\sigma}_n^2 / (\hat{J}_n)^2 . \tag{29}$$

5 Application to Fluorescence Measurements

When using the TCSPC fluorescence technique, the arrival time Y_i of a photon on the detector is the fluorescence lifetime, say \tilde{Y}_i , plus some noise E_i , which is due to the measuring instrument. That is, $Y_i = \tilde{Y}_i + E_i$ for independent variables \tilde{Y}_i and E_i . The noise, which is mainly due to the lamp, is called the *instrumental response function*. It is individual for every measuring instrument and does not belong to a common parametric family. However, as the probability density function η of the instrumental response function can easily be measured, η is assumed known in the following. The target density f_θ is given by (24).

For an observation $Y = \tilde{Y} + E$ let (S, E) be the latent variables, where S is the label of the exponential component that has generated \tilde{Y} , and E is the noise with density η . The joint density of (Y, S, E) is given by

$$\pi_\theta(y, s, x) = \eta(x)\alpha_s\nu_s e^{-\nu_s(y-x)}, \quad 0 < x < y, \quad s = 1, \dots, K.$$

The maximization (19) of $Q(\theta, \theta'; Z_1, \dots, Z_n)$ defined by (18) has the solutions

$$\alpha_l^{(t+1)} = \frac{\sum_{i=1}^n w(i/n) p(l | Z_{(i,n)}, \theta^{(t)})}{\sum_{i=1}^n w(i/n)},$$

$$\nu_l^{(t+1)} = \frac{\sum_{i=1}^n w(i/n) p(l | Z_{(i,n)}, \theta^{(t)})}{\sum_{i=1}^n w(i/n) p(l | Z_{(i,n)}, \theta^{(t)}) \{Z_{(i,n)} - \mathbb{E}_{\theta^{(t)}}[E | Y = Z_{(i,n)}, S = l]\}}$$

where $p(l | z, \theta) = \mathbb{P}_{\theta^{(t)}}(S = l | Y = Z)$. It remains to evaluate

$$p(l | z, \theta) = \frac{\alpha_l \nu_l e^{-\nu_l z} \int_0^z \eta(x) e^{\nu_l x} dx}{\sum_{k=1}^K \alpha_k \nu_k e^{-\nu_k z} \int_0^z \eta(x) e^{\nu_k x} dx}, \quad (30)$$

$$\mathbb{E}_\theta[E | Y = z, S = l] = \frac{\int_0^z x \eta(x) e^{\nu_l x} dx}{\int_0^z \eta(x) e^{\nu_l x} dx}. \quad (31)$$

In the TCSPC-set up the instrumental response function is approximated by a histogram, so that the integrals in (30) and (31) are Riemann sums.

We apply the corrected MLE to real TCSPC measurements. Figure 1 shows the histogram of photon arrival times and the instrument response function η of the measuring instrument. Data were obtained at a laser intensity corresponding to $\lambda = 0.166$. Hence, about 8% of the arrival times are the minimum of two or more photons. Consequently, the pile-up effect is not negligible. The sample size is $n = 1,743,811$ and there is a single-exponential

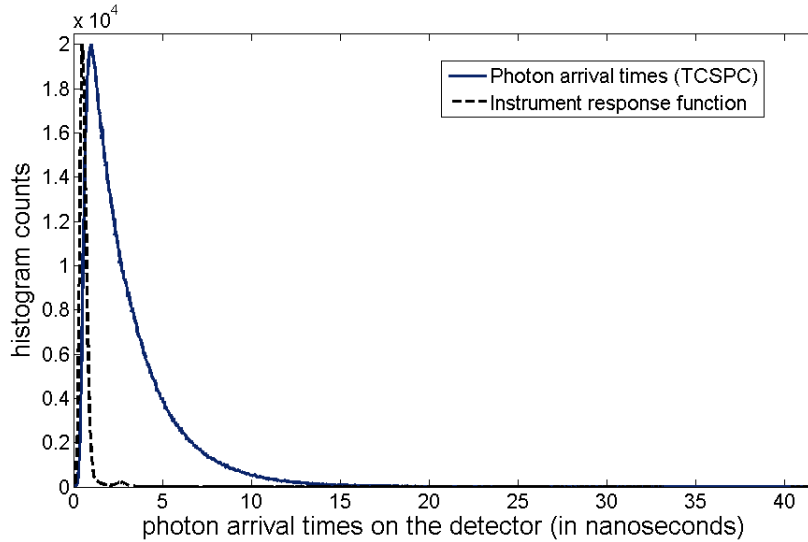


Figure 1: Histogram of TCSPC measurements and instrumental response function η of a single exponential component with Poisson parameter $\lambda = 0.166$.

component, $K = 1$. In this experiment the lifetime constant of the molecule is known to be $\tau = 1/\nu = 2.54$ ns. For more details on the data we refer to Patting *et al.* (2007).

A traditional estimator of the exponential parameter that does not take into account the pile-up effect yields the value $\tilde{\tau} = 2.40$ ns which is significantly shorter than the expected value. Applying the corrected MLE of the preceding paragraph provides the estimated value $\hat{\tau} = 1/\hat{\nu} = 2.5393$ ns. Moreover, using an estimator of the Fisher information and the variance estimator given in (29), the following confidence intervals with confidence level 0.95 are obtained: $IC_{\text{traditional}} = [2.3997, 2.4003]$ (interval length $6.75 \cdot 10^{-4}$) for the traditional estimator and $IC_{\text{pile-up}} = [2.5397, 2.5403]$ (interval length $6.89 \cdot 10^{-4}$) for the corrected MLE. As both confidence intervals are very short (due to the large number of observations), it is clear that the corrected MLE is unbiased, while the traditional estimator is not. It is hence indispensable to take the pile-up effect into account in the estimation procedure. We draw the conclusion that the corrected MLE is well suited for the pile-up model and handles additive noise correctly.

6 Numerical Study

The numerical performance of the corrected MLE in the pile-up model is evaluated in a twofold study. First we show that the acquisition time can be reduced by increasing the Poisson parameter λ . Second we compare the variance of the estimator to the Cramér-Rao bound of the pile-up model. Several single and multi-exponential target models are considered, but, for simplicity, no additive noise.

6.1 Reduction of Acquisition Time

In the fluorescence set-up the acquisition time consists of all excitation cycles, including those where no photon is detected. Hence, in the following N is assumed to have classical Poisson distribution on \mathbb{N} with parameter λ (and not restricted on \mathbb{N}^* as in (4)). If $N = 0$ put the pile-up observation $Z = \infty$. Then Z admits a density g_θ with respect to $\bar{\mathcal{L}}_+$, defined as the measure on $\mathbb{R}_+ \cup \{+\infty\}$ which puts mass 1 at $+\infty$ and whose restriction on \mathbb{R}_+ is the Lebesgue measure \mathcal{L}_+ . As shown in Rebafka *et al.* (2008), Lemma 1 the pile-up density is given by

$$g_\theta(z) = \begin{cases} \lambda f_\theta(z) e^{-\lambda F_\theta(z)}, & \text{if } x \in \mathbb{R}_+ \\ e^{-\lambda}, & \text{if } x = \infty. \end{cases} \quad (32)$$

Obviously, only observations where a photon is detected ($Z < \infty$) contain information on θ_0 . The average number λ of fluorescence photons per light pulse depends on the laser intensity which is tuned by the user. It increases, when λ does.

The current practice in fluorescence is to avoid pile-up by collecting data at a very low intensity λ , such that it is unlikely to have more than one photon per laser pulse, e.g. $\mathbb{P}(N > 1) = 0.0012$ if $\lambda = 0.05$. Observed photon arrival times may then be considered as realizations of the target distribution f_{θ_0} and not of the pile-up distribution g_{θ_0} . Thus, if f_θ is the multi-exponential density, standard estimators for exponential mixtures apply as the classical EM algorithm, which we consider in the following simulations.

Synthetic data are drawn from the pile-up density g_{θ_0} in (32). For the classical EM data are generated at $\lambda = 0.05$, while we use higher intensities for the corrected MLE ($\lambda = 1.32, 1.5, 2$). Let m denote the sample size, or better, the total number of excitation cycles. We stress that the observations where no photon event occurs ($Z = \infty$) are discarded from the samples, such that the effective number of observations used in the algorithms is much smaller than m . For $\lambda = 0.05$ we have $\mathbb{P}(N = 0) = 0.951$ compared to only

$K = 1, \nu_1 = 2$							
sample size m		100		500		1,000	
classical EM	ν_1	0.8095	(3.1687)	0.1321	(0.4728)	0.0769	(0.3096)
$\lambda = 0.05$							
corr. MLE	ν_1	0.0511	(0.3061)	0.0083	(0.1238)	0.0033	(0.0884)
$\lambda = 1.32$							

$K = 2, \alpha_1 = 0.25, \nu_1 = 0.2, \nu_2 = 2$							
sample size m		500		1,000		5,000	
classical EM	α_1	0.0611	(0.1702)	0.0171	(0.1122)	0.0037	(0.0441)
$\lambda = 0.05$	ν_1	0.1023	(0.2742)	0.0295	(0.1233)	0.0039	(0.0352)
	ν_2	0.9297	(6.3295)	0.2686	(1.0485)	0.0558	(0.2293)
corrected	α_1	0.0040	(0.0489)	0.0020	(0.0335)	0.0003	(0.0151)
MLE	ν_1	0.0065	(0.0429)	0.0024	(0.0280)	0.0001	(0.0122)
$\lambda = 2$	ν_2	0.0226	(0.2061)	0.0107	(0.1394)	0.0018	(0.0641)

$K = 3, \alpha_1 = 0.4, \alpha_2 = 0.3, \nu_1 = 0.1, \nu_2 = 0.5, \nu_3 = 2$							
sample size m		1,000		5,000		10,000	
classical EM	α_1	0.0218	(0.1589)	0.0171	(0.1110)	0.0210	(0.0883)
$\lambda = 0.05$	α_2	0.0850	(0.1381)	0.0803	(0.1031)	0.0593	(0.0965)
	α_3	0.1068	(0.1060)	0.0632	(0.1245)	0.0383	(0.1171)
	ν_1	0.0095	(0.0398)	0.0011	(0.0204)	0.0014	(0.0150)
	ν_2	0.5903	(0.7436)	0.1939	(0.3714)	0.0926	(0.2717)
	ν_3	8.7287	(91.414)	2.6270	(13.257)	1.5467	(7.9115)
corrected	α_1	0.0173	(0.1019)	0.0086	(0.0462)	0.0037	(0.0286)
MLE	α_2	0.0542	(0.0927)	0.0168	(0.0485)	0.0093	(0.0338)
$\lambda = 1.5$	α_3	0.0369	(0.1109)	0.0083	(0.0617)	0.0056	(0.0429)
	ν_1	0.0018	(0.0179)	0.0011	(0.0074)	0.0006	(0.0049)
	ν_2	0.0872	(0.2851)	0.0096	(0.1314)	0.0069	(0.0909)
	ν_3	0.7603	(4.0176)	0.0839	(0.3452)	0.0429	(0.2178)

Table 1: Empirical bias and standard deviation (in parentheses) of classical EM estimator and corrected MLE for different choices of parameters.

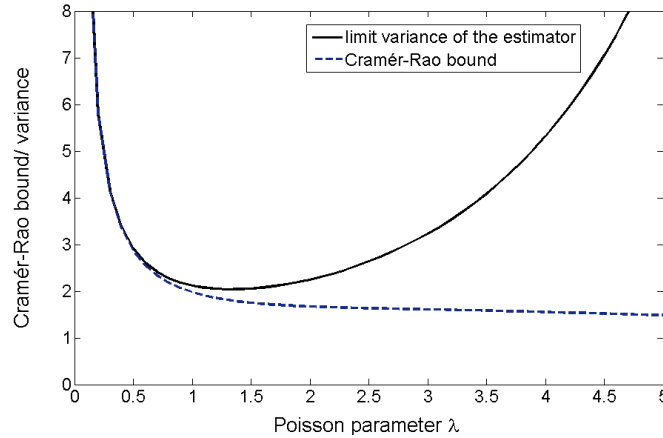


Figure 2: Limit variance of the corrected MLE versus Cramér-Rao bound in the exponential case with $\nu_1 = 1$.

$\mathbb{P}(N = 0) = 0.368$ when $\lambda = 1$. The empirical bias and standard deviations of both estimators are evaluated for different sample sizes m , each based on 1,000 repetitions, see Table 1.

The results in Table 1 show that the estimation quality, measured in terms of bias and standard deviation, obtained by the classical EM is achieved with only 10% of the observations with the corrected MLE. Thus a significant reduction of 90% of the acquisition time can be obtained by using data collected at a higher intensity and by applying the new corrected MLE.

6.2 Comparison to Cramér-Rao Bound

A comparison of the variance of the corrected MLE to the Cramér-Rao bound provides an explanation of the significant reduction of the acquisition time in the TCSPC application. In Rebafka *et al.* (2008) the Cramér-Rao bound, which is a lower bound of the variance of non biased estimators of θ_0 , is studied for the pile-up model defined in (32), where the no-photon events are taken into consideration. In the single-exponential case with known λ , the Cramér-Rao bound of the exponential parameter θ_0 decreases when λ increases, see Figure 2. Hence, for small λ , any estimator has large variance and many data are necessary for reliable estimation of θ_0 . In the exponential case the limit variance of the corrected MLE is given by

$$s^2 = \frac{\theta_0^2 \tau}{\mathbb{P}(N > 0)} = \frac{\theta_0^2 \tau}{1 - e^{-\lambda}},$$

m	corrected MLE				classical MLE			
	bias	std	length	cov.	bias	std	length	cov.
100	0.0511	0.3061	1.0641	0.927	0.0335	0.2800	0.9328	0.9140
500	0.0083	0.1238	0.4812	0.947	0.0051	0.1206	0.4109	0.9120
1,000	0.0033	0.0884	0.3417	0.944	0.0027	0.0828	0.2901	0.9210

Table 2: Comparison of the corrected MLE and the classical MLE associated with the pile-up model for synthetic data in the case of a single-exponential target distribution with exponential parameter $\nu_1 = 2$ and Poisson parameter $\lambda = 1.32$. Comparison of estimates of the bias, standard deviation, length and coverage of confidence interval with confidence level 0.95 for different sample sizes m .

where τ is defined in (28). Figure 2 shows that the limit variance s^2 attains the Cramér-Rao bound when $\lambda \leq 0.7$. The minimum is attained at $\lambda_{\min} \approx 1.329$, where the variance is ten times smaller than at $\lambda = 0.05$. This difference coincides with the reduction of the acquisition time of a factor ten by using the corrected MLE observed in Subsection 6.1. The point of minimum λ_{\min} is independent of θ_0 .

From the comparison with the Cramér-Rao bound, we see that the corrected MLE is not efficient at λ_{\min} . Table 2 provides more details on the loss by comparing the corrected MLE to the classical MLE associated with the pile-up model. Note that in the special one-dimensional case of a single-exponential target distribution the classical MLE defined in (17) can be computed numerically. The classical MLE is evaluated on the same synthetic datasets as in Table 1 and bias, standard deviation and length and coverage of confidence intervals are estimated. We observe that in terms of bias and standard deviation the classical MLE performs slightly better than the corrected MLE, while the confidence intervals based on the corrected MLE have a better coverage. The latter difference may be related to the different estimates of the variance, where the one of the corrected MLE seems to be more appropriate to finite samples.

In the multi-exponential case the Cramér-Rao bound is not explicitly known, but it can be estimated by the inverse of an approximation of the Fisher information matrix obtained by Monte-Carlo. For the two-component model used in Table 1, that is $\theta_0 = (\alpha_1, \nu_1, \nu_2) = (0.25, 0.2, 2)$ and $\lambda = 2$, the Cramér-Rao bound CRB and the covariance matrix Cov, both obtained by

Monte-Carlo, are approximately

$$\text{CRB} = \begin{pmatrix} 1.151 & 0.483 & 2.943 \\ 0.483 & 0.832 & 1.388 \\ 2.943 & 1.388 & 17.494 \end{pmatrix} \quad \text{and} \quad \text{Cov} = \begin{pmatrix} 1.310 & 0.576 & 3.453 \\ 0.576 & 0.944 & 1.794 \\ 3.453 & 1.794 & 20.517 \end{pmatrix}$$

From the relative closeness of the two matrices we conclude that the corrected MLE almost attains the Cramér-Rao bound and is nearly optimal in this sense. For completeness the covariance matrix can be compared to the limit variance obtained by Theorem 4 which is

$$\begin{pmatrix} 1.193 & 0.471 & 2.794 \\ 0.471 & 0.804 & 1.535 \\ 2.794 & 1.535 & 19.113 \end{pmatrix}.$$

We close the section with a heuristic explanation of the bad performance of the estimator for large λ . The corrected MLE is based on weighting the ordered observations. When a rather long arrival time is observed, it is very likely that there is no other photon and hence that we observe the ‘true’ distribution f_{θ_0} . That is why we associate higher weights with large observations. From (15) we note that the weights for large observations grow exponentially with λ while the weights for short observations decrease even more. It follows that if λ is large enough, the estimator relies almost entirely on the largest observations. It is as if the sample size diminishes. This movement is contrary to the augmentation of the proportion of events where a photon is detected when λ increases. Obviously, for small λ the latter augmentation is dominant, but for large λ the effect of large weights becomes predominant and increases the variance of the estimator.

7 Discussion

In this paper we proposed a new estimator for the pile-up model and, more generally, for nonlinear transformation models of the form $\tilde{G} = \gamma \circ \bar{F}$. The specific working assumption for the corrected MLE is that the function γ is known or can be estimated beforehand. This may often be the case in so-called *cure models* encountered in many clinical settings (see e.g. Tsodikov *et al.* (2003)) where G is an improper distribution function (i.e. $\lim_{t \rightarrow \infty} G(t) < 1$). More precisely, G denotes the distribution of the recurrence duration of some disease (or more generally the duration until the experience of the event of interest). However, there may be cured patients for whom no recurrence

duration is observed. In this case one may set the observation to some default value as $+\infty$. Obviously, cured patients do not provide any information on the the target distribution F . However, the proportion of cured patients may be used to determine the function γ , if we suppose a parametric setting where $\gamma \in \{\gamma_\alpha, \alpha \in A \subset \mathbb{R}\}$. Indeed, this is the case in the TCSPC application, where the number of photons follows a Poisson distribution with values in $\{0, 1, \dots\}$, and when $N = 0$, then no photon arrival time is measured. In practice, the proportion of excitations where no photon arrival is detected is used to estimate the Poisson parameter which determines the function γ . Then these observations are discarded from the sample to compute the corrected MLE. In the same way the corrected MLE may be used in other settings like cure models.

A Technical Arguments

A.1 Preliminary Results

Lemma 4. *Under Assumption 1, the weight function w defined in (8) is uniformly continuous on $[0, 1]$ taking its values in $[1/\max_{u \in [0,1]} \dot{\gamma}(u), 1/\min_{u \in [0,1]} \dot{\gamma}(u)]$. In addition, as $n \rightarrow \infty$,*

$$\sup_{u \geq 0} \left| w \circ \hat{G}_n(u) - w \circ G_{\theta_0}(u) \right| \rightarrow 0, \quad a.s. \quad (33)$$

Proof. The first assertions are clear. The Glivenko-Cantelli theorem gives that $\sup_{u > 0} \left| \hat{G}_n(u) - G_{\theta_0}(u) \right| \rightarrow 0$ a.s. and the uniform continuity of w yields (33). \square

A.2 Proof of Theorem 3

It suffices to show that $\hat{\theta}_n$ verifies (20). Denote for all t

$$\tilde{L}_n(t, \theta_0) = \tilde{L}_n(t) - \tilde{L}_n(\theta_0) \quad \text{and} \quad \hat{L}_n(t, \theta_0) = \hat{L}_n(t) - \hat{L}_n(\theta_0).$$

Lemma 4 and Condition (21) imply that

$$\begin{aligned}
 \left| \tilde{L}_n(\hat{\theta}_n, \theta_0) - \hat{L}_n(\hat{\theta}_n, \theta_0) \right| &= \left| \frac{1}{n} \sum_{i=1}^n \left\{ w \circ \hat{G}_n(Z_i) - w \circ G_{\theta_0}(Z_i) \right\} \log \frac{f_{\hat{\theta}_n}(Z_i)}{f_{\theta_0}(Z_i)} \right| \\
 &\leq \sup_z |w \circ \hat{G}_n(z) - w \circ G_{\theta_0}(z)| \sup_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^n \left| \log \frac{f_{\theta}(Z_i)}{f_{\theta_0}(Z_i)} \right| \\
 &= o_p(1) .
 \end{aligned} \tag{34}$$

It follows together with the fact that $\hat{\theta}_n$ is a maximizer of $\theta \mapsto \hat{L}_n(\theta, \theta_0)$ that

$$\begin{aligned}
 \tilde{L}_n(\hat{\theta}_n, \theta_0) &= \tilde{L}_n(\hat{\theta}_n, \theta_0) - \hat{L}_n(\hat{\theta}_n, \theta_0) + \hat{L}_n(\hat{\theta}_n, \theta_0) \\
 &\geq \sup_t \hat{L}_n(t, \theta_0) + o_p(1) \\
 &\geq \sup_t \tilde{L}_n(t, \theta_0) + o_p(1) ,
 \end{aligned} \tag{35}$$

where the last inequality is obtained by (34). Finally (35) is equivalent to (20).

A.3 Proof of Corollary 1

Using that $\min_u \dot{\gamma}(u) > 0$, we have $\sup_u w(u) \leq 1/\min_u \dot{\gamma}(u) < \infty$ and Condition (22) gives that, for any $\theta, \theta' \in \Theta$ and $z \in \mathbb{R}_+$,

$$\left| w \circ G_{\theta_0}(z) \log \frac{f_{\theta}(z)}{f_{\theta_0}(z)} - w \circ G_{\theta_0}(z) \log \frac{f_{\theta'}(z)}{f_{\theta_0}(z)} \right| \leq \frac{1}{\min_u \dot{\gamma}(u)} K(z) \|\theta - \theta'\| .$$

This and the compactness of Θ imply (3-i). By the identifiability condition of Assumption 1, we know that $\theta \mapsto \text{KL}(f_{\theta_0} \| f_{\theta})$ has a unique minimum at $\theta = \theta_0$. Moreover, observing that

$$\text{KL}(f_{\theta_0} \| f_{\theta}) - \text{KL}(f_{\theta_0} \| f_{\theta'}) = \mathbb{E}_{\theta_0} [\log f_{\theta'}(Y) - \log f_{\theta}(Y)] ,$$

we see that (22) and $\mathbb{E}_{\theta_0}[K(Y)] \leq \mathbb{E}_{\theta_0}[K(Z)]/\min_u \dot{\gamma}(u) < \infty$ imply the continuity of $\theta \mapsto \text{KL}(f_{\theta_0} \| f_{\theta})$ over the compact Θ , so that (3-ii) holds. Thus Assumption 3 holds. To apply Theorem 3, we now need to verify Condition (21), which follows from (22), the compactness of Θ and the fact that $\frac{1}{n} \sum_{i=1}^n K(Z_i)$ has finite mean $\mathbb{E}_{\theta_0}[K(Z)]$.

A.4 Proof of Lemma 1

By $\hat{P}_n^c(h) = \tilde{P}_n(w \circ \hat{G}_n \times h)$ and by (9), we observe that,

$$\nu_n^c r(\cdot, t) - \tilde{\nu}_n(w \circ G \times r(\cdot, t)) = \sqrt{n} \tilde{P}_n \left\{ (w \circ \hat{G}_n - w \circ G) \times r(\cdot, t) \right\} .$$

Hence

$$|\nu_n^c r(\cdot, t) - \tilde{\nu}_n(w \circ G \times r(\cdot, t))| \leq \sqrt{n} \|w \circ \hat{G}_n - w \circ G\|_\infty \tilde{P}_n |r(\cdot, t)| .$$

By (Kallenberg, 2002, Theorem 14.15), we have $\sqrt{n} \|\hat{G}_n - G\|_\infty = O_P(1)$ and by assumption on w , $\sqrt{n} \|w \circ \hat{G}_n - w \circ G\|_\infty = O_P(1)$. Hence Condition (v) follows from (26).

A.5 Proof of Lemma 2

Since $\dot{h}(z, \cdot)$ denotes the gradient of $h(z, \cdot)$ in U , we have, for all $s, t \in U$,

$$h(z, t) = h(z, s) + \int_{u=0}^1 (t-s)^t \dot{h}(z, tu + s(1-u)) du .$$

By definition of r in (25) with $\Delta = \dot{h}(\cdot, \theta_0)$, we obtain

$$\begin{aligned} r(z, t) &= |t - \theta_0|^{-1} \left(h(z, t) - h(z, \theta_0) - (t - \theta_0)^T \dot{h}(z, \theta_0) \right) \\ &= \frac{(t - \theta_0)^T}{|t - \theta_0|} \int_{u=0}^1 \left(\dot{h}(z, tu + \theta_0(1-u)) - \dot{h}(z, \theta_0) \right) du . \end{aligned} \quad (36)$$

It also follows that

$$\begin{aligned} r(z, t) - r(z, s) &= \left(\frac{(t - \theta_0)^T}{|t - \theta_0|} - \frac{(s - \theta_0)^T}{|s - \theta_0|} \right) \int_0^1 \dot{h}(z, tu + \theta_0(1-u)) - \dot{h}(z, \theta_0) du \\ &\quad + \frac{(s - \theta_0)^T}{|s - \theta_0|} \int_0^1 \dot{h}(z, tu + \theta_0(1-u)) - \dot{h}(z, su + \theta_0(1-u)) du . \end{aligned} \quad (37)$$

We apply Lemma 1 so that it is sufficient to verify that (26) holds. Condition (26) holds because, by (27) and (36),

$$|r(z, t)| \leq K(z) |t - \theta_0| , \quad (38)$$

$\tilde{\nu}_n(K) = O_P(1)$. We now prove the second part of Condition (26) as an application of (Pollard, 1985, Lemma 4). Using (27), (37) and

$$\left| \frac{(t - \theta_0)^T}{|t - \theta_0|} - \frac{(s - \theta_0)^T}{|s - \theta_0|} \right| \leq 2 \frac{|t - s|}{|t - \theta_0|} ,$$

we have

$$|r(z, t) - r(z, s)| \leq 3K(z)|t - s| .$$

Thus, since $\mathbb{E} [|w \circ G(Z) K(Z)|]$ is bounded by (27) and the assumption on w , the class of functions $\mathcal{F} = \{w \circ G \times r(\cdot, t), t \in U\}$ satisfies the bracketing condition of Pollard (1985). The other condition for applying (Pollard, 1985, Lemma 4) is

$$\mathbb{E} \left[\sup_{|t-\theta_0| \leq R} |w \circ G(Z) r(Z, t)|^2 \right] \rightarrow 0 \quad \text{as } R \rightarrow 0 .$$

This follows from (38) and (27). This concludes the proof.

B A New CLT for L -statistics

To verify Condition (iv) of Theorem 4 a central limit theorem for linear combinations of (transformed) order statistics, so-called L -statistics, is required. The existing theorems can be divided into two groups. One approach consists in approximating the L -statistic by a sum of i.i.d. random variables such that the classical central limit theorem can be applied. This can be accomplished by using Hájek projection (van der Vaart, 1998, p. 318), defining pseudo-random variables (Shorack, 2000) or using the influence function (Serfling, 1984). All these theorems require that the L -statistic has the form $L_n = \sum_{i=1}^n c_{i,n} h(X_{(i,n)})$, where h is a monotone function or even the identity. Alternatively the L -statistic can be represented as a functional φ evaluated at the empirical distribution function \hat{F}_n , that is $L_n = \varphi(\hat{F}_n)$. In van der Vaart (1998), p.322, the functional delta method is used to derive the asymptotic distribution, in the case where the weights have the form $c_{i,n} = \int_{(i-1)/n}^{i/n} J(t) dt$ with some function J that vanishes at the borders of the interval $[0, 1]$.

The L -statistic encountered in Condition (iv) of Theorem 4 is not covered by the existing theorems. Notably, h is not necessarily monotone and the weights do not vanish at the borders. In the following we consider L -statistics of the form

$$L_n(h) \equiv \frac{1}{n} \sum_{i=1}^n w(i/n) h(X_{(i,n)}) = \frac{1}{n} \sum_{i=1}^n w \circ \hat{F}_n(X_i) h(X_i) ,$$

for functions $h : \mathbb{R} \rightarrow \mathbb{R}$ and $w : [0, 1] \rightarrow \mathbb{R}$ and where \hat{F}_n denotes the empirical distribution function associated with the sample (X_1, \dots, X_n) with distribution F .

The following theorem gives conditions under which the empirical process

$$\nu_n(h) = \sqrt{n}(L_n(h) - \mathbb{E}[w \circ F(X)h(X)]) \tag{39}$$

is asymptotically normal.

Theorem 5. *Suppose that*

- (i) *h is of bounded variation on bounded intervals;*
- (ii) *w is Lipschitz continuous on [0, 1];*
- (iii) *F is a continuous distribution function;*
- (iv) $\mathbb{E}[|h(X)|] < \infty$, $\mathbb{E}[w^2 \circ F(X)h^2(X)] < \infty$ and $\sigma^2(h) < \infty$ where $\sigma^2(\cdot)$ is defined by (40).

Then $\nu_n(h) \xrightarrow{d} \mathcal{N}(0, \sigma^2(h))$ with limit variance given by

$$\sigma^2(h) = \int_{\mathbb{R}^2} w \circ F(x)w \circ F(y)(F(x \wedge y) - F(x)F(y)) dh(x)dh(y) . \tag{40}$$

Proof. Denote by \mathcal{S} the set of all right-continuous functions $f : \mathbb{R} \rightarrow \mathbb{R}$ with finite left limit at all points endowed with the uniform norm. Let $DF \subset \mathcal{S}$ be the set of all distribution functions defined on \mathbb{R} . Define the functional φ on DF by

$$\varphi(G) \equiv \int hw \circ G dG = \int h \circ G^{-1}(u)w \circ G \circ G^{-1}(u)du ,$$

where $G^{-1}(u) \equiv \inf\{x \in \mathbb{R} : G(x) \geq u\}$ denotes the quantile function associated with G . Then $\nu_n(h) = \sqrt{n}(\varphi(\hat{F}_n) - \varphi(F))$. We further define on DF the functional

$$\psi(G) \equiv \int h \circ G^{-1} dW = \int h \circ G^{-1}(u)w(u)du ,$$

where $W(t) = \int_0^t w(u)du$. If $G \in DF$ is continuous, then $G \circ G^{-1} = Id$, and thus $\varphi(G) = \psi(G)$ for all continuous G . Then for any continuous $F \in DF$ we have

$$\nu_n(h) = \sqrt{n}(\varphi(\hat{F}_n) - \psi(\hat{F}_n)) + \sqrt{n}(\psi(\hat{F}_n) - \psi(F)) . \tag{41}$$

Using $\sup_{u \in (0,1)} |\hat{F}_n \circ \hat{F}_n^{-1}(u) - u| \leq \frac{1}{n}$, it follows that

$$\begin{aligned} \sqrt{n}|\varphi(\hat{F}_n) - \psi(\hat{F}_n)| &= \sqrt{n} \left| \int (w \circ \hat{F}_n \circ \hat{F}_n^{-1}(u) - w(u))h \circ \hat{F}_n^{-1}dt \right| \\ &\leq \frac{1}{\sqrt{n}} \int |h| d\hat{F}_n = n^{-3/2} \sum_{i=1}^n |h(X_i)| \xrightarrow{P} 0 , \end{aligned}$$

since $\mathbb{E}|h(X)| < \infty$. Now we show that the second term of (41) converges in distribution. To start with, suppose that h is of bounded variation and has compact support. Then the conditions of Lemma 22.10, van der Vaart (1998) are verified and hence ψ is Hadamard differentiable at F with Hadamard derivative ψ'_F defined on the tangent space of bounded continuous functions by

$$\psi'_F(G) \equiv - \int G \times w \circ F \, dh ,$$

when F is continuous. Furthermore, we know that $\sqrt{n}(\hat{F}_n - F)$ converges weakly to an F -Brownian bridge, denoted by B , with covariance function $\text{Cov}(B(x), B(y)) = F(x \wedge y) - F(x)F(y)$. Hence, from the functional delta method it follows that

$$\sqrt{n}(\psi(\hat{F}_n) - \psi(F)) \xrightarrow{d} - \int B \times w \circ F \, dh .$$

The random variable on the right-hand side is normally distributed with zero mean and variance

$$\begin{aligned} \sigma^2(h) &= \mathbb{E} \left[\left(\int w \circ F(x) B(x) \, dh(x) \right)^2 \right] \\ &= \int \int w \circ F(x) w \circ F(y) (F(x \wedge y) - F(x)F(y)) \, dh(x)dh(y) . \end{aligned}$$

Thus, the theorem holds for h with compact support.

Now let h be a function with arbitrary support satisfying the assumptions of the theorem. The assertion follows from an approximation of h by a sequence of functions with compact support. More precisely, for a given compact interval A and its complement A^c define the functions $\zeta_A = h\mathbb{1}_A$ and $\zeta_{A^c} = h - \zeta_A$. Note that

$$\nu_n(h) = \nu_n(\zeta_A) + \nu_n(\zeta_{A^c}) .$$

The first term on the right-hand side is asymptotically normal since ζ_A has compact support and is of bounded variation. In addition, by a convenient choice of A the second term can be made arbitrarily small. This yields the asymptotic normality of $\nu_n(h)$. More formally, we have

$$\begin{aligned} \nu_n(\zeta_{A^c}) &= \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \left\{ w \circ \hat{F}_n(X_i) - w \circ F(X_i) \right\} \zeta_{A^c}(X_i) \right) \\ &\quad + \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n w \circ F(X_i) \zeta_{A^c}(X_i) - \mathbb{E} [w \circ F(X) \zeta_{A^c}(X)] \right) . \end{aligned}$$

By (iv) the second term on the right side converges weakly to a normal distribution with zero mean and variance $\text{Var}(w \circ F(X)\zeta_{A^c}(X))$. Furthermore,

$$\begin{aligned} & \sqrt{n} \left| \frac{1}{n} \sum_{i=1}^n \left\{ w \circ \hat{F}_n(X_i) - w \circ F(X_i) \right\} \zeta_{A^c}(X_i) \right| \\ & \leq L \sup_{z \in \mathbb{R}} \left\{ \sqrt{n} \left| \hat{F}_n(z) - F(z) \right| \right\} \frac{1}{n} \sum_{i=1}^n |\zeta_{A^c}(X_i)| \\ & = O_p(1) \frac{1}{n} \sum_{i=1}^n |\zeta_{A^c}(X_i)| . \end{aligned}$$

Since $\mathbb{E}[|h(X)|] < \infty$ and $\mathbb{E}[(w \circ F(X)h(X))^2] < \infty$ by (iv), the set A can be chosen such that $\mathbb{E}|\zeta_{A^c}(X)|$ and $\text{Var}(w \circ F(X)\zeta_{A^c}(X))$ are arbitrarily small. Hence, for every $\varepsilon > 0, \eta > 0$ there exists a compact interval A such that ζ_{A^c} satisfies

$$\limsup_{n \rightarrow \infty} \mathbb{P} (|\nu_n(\zeta_{A^c})| > \eta) < \varepsilon .$$

Now let $(A_m)_m$ be a sequence of compact sets such that $\zeta_m \equiv h\mathbb{1}_{A_m}$ tends to h . By the dominated convergence theorem and (iv) we obtain that $\lim_{m \rightarrow \infty} \sigma^2(\zeta_m) = \sigma^2(h)$. Then Billingsley (1999), Theorem 3.2, p.28, implies the weak convergence of $\nu_n(h)$ to a centered normal distribution with variance $\sigma^2(h)$ given by (40). \square

If w is continuously differentiable on $[0, 1]$ and $\mathbb{E}[|h(X)|] < \infty$, it can be shown that

$$\mathbb{E}[w \circ F(X)h(X)] = \mathbb{E}[L_n(h)] + o(n^{-1/2}) .$$

Thus, in this case the centering constant $\mathbb{E}[w \circ F(X)h(X)]$ in (39) can be replaced by $\mathbb{E}[L_n(h)]$. We finally provide an estimator of the limit variance, which is useful for constructing asymptotic confidence intervals.

Proposition 1. *If h is continuous and w is differentiable with derivative \dot{w} , then Condition (iv) of Theorem 5 can be replaced by*

$$(iv') \quad \mathbb{E}[|h(X)|] < \infty, \mathbb{E}[w^2 \circ F(X)h^2(X)] < \infty \text{ and } \mathbb{E}[|\dot{w} \circ F(X)h(X)|] < \infty ,$$

and the limit variance writes

$$\begin{aligned} \sigma^2(h) &= \mathbb{E} [w^2 \circ F(X)h^2(X)] \\ &+ 2\mathbb{E} [\dot{w}_1 \circ F(X_1)\dot{w}_2 \circ F(X_2)h(X_1)h(X_2)\mathbb{1}\{X_1 > X_2\}] , \end{aligned}$$

where $w_1(t) = (1 - t)w(t)$ and $w_2(t) = tw(t)$ with derivatives \dot{w}_1 and \dot{w}_2 . If moreover $\mathbb{E}[w \circ F(X)\dot{w} \circ F(X)h^2(X)] < \infty$ and $\mathbb{E}[\dot{w}^2 \circ F(X)h^2(X)] < \infty$, then $\hat{\sigma}_n^2(h) \xrightarrow{P} \sigma^2(h)$ as $n \rightarrow \infty$, where

$$\begin{aligned} \hat{\sigma}_n^2(h) &= \frac{1}{n} \sum_{i=1}^n w^2(i/n)h^2(X_{(i,n)}) \\ &+ \frac{2}{n(n-1)} \sum_{i>j} \dot{w}_1(i/n)\dot{w}_2(j/n)h(X_{(i,n)})h(X_{(j,n)}) . \end{aligned} \quad (42)$$

References

- BILLINGSLEY, P. (1999). *Convergence of Probability Measures*. John Wiley & Sons Inc, 2nd ed.
- COATES, P. B. (1968). The correction of photon ‘pile-up’ in the measurement of radiative lifetimes. *Journal of Physics E: Scientific Instruments* **1**, 878–879.
- CRISSMAN, H. A. & STEINKAMP, J. A. (2000). Flow cytometric fluorescence lifetime measurements. In: *Cytometry* (DARZYNKIEWICZ, Z., ROBINSON, J. P. & CRISSMAN, H. A., eds.), vol. 63, chap. 6. Academic Press, 3rd ed.
- DEMPSTER, A. P., LAIRD, N. M. & RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)* **39**(1), 1–38.
- HALL, P. & SELINGER, B. (1981). Better estimates of exponential decay parameters. *Journal of Physical Chemistry* **85**(20), 2941–2946.
- HOUGAARD, P. (1984). Life table methods for heterogeneous populations: Distributions describing the heterogeneity. *Biometrika* **71**(1), 75–83.
- KALLENBERG, O. (2002). *Foundations of modern probability*. Probability and its Applications (New York). New York: Springer, second ed.
- KOSOROK, M. R., LEE, B. L. & FINE, J. P. (2004). Robust inference for univariate proportional hazards frailty regression models. *The Annals of Statistics* **32**(4), 1448–1491.
- LAKOWICZ, J. R. (1999). *Principles of Fluorescence Spectroscopy*. New York: Academic/Plenum.
- O’CONNOR, D. V. & PHILLIPS, D. (1984). *Time-correlated single photon counting*. London: Academic Press.
- OKANO, L., TEDESCO, A., TOUSO, A., OLIVEIRA, D., MAEDER, M. & BRINN, I. (2005). Abstracting reliable parameters from time-correlated

- single photon counting experiments. *Journal of Photochemistry and Photobiology A: Chemistry* **175**, 221–225.
- PATTING, M., WAHL, M., KAPUSTA, P. & ERDMANN, R. (2007). Dead-time effects in TCSPC data analysis. In: *Proceedings of SPIE*, vol. 6583.
- PIN, J.-P., AYOUB, M.-A., MAUREL, D., PERROY, J. & TRINQUET, E. (2007). Energy transfer technologies to monitor the dynamics and signaling properties of G-protein-coupled receptors in living cells. In: *Biophysical analysis of membrane proteins: investigating structure and function* (PEBAY-PEYROULA, E., ed.), chap. 13. Wiley-VCH.
- POLLARD, D. (1985). New ways to prove central limit theorems. *Econometric Theory* **1**(3), 295–313.
- REBAFKA, T. (2007). An MCMC approach for estimating a fluorescence lifetime with pile-up distortion. In: *GRETSI 2007: 21ème colloque sur le traitement du signal et des images*. Troyes, France.
- REBAFKA, T., ROUEFF, F. & SOULOUMIAC, A. (2008). Information bounds and MCMC parameter estimation for the pile-up model. Submitted.
- SERDYUK, I. N., ZACCAI, N. R. & ZACCAI, J. (2007). *Methods in molecular biophysics: structure, dynamics, function*. Cambridge University Press.
- SERFLING, R. J. (1984). Generalized L-, M-, and R-statistics. *The Annals of Statistics* **12**(1), 76–86.
- SHORACK, G. R. (2000). *Probability for Statisticians*. New York: Springer Verlag.
- SOULOUMIAC, A. (2007). Apparatus and method for characterizing a system for counting elementary particles. Patent number WO2007088173 (A1).
- TEICHER, H. (1961). Identifiability of mixtures. *The Annals of Mathematical Statistics* **32**(1), 244–248.
- TSODIKOV, A. (2003). A generalized self-consistency approach. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **65**(3), 759–774.
- TSODIKOV, A., IBRAHIM, J. & YAKOVLEV, A. (2003). Estimating cure rates from survival data: An alternative to two-component mixture models. *Journal of the American Statistical Association* **98**(464), 1063–1078.
- VALEUR, B. (2002). *Molecular Fluorescence*. Weinheim: Wiley-VCH.
- VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.
- WALKER, J. G. (2002). Iterative correction for 'pile-up' in single-photon lifetime measurements. *Optics Communications* **201**, 271–277.
- WARE, W. R., DOEMENY, L. J. & NEMZEK, T. L. (1973). Deconvolution

of fluorescence and phosphorescence decay curves. Least-squares method.
The Journal of Physical Chemistry **77**(17), 2038–2048.

YAKOVLEV, A. & TSODIKOV, A. (1996). *Stochastic Models of Tumor Latency and their Biostatistical Applications*. World Scientific.