Information bounds and MCMC parameter estimation for the pile-up model

Tabea Rebafka, François Roueff, Antoine Souloumiac

Information bounds and MCMC parameter estimation for the pile-up model

Tabea Rebafka[a,b,☆], François Roueff[a], Antoine Souloumiac[b]

[a] *Télécom Paristech, Département TSI / CNRS LTCI, 46 rue Barrault, 75634 Paris, France*
[b] *CEA, LIST, Laboratoire Processus Stochastiques et Spectres, 91191 Gif-sur-Yvette, France*

Abstract

This paper is concerned with the pile-up model defined as a nonlinear transformation of a distribution of interest. An observation of the pile-up model is the minimum of a random number of independent variables from the distribution of interest. One specific pile-up model is encountered in time-resolved fluorescence where only the first photon of a random number of photons is observed. In the first part of the paper the Cramér-Rao bound is studied to optimize the experimental conditions by choosing the best tuning parameter which is the average number of variables over which the minimum is taken. The implication is that the tuning parameter currently used in fluorescence does not minimize the acquisition time. However, data obtained at the optimal choice of the tuning parameter require an estimator adapted to the pile-up effect, therefore, an appropriate Gibbs sampler is presented. The covariance matrix of this estimator turns out to be close to the Cramér-Rao bound and hence the acquisition time may be reduced considerably.

Key words: Cramér-Rao bound; exponential mixture; fluorescence lifetime measurements; Gibbs sampler; Hellinger differentiability; TCSPC.

1. Introduction

There are applications where parameter estimation is difficult since the observed data do not form a sample from the distribution of interest due to noisy, missing, censored or truncated observations. In this paper a new type of problem involving missing data is studied. We suppose that the observed sample is distributed as the minimum of a random number of independent variables from the distribution of interest. Thus, the distribution of the observations results in a nonlinear distortion of the distribution of interest, usually called the pile-up effect. A formal definition of the pile-up model is given in Section 2.

The motivation for considering this type of problem arises from data encountered in time-resolved fluorescence, where the time from exciting a fluorescent molecule to the emission of a photon is called the fluorescence lifetime. This random time follows an exponential distribution whose parameter is called the lifetime constant and which depends on the molecule as well as its environment like pH, viscosity, polarity. Measurements for determining lifetime constants are obtained by the Time-Correlated Single-Photon Counting (TCSPC) technique, also called Single-Photon Timing (see O'Connor & Phillips, 1984; Valeur, 2002). Here a laser flash excites a random number of molecules but for technical reasons only the arrival time of the first fluorescent photon hitting the detector can be measured. Currently, the TCSPC method is used with a very low laser intensity such that the probability of detecting two or more photons per laser pulse becomes negligible, that is, only two events are likely: either no photon at all or exactly one photon hits the detector (see Lakowicz, 1999). Hence, no photons 'get lost' and the observations, where a photon is detected, may be considered as a sample distributed according to a mixture of exponential distributions. In this case, classical methods such as the EM algorithm (Dempster *et al.*, 1977) or

---

☆Corresponding author. Tel.: ++33 145817138; fax: +33 169087819.
Email address: tabea.rebafka@telecom-paristech.fr (T. Rebafka)

a Gibbs sampler (Gruet *et al.*, 1999) can be applied to estimate the parameters of interest.

The number of photons hitting the detector per excitation pulse is modeled by a Poisson distribution with parameter λ. The literature advises a low laser intensity such that the average number of photons λ hitting the detector is less than 0.05, see Valeur (2002). However, a small λ has the drawback that more than 95% of the measurements provide no information about the distribution of interest since there is no photon detected. Thus, the acquisition time is long which is unfavourable since fluorescence measurements are highly sensitive to unstable experimental conditions such as variations of the laser intensity (see Walker, 2002). We believe that the acquisition time can be greatly reduced by increasing the laser intensity which results in an increase in the proportion of observations containing information about the parameters. However, the pile-up effect can no longer be disregarded and we have to know how to handle pile-up affected data. In consequence, two specific questions arise: How does the pile-up affect the information contained in the data? How to handle the pile-up in the estimation procedure? The goal of this paper is to answer these two questions. In Section 3, we study the behaviour of standard information bounds as λ varies. In Section 4 we develop a Gibbs sampler to estimate a pile-up affected exponential mixture. This estimator is an adaption of the Gibbs sampler proposed by Gruet *et al.* (1999), using a new set of missing variables.

2. General setting

2.1. The pile-up model

Let $\{Y_k, \ k \geq 1\}$ be a sequence of independent positive random variables with distribution function F, and let N be a random variable taking its values in $\mathbb{N}$ independently of this sequence. We define the pile-up observation Z as the random variable taking its values in $\overline{\mathbb{R}_+} = \mathbb{R}_+ \cup \{\infty\}$ by

$$Z = \begin{cases} \min\{Y_1, \ldots, Y_N\} & \text{if } N > 0 \\ \infty & \text{if } N = 0 \ . \end{cases} \tag{1}$$

Lemma 1. *The distribution function G of Z is given by*

$$G(z) = 1 - M(1 - F(z)) \ , \quad z \in \mathbb{R}_+ \ , \tag{2}$$

where M is the moment generating function associated with N, defined as $M(u) = \mathrm{E}(u^N)$ for all $u \in [0, 1]$. If F admits a density f, G admits a density g with respect to the measure $\overline{\mathcal{L}_+}$, defined as the measure on $\overline{\mathbb{R}_+}$ which puts mass 1 at $\{\infty\}$ and whose restriction on $\mathbb{R}_+$ is the Lebesgue measure $\mathcal{L}_+$. Setting $\dot{M}(u) = \mathrm{E}[Nu^{N-1}]$ for all $[0, 1]$, the density g is given by

$$g(z) = \begin{cases} f(z)\dot{M}(1 - F(z)) & \text{if } z \in \mathbb{R}_+ \\ M(0) & \text{if } z = \infty \ . \end{cases} \tag{3}$$

PROOF. Relation (3) is obtained by computing $\mathrm{P}(Z > z) = \mathrm{E}[\mathrm{P}(Z < z|N)]$.

The general case is obtained by a density argument. Let (f_n) be a sequence of continuous densities tending to f in the L^1 sense, $\int |f_n - f| \, d\mathcal{L}_+ \to 0$ as $n \to \infty$. For each n we denote by F_n the distribution function of the density f_n, and by G_n and g_n the distribution function and density defined by (2) and (3), respectively. We get that F_n converge to F uniformly, $\sup_y |F_n(y) - F(y)| \to 0$ as $n \to \infty$ and, using the properties of M and $\dot{M}$, that $\sup_z |G_n(z) - G(z)| \to 0$ and $\int_z^\infty |g_n - g| \, d\mathcal{L}_+ \to 0$ as $n \to \infty$ for any z such that $F(z) > 0$. It follows that G admits g as a density. $\square$

We will compute Cramér-Rao information bounds for the pile-up model in a parametric setting. The most general setting that will be considered is the following one.

Assumption 1. *The distribution F belongs to a 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 distribution of N belongs to a parametric family $\{p_\lambda, \ \lambda \in \Lambda\}$ where $\Lambda \subset \mathbb{R}$ and, for a given λ, $p_\lambda(k) = \mathrm{P}(N = k)$ for all $k \in \mathbb{N}$.*

Definition 1. *Under Assumption 1, let us denote by $\{g_{\lambda,\theta}, \ (\lambda,\theta) \in \Lambda \times \Theta\}$ the corresponding collection of densities obtained by Relation (3). We call this model, that is dominated by $\overline{\mathcal{L}_+}$, the pile-up model associated with $\{f_\theta, \ \theta \in \Theta\}$ and $\{p_\lambda, \ \lambda \in \Lambda\}$.*

Remark 1. In an experimental setup it is impossible to wait indefinitely for some outcome. To obtain a more realistic data model, we could introduce a finite maximal observation time T, that is, instead of considering ∞, one observes finite T whenever there are no photons detected or when all photons arrive after T. Generally, T would be chosen rather large to avoid the censoring of events occurring after T. In the following we consider $T = \infty$ for simplicity except in Section 4 where we consider an estimation using censored data.

Remark 2. A case of particular interest, mainly motivated by the TCSPC application, is when N possesses a Poisson distribution with parameter $\lambda > 0$. Then

$$M(u) = e^{\lambda(u-1)} \quad \text{and} \quad \dot{M}(u) = \lambda e^{\lambda(u-1)}, \quad u \in [0,1] \,, \tag{4}$$

and we obtain the pile-up distribution function

$$G(z) = 1 - \exp\{-\lambda F(z)\} \,, \quad z \in \mathbb{R}_+ \,.$$

and the pile-up density

$$g(z) = \begin{cases} \lambda f(z) e^{-\lambda F(z)} & \text{if } z \in \mathbb{R}_+ \\ e^{-\lambda} & \text{if } z = \infty \,. \end{cases} \tag{5}$$

Concerning the family $\{f_\theta, \theta \in \Theta\}$, three particular cases will be considered: a scale family $f_\theta(t) = \theta f_1(\theta t)$ defined for all $\theta > 0$, the exponential family which is the particular scale family with $f_1(t) = e^{-t}$, and the family of mixtures of exponential density functions with given positive order, referred to as the multi-exponential case.

2.2. Hellinger differentiability

In maximum likelihood theory information bounds are well-known for regular models. The regular conditions involve the existence of two derivatives of the density function, together with interchangeability of derivation and integration (see e.g. Bickel & Doksum, 2001). However, even in simple models as the double exponential density, $\frac{1}{2}\exp(-|x - \theta|)$, some of the conditions are violated. For a gain in elegance and economy of assumptions the classical regularity conditions can be replaced by assumptions of Hellinger differentiability, that is, differentiability in norm of the square root of the density as an element of the L^2 space, defined formally hereafter.

Definition 2. *A collection of densities $\mathcal{P} = \{f_\theta, \theta \in \Theta\}$ with dominating measure ν and parameter set $\Theta \in \mathbb{R}^d$ is said to be Hellinger differentiable with respect to ν at a point θ_0, if the map $\theta \mapsto \xi_\theta(x) := \sqrt{f_\theta(x)}$ is differentiable in $L^2(\nu)$-norm at θ_0. That is, there exists a vector $\dot{\xi}_{\theta_0}(x)$ of $\mathbb{R}^d$–valued functions in $L^2(\nu)$ such that*

$$\xi_\theta(x) = \xi_{\theta_0}(x) + (\theta - \theta_0)^T \dot{\xi}_{\theta_0}(x) + r_\theta(x) \,,$$

where the $L^2(\nu)$-norm of r_θ satisfies $\|r_\theta\|_2 = o(|\theta - \theta_0|)$ as $\theta \to \theta_0$.

The information inequality for a model $\mathcal{P} = \{f_\theta, \theta \in \Theta\}$ bounds the variance of a parameter estimator $T(X)$ from below by an expression involving the expected value of the statistic and the Fisher information. Under classical regularity conditions this result is known as the Cramér-Rao bound.

Theorem 1 (Pollard, 2001). *Suppose that $\mathcal{P}$ is Hellinger differentiable at an interior point θ_0 of Θ with derivative $\dot{\xi}_{\theta_0}$. Suppose that an estimator $T(X)$ satisfies*

$$\sup_{\theta \in U_0} \mathrm{E}_\theta \left\{ T(X)^2 \right\} < \infty \,,$$

for some neighbourhood U_0 of θ_0. Then $\theta \mapsto \gamma_\theta := \mathrm{E}_\theta\{T(X)\}$ has a derivative at θ_0 given by

$$\nabla_\theta \gamma_{\theta_0} = 2 \int \xi_{\theta_0} \dot{\xi}_{\theta_0} T \mathrm{d}\nu \ .$$

Define the Fisher information matrix by $I_\theta = 4 \int \dot{\xi}_\theta \dot{\xi}_\theta^T \mathrm{d}\nu$. If, in addition, I_{θ_0} is nonsingular, then

$$\mathrm{Var}_{\theta_0}\{T(X)\} \geq \nabla_\theta \gamma_{\theta_0} I_{\theta_0}^{-1} \nabla_\theta \gamma_{\theta_0} \ . \tag{6}$$

We refer to Pollard (2001) or Bickel *et al.* (1998) for a proof of this result.

We now give some simple sufficient conditions for the pile-up model to be Hellinger differentiable and determine its Hellinger derivative. The following study of information bounds will mainly rely on this result.

Theorem 2. *Under Assumption 1 let $\{g_{\lambda,\theta}, \ (\lambda,\theta) \in \Lambda \times \Theta\}$ be the associated pile-up model. Define $\xi_\theta = \sqrt{f_\theta}$ and $\zeta_{\lambda,\theta} = \sqrt{g_{\lambda,\theta}}$ for all $\lambda \in \Lambda$ and $\theta \in \Theta$. Let $(\lambda_0, \theta_0) \in \Lambda \times \Theta$. If*

(i) $\{f_\theta, \ \theta \in \Theta\}$ is Hellinger differentiable with respect to $\mathcal{L}_+$ at θ_0 with Hellinger derivative $\dot{\xi}_{\theta_0}$,

(ii) $\mathrm{E}_{\lambda_0}(N^3) = \sum_{k \geq 0} k^3 p_{\lambda_0}(k) < \infty$,

(iii) $p_{\lambda_0}(1) > 0$,

(iv) the function $\lambda \mapsto \{p_\lambda(k)\}_{k \geq 1}$ is differentiable at λ_0 as a function defined on Λ with values in the space of sequences $\ell_1 = \{(u_k)_{k \in \mathbb{N}} : \ \sum_{k \geq 1} k|u_k| < \infty\}$,

(v) either (a) $p_{\lambda_0}(0) > 0$ and $\lambda \mapsto p_\lambda(0)$ is differentiable at λ_0 or (b) $p_\lambda(0) = 0$ for all λ in a neighbourhood of λ_0,

then $\{g_{\lambda,\theta}, \ (\lambda,\theta) \in \Lambda \times \Theta\}$ is Hellinger differentiable with respect to $\overline{\mathcal{L}_+}$ at (λ_0, θ_0) with Hellinger derivative

$$\dot{\zeta}_{\lambda_0,\theta_0} = [\partial_\theta \zeta_{\lambda_0,\theta_0} \ \ \partial_\lambda \zeta_{\lambda_0,\theta_0}]^T \ , \tag{7}$$

where $\partial_\theta \zeta_{\lambda_0,\theta_0}$ is given by

$$\partial_\theta \zeta_{\lambda_0,\theta_0}(z) = \begin{cases} \dot{\xi}_{\theta_0}(z)\dot{M}_{\lambda_0}^{1/2}(1 - F_{\theta_0}(z)) + \frac{\xi_{\theta_0}(z)\ddot{M}_{\lambda_0}(1-F_{\theta_0}(z))}{\dot{M}_{\lambda_0}^{1/2}(1-F_{\theta_0}(z))} \int_z^\infty \xi_{\theta_0}(t)\dot{\xi}_{\theta_0}(t)\mathrm{d}t & \text{if } z \in \mathbb{R}_+ \\ 0 & \text{if } z = \infty \ , \end{cases} \tag{8}$$

with $\dot{M}_\lambda(u) = \mathrm{E}_\lambda(Nu^{N-1})$, $\ddot{M}_\lambda(u) = \mathrm{E}_\lambda\{N(N-1)u^{N-2}\}$ and

$$\partial_\lambda \zeta_{\lambda_0,\theta_0}(z) = \begin{cases} \frac{\xi_{\theta_0}(z)\check{M}_{\lambda_0}(\bar{F}_{\theta_0}(z))}{2\dot{M}_{\lambda_0}^{1/2}(\bar{F}_{\theta_0}(z))} & \text{if } z \in \mathbb{R}_+ \\ \frac{\check{p}_{\lambda_0}(0)}{2p_{\lambda_0}^{1/2}(0)} & \text{if } z = \infty \ , \end{cases} \tag{9}$$

with the convention $0/0 = 0$ and where $\check{p}_{\lambda_0}(k)$ denotes the derivative of $\lambda \mapsto p_\lambda(k)$ at λ_0 and

$$\check{M}_{\lambda_0}(u) = \sum_{k=1}^\infty k \, u^{k-1} \check{p}_{\lambda_0}(k), \quad u \in [0,1] \ . \tag{10}$$

Remark 3. The pile-up model $\{g_{\theta,\lambda_0}, \theta \in \Theta\}$ parameterized by θ with a given λ_0 is Hellinger differentiable at a point θ_0 under the conditions (i)–(iii) and the Hellinger derivative $\partial_\theta \zeta_{\lambda_0,\theta_0}$ is given by (8), see the proof of the Theorem.

4

Remark 4. Condition (iv) implies that, for all $k \geq 1$, $\lambda \mapsto p_\lambda(k)$ is differentiable at λ_0, and that the sequence made of their derivatives $\{\breve{p}_{\lambda_0}(k)\}_{k \geq 1}$ belongs to ℓ_1. Hence $\breve{M}_{\lambda_0}(u)$ is well defined in (10) and is equal to the derivative of $\lambda \mapsto \dot{M}_\lambda(u)$ at λ_0. Conversely, if $\lambda \mapsto p_\lambda(k)$ is differentiable in a neighbourhood U_0 of λ_0 for all $k \geq 1$ and there exist $k_0 > 0$, $(h_k)_{k \geq k_0}$ and $v_0 : \mathbb{R}_+ \to \mathbb{R}_+$ such that, for all $k \geq k_0$ and $\lambda \in U_0$,

$$|\breve{p}_{\lambda_0}(k) - \breve{p}_\lambda(k)| \leq v_0(|\lambda - \lambda_0|)h_k \ , \tag{11}$$

with $v_0(r) \to 0$ as $r \downarrow 0$ and $\sum_{k>0} kh_k < \infty$, then (iv) is satisfied. The bound (11) holds for instance when $\lambda \mapsto p_\lambda(k)$ is twice differentiable on U_0 with $v_0(r) = r$ and h_k defined as a local bound of the second derivative $|\partial_\lambda^2 p_\lambda(k)|$, for instance

$$h_k = \sup_{\lambda \in U_0} |\partial_\lambda^2 p_\lambda(k)| \ .$$

Remark 5. In particular, in the Poisson case, $p_\lambda(k) = \mathrm{e}^{-\lambda}\lambda^k/k!$, Conditions (ii)–(v) hold. Then Theorem 2 implies that the Hellinger differentiability of the pile-up model $\{g_{\lambda,\theta}, \ (\lambda,\theta) \in \Lambda \times \Theta\}$ at any (λ_0, θ_0) follows from that of the model $\{f_\theta, \ \theta \in \Theta\}$ at θ_0. Moreover, we have

$$\breve{M}_\lambda(u) = \mathrm{e}^{\lambda(u-1)}\{1 + \lambda(u-1)\} \ . \tag{12}$$

PROOF (THEOREM 2). We consider below the differentiability of $\zeta_{\lambda,\theta}$ as an element of $L^2(\mathcal{L}_+)$. The extension to $L^2(\overline{\mathcal{L}_+})$ is immediate in (8) because $\zeta_{\lambda,\theta}(\infty)$ does not depend on θ and in (9) as an application of Condition (v). Let us first show that the model $\{g_{\theta,\lambda_0}, \theta \in \Theta\}$ is Hellinger differentiable at θ_0 for some given λ_0. Let $L^\infty(\mathcal{L}_+)$ denote the space of functions which are $\mathcal{L}_+$-essentially bounded, and $\|\cdot\|_\infty$ denote the corresponding norm. By Lemma 1, we have

$$\zeta_{\lambda_0,\theta} = \xi_\theta \dot{M}_{\lambda_0}^{1/2}(1 - F_\theta) \ .$$

Consider the function $G : L^2(\mathcal{L}_+) \to L^\infty(\mathcal{L}_+)$ defined by

$$G(\xi)(z) = \int_z^\infty \xi^2(t) \, \mathrm{d}t \ ,$$

and the function H with values in $L^\infty(\mathcal{L}_+)$, defined on the unit ball of $L^\infty(\mathcal{L}_+)$ by

$$H(\mathbf{u}) = \dot{M}_{\lambda_0} \circ \mathbf{u} \ .$$

By (i) and using the differentiability of $(\mathbf{u}, \mathbf{v}) \mapsto \mathbf{u} \times \mathbf{v}$ defined from $L^2(\mathcal{L}_+) \times L^\infty(\mathcal{L}_+)$ to $L^2(\mathcal{L}_+)$, the differentiability (8) at θ_0 follows from that of $\theta \mapsto \sqrt{H} \circ G \circ \xi_\theta$ as a function valued in $L^\infty(\mathcal{L}_+)$. Observe that G is differentiable at every point $\xi \in L^2(\mathcal{L}_+)$, with derivative function $\dot{G}_\xi : L^2(\mathcal{L}_+) \to L^\infty(\mathcal{L}_+)$ given by

$$\dot{G}_\xi(\mathbf{u})(z) = 2 \int_z^\infty \xi(t)\mathbf{u}(t) \, \mathrm{d}t, z \in \mathbb{R}_+ \ .$$

We next show that $\sqrt{H}$ is differentiable at any point of the set of non-negative functions in the unit ball of $L^\infty(\mathcal{L}_+)$ and conclude (11) by using that the derivative of a composition is the composition of derivatives. By (ii), the function $\dot{M}_{\lambda_0}$ is twice continuously differentiable on $[-1, 1]$ with first derivative $\ddot{M}_{\lambda_0}$ and second derivative bounded by $\mathrm{E}_{\lambda_0}\{N(N-1)(N-2)\}$. Hence, for all $t, s \in [-1, 1]$,

$$\dot{M}_{\lambda_0}(t) = \dot{M}_{\lambda_0}(s) + (t-s)\ddot{M}_{\lambda_0}(s) + r(t-s) \ , \tag{13}$$

where $|r(x)| \leq \mathrm{E}_{\lambda_0}\{N(N-1)(N-2)\}x^2/2$. Applying (13), we obtain that H is differentiable at any $\mathbf{u}$ in the unit ball of $L^\infty(\mathcal{L}_+)$ with derivative $\mathbf{v} \mapsto \mathbf{v} \times \ddot{M}_{\lambda_0} \circ \mathbf{u}$. Similarly, using (iii), since $\dot{M}_{\lambda_0}(u) \geq p_{\lambda_0}(1)$ for all $u \in [0, 1]$ and the square root function is twice continuously differentiable on

$\mathbb{R}_+$ with a second derivative uniformly bounded on $[p_{\lambda_0}(1), \infty)$, we get that $\sqrt{H}$ is differentiable at any non-negative $\mathbf{u}$ in the unit ball of $L^\infty(\mathcal{L}_+)$ with derivative $\mathbf{v} \mapsto (\mathbf{v}/2)\ddot{M}_{\lambda_0}\dot{M}_{\lambda_0}^{-1/2} \circ \mathbf{u}$.

Let us now prove the Hellinger differentiability of $\zeta_{\lambda,\theta}$ at (λ_0, θ_0). We write

$$\zeta_{\lambda,\theta} - \zeta_{\lambda_0,\theta_0} = \zeta_{\lambda,\theta} - \zeta_{\lambda_0,\theta} + \zeta_{\lambda_0,\theta} - \zeta_{\lambda_0,\theta_0} . \tag{14}$$

For any $\lambda \in \Lambda$ we have

$$\sup_{\theta \in \Theta} \left\| \dot{M}_\lambda(1 - F_\theta) - \dot{M}_{\lambda_0}(1 - F_\theta) - (\lambda - \lambda_0)\check{M}_{\lambda_0}(1 - F_\theta) \right\|_\infty$$
$$\leq \sum_{k \geq 1} |p_\lambda(k) - p_{\lambda_0}(k) - (\lambda - \lambda_0)\check{p}_{\lambda_0}(k)| = o(|\lambda - \lambda_0|) ,$$

where the o-term holds as $\lambda \to \lambda_0$ and follows from Condition (iv). Since for all $u \in [0, 1]$, $\dot{M}_{\lambda_0}(u) \geq p_{\lambda_0}(1) > 0$ and $|\check{M}_{\lambda_0}(u)| \leq \sum_{k \geq 1} k|\check{p}_{\lambda_0}(k)| < \infty$, we obtain that, as $\lambda \to \lambda_0$,

$$\sup_{\theta \in \Theta} \left\| \dot{M}_\lambda^{1/2}(1 - F_\theta) - \dot{M}_{\lambda_0}^{1/2}(1 - F_\theta) - (\lambda - \lambda_0)\frac{\check{M}_{\lambda_0}(1 - F_\theta)}{2\dot{M}_{\lambda_0}^{1/2}(1 - F_\theta)} \right\|_\infty = o(|\lambda - \lambda_0|) .$$

Thus, since $\|\xi_\theta\|_2 = 1$ for all $\theta \in \Theta$ and $\|\mathbf{uv}\|_2 \leq \|\mathbf{u}\|_2\|\mathbf{u}\|_\infty$, multiplying the previous equation by ξ_θ yields, as $\lambda \to \lambda_0$,

$$\sup_{\theta \in \Theta} \left\| \zeta_{\lambda,\theta} - \zeta_{\lambda_0,\theta} - (\lambda - \lambda_0)\frac{\xi_\theta \check{M}_{\lambda_0}(1 - F_\theta)}{2\dot{M}_{\lambda_0}^{1/2}(1 - F_\theta)} \right\|_2 = o(|\lambda - \lambda_0|) .$$

Observe that $\dot{M}_{\lambda_0}$ is continuous and positively lower bounded on $[0, 1]$, $\check{M}_{\lambda_0}$ is continuous on $[0, 1]$ since $\sum_{k \geq 1} k|\check{p}_{\lambda_0}(k)| < \infty$, $\theta \mapsto \xi_\theta$ is continuous at θ_0 as a map valued in $L^2(\mathcal{L}_+)$ and $\theta \mapsto 1 - F_\theta$ is continuous at θ_0 as a map valued in $L^\infty(\mathcal{L}_+)$. Hence, as $\theta \to \theta_0$,

$$\left\| \frac{\xi_\theta \check{M}_{\lambda_0}(1 - F_\theta)}{2\dot{M}_{\lambda_0}^{1/2}(1 - F_\theta)} - \frac{\xi_{\theta_0} \check{M}_{\lambda_0}(1 - F_{\theta_0})}{2\dot{M}_{\lambda_0}^{1/2}(1 - F_{\theta_0})} \right\|_2 \to 0 .$$

On the other hand, the previously proved Hellinger differentiability of $\zeta_{\lambda_0,\theta_0}$ at θ_0 implies that, as $\theta \to \theta_0$,

$$\left\| \zeta_{\lambda_0,\theta} - \zeta_{\lambda_0,\theta_0} - (\theta - \theta_0)^T \partial_\theta \zeta_{\lambda_0,\theta_0} \right\|_2 = o(\|\theta - \theta_0\|) .$$

The two last displays and (14) conclude the proof. $\square$

3. The effect of pile-up on information bounds

This section provides a study of the information inequality for the pile-up model. Our special interest lies in the influence of the parameter λ on the information contained in the data, since in the context of fluorescence measurements λ presents the average number of photons hitting the detector and can be coarsely tuned in the experimental setup. At a low intensity λ the acquisition time is long since most of the time no fluorescent photon is detected ($N = 0$ in (1)). At a high intensity the number of photons N hitting the detector is very large and the distribution of Z is almost degenerated. Intuitively, we may find that in both of these extreme cases the information about the parameter of interest θ_0 is quite poor. The question of finding an optimal value for λ arises.

The main result of this section is Theorem 4: If the underlying density f_{θ_0} belongs to a scale family, then the Cramér-Rao bound of the scale parameter θ_0 achieves a minimum at a point λ_{opt} which is the

same for all $\theta_0 > 0$. Evidently, this is of great practical use for fluorescence measurements when F_{θ_0} is considered to be the exponential distribution. Theorem 4 suggests tuning the parameter as close to λ_{opt} as possible independently of the unknown parameter θ_0, to minimize the asymptotic variance of the maximum likelihood estimator of θ_0. In Subsection 3.4 the optimal intensity is determined at $\lambda_{\mathrm{opt}} \approx 1.32$ using numerical evaluations of the information bounds. Carrying out the experiment at the intensity λ_{opt}, instead of $\lambda < 0.05$ as currently done, results in an enormous increase in information and hence reduces the acquisition time considerably. In the multi-exponential case the Cramér-Rao bound depends on the parameter θ_0. However, simulations reveal that an intensity in the range of $[1, 2]$ remains a good choice.

To assure the existence of the Fisher information we will make the following assumption.

Assumption 2. *In addition to Assumption 1 suppose that for a given (λ_0, θ_0) the density families $\{f_\theta, \; \theta \in \Theta\}$ and $\{p_\lambda, \; \lambda \in \Lambda\}$ satisfy Condition (i) and Conditions (ii)–(v) in Theorem 2, respectively. Moreover, we assume that Θ and Λ are open sets.*

The implication of Theorem 2 is that the pile-up model $\{g_{\lambda,\theta}, \; (\lambda, \theta) \in \Lambda \times \Theta\}$ is Hellinger differentiable at (λ_0, θ_0), with derivative $\dot{\zeta}_{\lambda_0,\theta_0}$ given by (7), (8) and (9). We denote the corresponding Fisher information matrix of the pile-up model by

$$
I(\lambda_0, \theta_0) = 4 \int \dot{\zeta}_{\lambda_0,\theta_0} \dot{\zeta}^T_{\lambda_0,\theta_0} \mathrm{d}\overline{\mathcal{L}_+} = \left[\begin{array}{cc} I_\lambda(\lambda_0, \theta_0) & I_{\lambda,\theta}(\lambda_0, \theta_0) \\ I^T_{\lambda,\theta}(\lambda_0, \theta_0) & I_\theta(\lambda_0, \theta_0) \end{array} \right], \tag{15}
$$

where, in the given block description, $I_\lambda(\lambda_0, \theta_0)$ is defined as a scalar, $I_{\lambda,\theta}(\lambda_0, \theta_0)$ as a row vector of length d and $I_\theta(\lambda_0, \theta_0)$ as a $d \times d$ matrix. If $I(\lambda_0, \theta_0)$ is nonsingular, its inverse, which appears in the information bound (6), is denoted by

$$
CR(\lambda_0, \theta_0) = I^{-1}(\lambda_0, \theta_0) = \left[\begin{array}{cc} CR_\lambda(\lambda_0, \theta_0) & CR_{\lambda,\theta}(\lambda_0, \theta_0) \\ CR^T_{\lambda,\theta}(\lambda_0, \theta_0) & CR_\theta(\lambda_0, \theta_0) \end{array} \right], \tag{16}
$$

with a similar block description as above, that is, $CR_\lambda(\lambda_0, \theta_0)$ is a scalar, $CR_{\lambda,\theta}(\lambda_0, \theta_0)$ is a row vector of length d and $CR_\theta(\lambda_0, \theta_0)$ is a $d \times d$ matrix.

3.1. Fisher information of λ

In the pile-up model $\{g_{\lambda,\theta}, \; (\lambda, \theta) \in \Lambda \times \Theta\}$ the Fisher information of λ is independent of the underlying distribution F_θ, as shown in the following result.

Theorem 3. *Under Assumption 2 the Fisher information term $I_\lambda(\lambda_0, \theta_0)$ of the pile-up model $\{g_{\lambda,\theta}, \; (\lambda, \theta) \in \Lambda \times \Theta\}$ defined by (15) is independent of θ_0 and $\{f_\theta, \; \theta \in \Theta\}$. More precisely,*

$$
I_\lambda(\lambda_0, \theta_0) = \int_0^1 \frac{\check{M}^2_{\lambda_0}(u)}{\dot{M}_{\lambda_0}(u)} \mathrm{d}u + \frac{\check{p}^2_{\lambda_0}(0)}{p_{\lambda_0}(0)} . \tag{17}
$$

PROOF. Using (9), we obtain

$$
I_\lambda(\lambda_0, \theta_0) = 4 \int (\partial_\lambda \zeta_{\lambda_0,\theta_0})^2 \mathrm{d}\overline{\mathcal{L}_+} = \int_0^\infty \frac{\xi^2_{\theta_0}(z)\check{M}^2_{\lambda_0}(1 - F_{\theta_0}(z))}{\dot{M}_{\lambda_0}(1 - F_{\theta_0}(z))} \mathrm{d}z + \frac{\check{p}^2_{\lambda_0}(0)}{p_{\lambda_0}(0)} .
$$

Let X be a random variable with density $\xi^2_{\theta_0}$. As its distribution function F_{θ_0} is continuous, $F_{\theta_0}(X)$ has uniform distribution on $[0, 1]$ (see Devroye, 1986). Consequently,

$$
\int_0^\infty \frac{\xi^2_{\theta_0}(z)\check{M}^2_{\lambda_0}(1 - F_{\theta_0}(z))}{\dot{M}_{\lambda_0}(1 - F_{\theta_0}(z))} \mathrm{d}z = \mathrm{E}\left\{ \frac{\check{M}^2_{\lambda_0}(1 - F_{\theta_0}(X))}{\dot{M}_{\lambda_0}(1 - F_{\theta_0}(X))} \right\} = \int_0^1 \frac{\check{M}^2_{\lambda_0}(u)}{\dot{M}_{\lambda_0}(u)} \mathrm{d}u .
$$

This completes the proof. $\square$

Remark 6. In the Poisson case where $\breve{M}$ and $\dot{M}$ are given by (4) and (12), we have

$$I_\lambda(\lambda_0, \theta_0) = \frac{1}{\lambda_0} \int_0^1 e^{\lambda_0(u-1)}(1 + \lambda_0(u-1))du + e^{-\lambda_0} = \frac{1 - e^{-\lambda_0}}{\lambda_0^2} . \qquad (18)$$
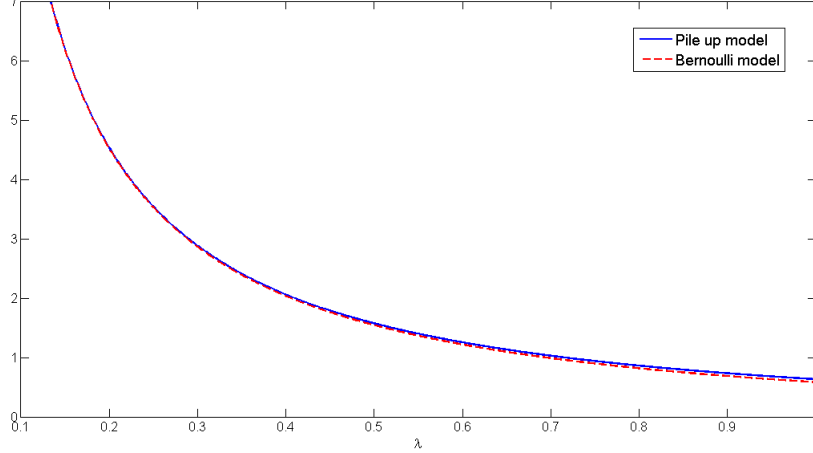

Figure 1: Fisher information $\lambda \mapsto I_\lambda(\theta_0, \lambda)$ of λ for the pile-up model in the Poisson case versus the Fisher information of the Bernoulli model with parameter $e^{-\lambda}$.

The independence of I_λ of the underlying distribution F_{θ_0} may appear surprising, but it indicates that the main part of information about λ is the proportion of observations equal to infinity. This quantity is a natural estimator of $p_\lambda(0)$, the probability of detecting no photons, which is independent of F_{θ_0}. A comparison of the information I_λ in the Poisson case to the Fisher information of a Bernoulli model with parameter $p_\lambda(0) = e^{-\lambda}$, equal to $\tilde{I}_\lambda = (e^\lambda - 1)^{-1}$, indicates that for values of λ close to 0 the two information values almost coincide, see Figure 1. Furthermore, observe that the Fisher information I_λ of the pile-up model slightly exceeds the information $\tilde{I}_\lambda$ of the Bernoulli model, for all $\lambda > 0$. This is clear because a Bernoulli model can be obtained from the pile-up observations by projecting them on the pair $\{0, 1\}$.

3.2. Scale family

We now consider the pile-up model where the underlying density f_{θ_0} belongs to a scale family $\{\theta f_1(\theta \cdot), \theta > 0\}$ with $d = 1$. We show that the Cramér-Rao bound $\lambda \mapsto CR_\theta(\lambda, \theta_0)$ achieves a minimum at a point λ_{opt} which is the same for all $\theta_0 > 0$.

Lemma 2. *Let $\{f_\theta, \theta > 0\}$ be a scale family. Then, for any $\theta_0 > 0$, $\{f_\theta, \theta > 0\}$ is Hellinger differentiable at $\theta = \theta_0$ if and only if $\{f_\theta, \theta > 0\}$ is Hellinger differentiable at $\theta = 1$, and $\dot{\xi}_{\theta_0}(x) = \frac{1}{\sqrt{\theta_0}}\dot{\xi}_1(\theta_0 x)$.*

PROOF. Let $\{f_\theta, \theta > 0\}$ be Hellinger differentiable at $\theta = 1$ with derivative $\dot{\xi}_1(x)$ and let $\theta_0 > 0$. Denote

$$r_\theta(x) = \xi_\theta(x) - \xi_1(x) - (\theta - 1)\dot{\xi}_1(x) ,$$

then $\|r_\theta\|_2 = o(|\theta - 1|)$ as $\theta \to 1$. As $\{f_\theta, \theta > 0\}$ is a scale family, one can show by changing variables that

$$\left\| \xi_\theta(x) - \xi_{\theta_0}(x) - \frac{(\theta - \theta_0)}{\sqrt{\theta_0}}\dot{\xi}_1(\theta_0 x) \right\|_2^2 = \int_0^\infty \left(\xi_{\theta/\theta_0}(u) - \xi_1(u) - \left(\frac{\theta}{\theta_0} - 1\right)\dot{\xi}_1(u) \right)^2 du$$

$$= \|r_{\theta/\theta_0}\|_2^2 .$$

8

Since $\|r_{\theta/\theta_0}\|_2 = o(|\theta/\theta_0 - 1|) = o(|\theta - \theta_0|)$, as $\theta \to \theta_0$, we obtain the Hellinger differentiability at θ_0 with the given derivative. The converse is shown in the same way. $\square$

Theorem 4. *Let $\{f_\theta, \theta > 0\}$ be a scale family. If Assumption 2 holds for some (λ_0, θ_0) and the Fisher information matrix $I(\lambda_0, \theta_0)$ is nonsingular, then Assumption 2 and the invertibility of $I(\lambda_0, \theta_0)$ hold for any positive value of θ_0 and the Cramér-Rao bound is given by*

$$CR(\lambda_0, \theta_0) = \begin{bmatrix} CR_\lambda(\lambda_0, 1) & \theta_0 CR_{\lambda,\theta}(\lambda_0, 1) \\ \theta_0 CR_{\lambda,\theta}^T(\lambda_0, 1) & \theta_0^2 CR_\theta(\lambda_0, 1) \end{bmatrix} . \tag{19}$$

Consequently, $\arg\min\limits_{\lambda_0} CR_\theta(\lambda_0, \theta_0) = \arg\min\limits_{\lambda_0} CR_\theta(\lambda_0, 1)$, *and thus the value of λ_0 minimizing the Cramér-Rao bound $CR_\theta(\lambda_0, \theta_0)$ does not depend on θ_0.*

PROOF. Lemma 2 implies the Hellinger differentiablity of $\{f_\theta, \theta > 0\}$ at any $\theta_0 > 0$ with derivative given by

$$\dot{\xi}_{\theta_0}(x) = \frac{1}{\sqrt{\theta_0}} \dot{\xi}_1(\theta_0 x) .$$

By applying Theorem 2, we have

$$\partial_\theta \zeta_{\lambda_0, \theta_0}(\infty) = \partial_\theta \zeta_{\lambda_0, 1}(\infty) = 0 \quad \text{and} \quad \partial_\lambda \zeta_{\lambda_0, \theta_0}(\infty) = \partial_\lambda \zeta_{\lambda_0, 1}(\infty) .$$

Further, using that $F_{\theta_0}(z) = F_1(\theta_0 z)$, $\xi_{\theta_0}(x) = \sqrt{\theta_0}\xi_1(\theta_0 x)$ and

$$\int_z^\infty \xi_1(\theta_0 t)\dot{\xi}_1(\theta_0 t) \, \mathrm{d}t = \frac{1}{\theta_0} \int_{\theta_0 z}^\infty \xi_1(t)\dot{\xi}_1(t) \, \mathrm{d}t, \quad z \in \mathbb{R}_+ ,$$

gives for all $z \in \mathbb{R}_+$,

$$\partial_\theta \zeta_{\lambda_0, \theta_0}(z) = \frac{1}{\sqrt{\theta}} \partial_\theta \zeta_{\lambda_0, 1}(\theta z) \quad \text{and} \quad \partial_\lambda \zeta_{\lambda_0, \theta_0}(z) = \sqrt{\theta} \partial_\lambda \zeta_{\lambda_0, 1}(\theta z) .$$

Therefore,

$$I_\theta(\lambda_0, \theta_0) = 4 \int_0^\infty (\partial_\theta \zeta_{\lambda_0, \theta_0}(z))^2 \mathrm{d}z = \frac{4}{\theta_0^2} \int_0^\infty (\partial_\theta \zeta_{\lambda_0, 1}(z))^2 \mathrm{d}z = \frac{1}{\theta_0^2} I_\theta(\lambda_0, 1) , \tag{20}$$

and, likewise, $I_{\lambda,\theta}(\lambda_0, \theta_0) = \frac{1}{\theta_0} I_{\lambda,\theta}(\lambda_0, 1)$. Note that $I_\lambda(\lambda_0, \theta_0) = I_\lambda(\lambda_0, 1)$, by Theorem 3. The theorem follows by applying the preceding relations to the inverse of the 2×2 Fisher information matrix

$$CR(\lambda_0, \theta_0) = \frac{1}{I_\lambda(\lambda_0, \theta_0)I_\theta(\lambda_0, \theta_0) - I_{\lambda,\theta}(\lambda_0, \theta_0)^2} \begin{bmatrix} I_\theta(\lambda_0, \theta_0) & -I_{\lambda,\theta}(\lambda_0, \theta_0) \\ -I_{\theta,\lambda}(\lambda_0, \theta_0) & I_\lambda(\lambda_0, \theta_0) \end{bmatrix} .$$

$\square$

3.3. Asymptotic behaviour in the Poisson case

Our main motivation is to investigate whether, from an information point of view, it is legitimate to collect fluorescence lifetime data at a very low intensity λ. Recall that in the fluorescence context N follows a Poisson distribution $\mathrm{Poi}(\lambda_0)$ and literature advises to use $\lambda_0 < 0.05$. The question is, whether the currently used low intensity λ_0 is optimal. We now provide an approximation of the information bound of the parameter θ_0 as λ_0 tends to 0 in the Poisson case, and deduce that the Cramér-Rao bound increases greatly as $\lambda_0 \to 0$.

In the Poisson case the parameter λ coincides with the expected number of photons hitting the detector per light pulse. If λ is sufficiently small the probability of detecting two or more photons becomes

9

negligible and the model is similar to the distribution where the value ∞ is observed with probability $\mathrm{e}^{-\lambda}$ and otherwise a random variable Y with density f_θ. Formally, we define the following random variable

$$X = \begin{cases} Y & \text{with probability } 1 - \mathrm{e}^{-\lambda} \\ \infty & \text{with probability } \mathrm{e}^{-\lambda} . \end{cases} \tag{21}$$

Hence, for a small λ_0, the density of X is a good approximation of g_{λ_0,θ_0}. Moreover, the corresponding information matrix is easier to calculate and to inverse.

Lemma 3. *Let, for all $\lambda > 0$,*

$$A_\lambda = \begin{bmatrix} a_\lambda & \lambda b_\lambda^T \\ \lambda b_\lambda & \lambda d_\lambda \end{bmatrix} ,$$

where $a_\lambda \in \mathbb{R}$, $b_\lambda = (b_{\lambda,1}, \ldots, b_{\lambda,p})^T \in \mathbb{R}^p$ and $d_\lambda \in \mathbb{R}^{p \times p}$. Suppose that

$$b_\lambda = O(1), \quad \lambda a_\lambda \to 1 \text{ and } d_\lambda \to d \text{ when } \lambda \to 0 ,$$

where d is an invertible matrix. Then, when $\lambda \to 0$,

$$\det(A_\lambda) \sim \lambda^{p-1} \det(d_\lambda) .$$

Further, denote the two diagonal blocks of the inverse of A_λ by m_λ and M_λ, where m_λ is a scalar and M_λ a $p \times p$ matrix,

$$A_\lambda^{-1} = \begin{bmatrix} m_\lambda & \cdots \\ \vdots & M_\lambda \end{bmatrix} .$$

Then, when $\lambda \to 0$,

$$m_\lambda \sim \lambda \quad \text{and} \quad M_\lambda \sim \frac{d^{-1}}{\lambda} .$$

PROOF. Using the Laplace expansion for the determinant of A_λ along the first row, we obtain with the notation $d_\lambda^{-(j)}$ for the $p \times (p-1)$ matrix that results from d_λ by removing the j-th column,

$$\det(A_\lambda) = a_\lambda \lambda^p \det(d_\lambda) + \sum_{j=1}^{p} (-1)^{j+1} b_{\lambda,j} \lambda^{p+1} \det\left(\left[b_\lambda^T, d_\lambda^{-(j)}\right]\right) .$$

Further,

$$\det\left(\left[b_\lambda^T, d_\lambda^{-(j)}\right]\right) = \sum_{k=1}^{p} (-1)^{k+1} b_{\lambda,k} \det\left(d_\lambda^{-(k,j)}\right) ,$$

where the $(p-1) \times (p-1)$ matrix $d_\lambda^{-(k,j)}$ results from d_λ by removing the k-th row and the j-th column. By the assumptions of the theorem it follows $\det(A_\lambda) = \lambda^p a_\lambda \det(d_\lambda) + O(\lambda^{p+1}) \sim \lambda^{p-1} \det(d)$, as $\lambda \to 0$. For the second part of the theorem, we use the representation of the inverse matrix using minors. For all $i, j > 1$ we have, similarly as above,

$$\left(A_\lambda^{-1}\right)_{i,j} = \frac{(-1)^{i+j}}{\det(A_\lambda)} \left(a_\lambda \lambda^{p-1} \det\left(d_\lambda^{-(i-1,j-1)}\right) + O(\lambda^p)\right)$$

$$= \frac{(-1)^{i+j} \det\left(d_\lambda^{-(i-1,j-1)}\right)}{\lambda \det(d)} + o\left(\frac{1}{\lambda}\right) = \frac{(d^{-1})_{i,j}}{\lambda} + o\left(\frac{1}{\lambda}\right) .$$

And finally, to complete the proof, we get $\left(A_\lambda^{-1}\right)_{1,1} = \frac{\lambda^p \det(d_\lambda)}{\det(A_\lambda)} \sim \lambda$, as $\lambda \to 0$. $\square$

Lemma 4. *If a collection of densities $\mathcal{P} = \{f_\theta, \theta \in \Theta\}$ has a Hellinger derivative $\dot{\xi}_{\theta_0}$ in θ_0, and if θ_0 is an interior point of Θ, then $\int \dot{\xi}_{\theta_0}(z)\xi_{\theta_0}(z)\,\mathrm{d}z = 0$.*

PROOF. Differentiating the relationship $\int \xi_\theta^2(z)\mathrm{d}z = 1$ yields the Lemma (see Pollard, 2001; Bickel *et al.*, 1998). $\square$

Theorem 5. *Let p_λ be the Poisson distribution Poi(λ). Suppose that $\{f_\theta, \theta \in \Theta\}$ is Hellinger differentiable at θ_0. Denote by $I(\lambda_0, \theta_0)$ the Fisher information of the pile-up model, and by $I^f(\theta_0)$ the Fisher information associated with the original model $\{f_\theta, \theta \in \Theta\}$. Assume that $I^f(\theta_0)$ is nonsingular. Then for every λ_0 small enough, $I(\lambda_0, \theta_0)$ is nonsingular and the Cramér-Rao bound CR given by (16) satisfies the following asymptotic equivalences, as $\lambda_0 \to 0$,*

$$CR_\lambda(\lambda_0, \theta_0) \sim \lambda_0, \quad and \quad CR_\theta(\lambda_0, \theta_0) \sim \frac{(I^f)^{-1}(\theta_0)}{\lambda_0} .$$

PROOF. The random variable X defined in (21) has a density with respect to $\overline{\mathcal{L}_+}$ given by

$$h_{\lambda,\theta}(x) = \begin{cases} f_\theta(x)(1 - \mathrm{e}^{-\lambda}) & \text{if } x \in \mathbb{R}_+ \\ \mathrm{e}^{-\lambda} & \text{if } x = \infty . \end{cases} \tag{22}$$

First we show that the Cramér-Rao bound associated with the model $\{h_{\lambda,\theta}, (\lambda, \theta) \in \Lambda \times \Theta\}$ is given by

$$CR^h(\lambda_0, \theta_0) = \begin{bmatrix} \mathrm{e}^{\lambda_0} - 1 & 0 \\ 0 & \frac{(I^f)^{-1}(\theta_0)}{1 - \mathrm{e}^{-\lambda_0}} \end{bmatrix} . \tag{23}$$

The family $\{h_{\lambda,\theta}, (\lambda, \theta) \in \Lambda \times \Theta\}$ is a pile-up model with N having Bernoulli distribution with parameter $(1 - \mathrm{e}^{\lambda_0})$. Set $\rho_{\lambda_0,\theta_0} := \sqrt{h_{\lambda_0,\theta_0}}$. Applying Theorem 2 gives the Hellinger differentiability at (λ_0, θ_0) with derivative

$$\partial_\lambda \rho_{\lambda_0,\theta_0}(x) = \begin{cases} \frac{-\xi_{\theta_0}(x)\mathrm{e}^{-\lambda_0}}{2\sqrt{1 - \mathrm{e}^{-\lambda_0}}} & \text{if } x \in \mathbb{R}_+ \\ -\frac{\mathrm{e}^{-\lambda_0/2}}{2} & \text{if } x = \infty , \end{cases}$$

$$\partial_\theta \rho_{\lambda_0,\theta_0}(x) = \begin{cases} \sqrt{1 - \mathrm{e}^{-\lambda_0}}\,\dot{\xi}_{\theta_0}(x) & \text{if } x \in \mathbb{R}_+ \\ 0 & \text{if } x = \infty . \end{cases}$$

Obviously, we obtain for the Fisher information $I^h(\lambda_0, \theta_0)$ of $\{h_{\lambda,\theta}, (\lambda, \theta) \in \Lambda \times \Theta\}$

$$I_\lambda^h(\lambda_0, \theta_0) = 4 \int (\partial_\lambda \rho_{\lambda_0,\theta_0})^2 \mathrm{d}\overline{\mathcal{L}_+} = \frac{\mathrm{e}^{-2\lambda_0}}{1 - \mathrm{e}^{-\lambda_0}} \int_0^\infty \xi_{\theta_0}^2(x)\mathrm{d}x + \mathrm{e}^{-\lambda_0} = \frac{1}{\mathrm{e}^{\lambda_0} - 1} , \tag{24}$$

and

$$I_\theta^h(\lambda_0, \theta_0) = 4(1 - \mathrm{e}^{-\lambda_0}) \int_0^\infty \dot{\xi}_{\theta_0}(x)\dot{\xi}_{\theta_0}^T(x)\mathrm{d}x = (1 - \mathrm{e}^{-\lambda_0})I^f(\theta_0) . \tag{25}$$

Finally, by applying Lemma 4, it follows that

$$I_{\lambda,\theta}^h(\lambda_0, \theta_0) = 4 \int_0^\infty \partial_\lambda \rho_{\lambda_0,\theta_0}(x)\partial_\theta \rho_{\lambda_0,\theta_0}(x)\mathrm{d}x = 2\mathrm{e}^{-\lambda_0} \int_0^\infty \xi_{\theta_0}(x)\dot{\xi}_{\theta_0}(x)\mathrm{d}x = 0 . \tag{26}$$

Inverting the Fisher information matrix

$$I^h(\lambda_0, \theta_0) = \begin{pmatrix} I_\theta^h(\lambda_0, \theta_0) & 0 \\ 0 & I_\lambda^h(\lambda_0, \theta_0) \end{pmatrix}$$

11

yields (23). Now we show that the following asymptotic equivalences hold, as $\lambda_0 \to 0$,

$$I_\lambda^h(\lambda_0, \theta_0) \sim \lambda_0^{-1} \quad \text{and} \quad I_\lambda^g(\lambda_0, \theta_0) - I_\lambda^h(\lambda_0, \theta_0) = O(\lambda_0), \tag{27}$$

$$I_\theta^h(\lambda_0, \theta_0) \sim \lambda_0 I^f(\theta_0) \quad \text{and} \quad I_\theta^g(\lambda_0, \theta_0) - I_\theta^h(\lambda_0, \theta_0) = O(\lambda_0^2), \tag{28}$$

$$I_{\lambda,\theta}^h(\lambda_0, \theta_0) = 0 \quad \text{and} \quad I_{\lambda,\theta}^g(\lambda_0, \theta_0) = O(\lambda_0) . \tag{29}$$

Combining (18) and (24) yields (27). To establish (28), applying Theorem 2 in the Poisson case (see (4) and (12)), and using the notation therein, we obtain, for all $z \in \mathbb{R}_+$,

$$\partial_\theta \zeta_{\lambda_0, \theta_0}(z) = \sqrt{\lambda_0} e^{-\lambda_0 F_{\theta_0}(z)/2} \left[\dot{\xi}_{\theta_0}(z) + \lambda_0 \xi_{\theta_0}(z) \int_z^\infty \dot{\xi}_{\theta_0}(t) \xi_{\theta_0}(t) \mathrm{d}t \right] .$$

By the Cauchy–Schwarz Inequality and since $\|\xi_{\theta_0}\|_2 = 1$, we have

$$\sup_{z \geq 0} \left| \int_z^\infty \dot{\xi}_{\theta_0}(t) \xi_{\theta_0}(t) \mathrm{d}t \right| \leq \|\dot{\xi}_{\theta_0}\|_2 .$$

Using $\sup_z |1 - e^{-\lambda_0 F_{\theta_0}(z)/2}| \leq |1 - e^{-\lambda_0/2}|$ and the two previous displays, we obtain, as $\lambda_0 \to 0$, $\left\| \partial_\theta \zeta_{\lambda_0, \theta_0} - \sqrt{\lambda_0} \dot{\xi}_{\theta_0} \right\|_2 = O(\lambda_0^{3/2})$. This combined with (25) yields (28). Further applying Theorem 2 and using similar arguments as above and Lemma 4 yields (29). Apply Lemma 3 and (24)–(26) complete the proof of the theorem. $\square$

The assertion of the theorem is that the information bound $CR_\theta(\lambda_0, \theta_0)$ for estimating the parameter of interest θ_0 increases as λ_0 decreases to 0, which indicates that a microscopic λ_0 results in a significant loss of information.

3.4. Exponential case

In view of the TCSPC application, we consider the special case where f_{θ_0} belongs to the family of exponential distributions $\{f_\theta(x) = \theta e^{-\theta x}, \theta > 0\}$ and p_λ is the Poisson distribution $\text{Poi}(\lambda)$. We will see that in this case it is relatively easy to determine the optimal value of λ_0 since $\{f_\theta, \theta \in \Theta\}$ is a scale family that is Hellinger differentiable at all $\theta > 0$. In view of Theorem 4, our study of the pile-up model can be restricted to the case of $\theta_0 = 1$. Furthermore, Theorem 5 implies that the Cramér-Rao bound becomes large when λ is small.

The following theorem describes the asymptotic behaviour of the Fisher information I_θ when λ tends to infinity. We state that the pile-up model with large λ contains (asymptotically) the same amount of information on the exponential parameter θ as a random variable with exponential distribution $\mathcal{E}(\theta)$.

Theorem 6. *Let $\{f_\theta(x) = \theta e^{-\theta x}, \theta > 0\}$ be the family of exponential densities and let p_λ be the Poisson distribution $\text{Poi}(\lambda)$. Denote the Fisher information of the parameter θ associated with the pile-up model by I_θ. Then, for all $\theta_0 > 0$,*

$$\lim_{\lambda_0 \to \infty} I_\theta(\lambda_0, \theta_0) = \frac{1}{\theta_0^2} . \tag{30}$$

Note that the following property holds for the functions $b_\lambda(m, n)$ defined by

$$b_\lambda(m, n) = \int_0^\infty z^m e^{-nz + \lambda e^{-z}} dz = \sum_{k=0}^\infty \frac{\lambda^k}{k!} \int_0^\infty z^m e^{-z(n+k)} dz = m! \sum_{k=0}^\infty \frac{\lambda^k}{k!(k+n)^{m+1}} .$$

Lemma 5. *For all $m \geq n \geq 1$, we have $b_\lambda(m, n) = o(\lambda^{-n} e^\lambda)$ as $\lambda \to \infty$.*

PROOF. The result follows by applying n times Hospital's rule to

$$\lim_{\lambda \to \infty} \frac{\lambda^n b_\lambda(m,n)}{e^\lambda} = \lim_{\lambda \to \infty} \frac{m! \sum_{k=0}^\infty \frac{\lambda^{k+n}}{k!(k+n)^{m+1}}}{e^\lambda} = \lim_{\lambda \to \infty} \frac{m! \sum_{k=0}^\infty \frac{\lambda^k(k+n)...(k+1)}{k!(k+n)^{m+1}}}{e^\lambda}$$

$$\leq m! \lim_{\lambda \to \infty} e^{-\lambda} \sum_{k=0}^\infty \frac{\lambda^k}{k!(k+n)^{m+1-n}} \leq m! \lim_{\lambda \to \infty} e^{-\lambda} \sum_{k=0}^\infty \frac{\lambda^k}{(k+1)!}$$

$$= m! \lim_{\lambda \to \infty} \frac{1 - e^{-\lambda}}{\lambda} = 0 \;.$$

$\square$

PROOF (THEOREM 6). Note that the Fisher information of θ has the following explicit form

$$I_\theta(\lambda_0, 1) = 4 \int (\partial_\theta \zeta_{\lambda_0,1}(z))^2 \mathrm{d}z = \lambda_0 e^{\lambda_0} \int_0^\infty e^{-z+\lambda_0 e^{-z}} \left(1 - z + \lambda_0 z e^{-z}\right)^2 dz$$

$$= 1 - e^{-\lambda_0} + \lambda_0 e^{-\lambda_0} \left\{ b_{\lambda_0}(2,1) + \lambda_0^2 b_{\lambda_0}(2,3) - 2 b_{\lambda_0}(1,1) - 2\lambda_0 b_{\lambda_0}(1,2) + 2\lambda_0 b_{\lambda_0}(2,2) \right\} \;.$$

Without loss of generality take $\theta_0 = 1$, by Theorem 4. By Lemma 5, the terms $e^{-\lambda_0}\lambda_0 b_{\lambda_0}(2,1)$, $e^{-\lambda_0}\lambda_0 b_{\lambda_0}(1,1)$ and $e^{-\lambda_0}\lambda_0^2 b_{\lambda_0}(2,2)$ tend to 0 as $\lambda_0 \to \infty$. By applying Hospital's rule to the first term three times and to the second term two times, we obtain

$$\lim_{\lambda_0 \to \infty} \left| e^{-\lambda_0}\lambda_0^3 b_{\lambda_0}(2,3) - 2e^{-\lambda_0}\lambda_0^2 b_{\lambda_0}(1,2) \right|$$

$$= 2 \left| \lim_{\lambda_0 \to \infty} e^{-\lambda_0} \sum_{k=0}^\infty \frac{\lambda_0^k(k+1)(k+2)}{k!(k+3)^2} - 2 \lim_{\lambda_0 \to \infty} e^{-\lambda_0} \sum_{k=0}^\infty \frac{\lambda_0^k(k+1)}{k!(k+2)} \right|$$

$$= 2 \lim_{\lambda_0 \to \infty} e^{-\lambda_0} \left| \sum_{k=0}^\infty \frac{\lambda_0^k}{k!} \frac{2k^2 + 7k + 5}{k^3 + 8k^2 + 21k + 18} \right|$$

$$\leq 28 \lim_{\lambda_0 \to \infty} e^{-\lambda_0} \left| \sum_{k=0}^\infty \frac{\lambda_0^k}{(k+1)!} \right|$$

$$= 0 \;.$$

Thus $e^{-\lambda_0}\lambda_0^3 b_{\lambda_0}(2,3) - 2e^{-\lambda_0}\lambda_0^2 b_{\lambda_0}(1,2)$ tends to 0 as $\lambda_0 \to \infty$. This implies that $I_\theta(\lambda_0, 1)$ tends to 1 as $\lambda_0 \to \infty$, and (20) gives (30). $\square$

From Figure 2 we see that the Cramér-Rao bound $\lambda \mapsto CR_\theta(\lambda, 1)$ is convex in λ. A numerical minimization yields a minimum at $\lambda_{opt} \approx 1.3275$ and Theorem 4 implies that λ_{opt} minimizes the Cramér-Rao bound $\lambda \mapsto CR_\theta(\lambda, \theta_0)$ for all $\theta_0 > 0$. The value of λ_{opt} is much larger than the intensity usually used in fluorescence measurements ($\lambda < 0.05$). We conclude that it is worthwhile to increase the intensity λ because this opens the way for estimators with smaller variance. In Section 4 we propose such an estimator that is adapted to the pile-up model.

3.5. Multi-exponential case

In experiments where there are more than one fluorescent molecule, a pile-up model with underlying exponential mixture distribution is considered. The density is given by

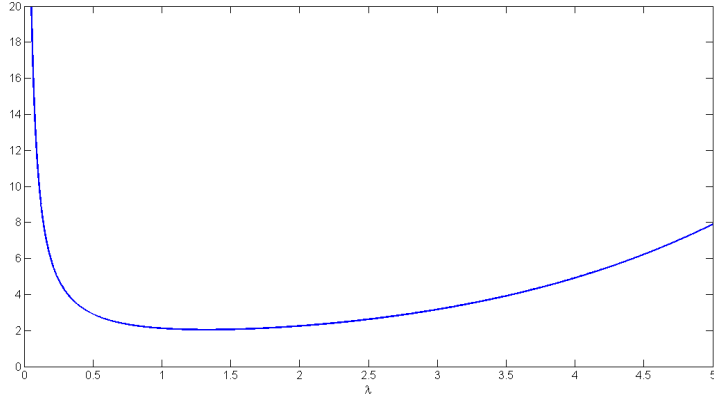$$f_\theta(y) = \sum_{k=1}^K \alpha_k \nu_k e^{\nu_k y}, \tag{31}$$

Figure 2: Cramér-Rao bound $\lambda \mapsto CR_\theta(\lambda, 1)$ of θ for the pile-up model with underlying exponential distribution in the Poisson case. The minimum is attained at $\lambda_{\text{opt}} \approx 1.3275$.
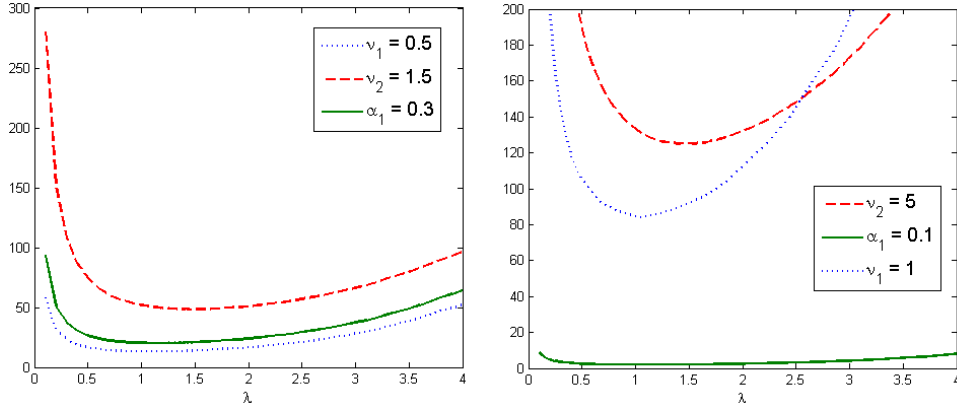

Figure 3: Cramér-Rao bounds $\lambda \mapsto CR(\lambda, \theta_0)$ for two-component models.

where $\nu_k > 0$ and $0 < \alpha_k < 1$. Note that α_K is given indirectly by $\alpha_K = 1 - \sum_{k=1}^{K-1} \alpha_k$. Denote by $\theta = (\alpha_1, \ldots, \alpha_{K-1}, \nu_1, \ldots, \nu_K)^T$ the parameter vector. We assume that the order of the mixture model K is fixed and known. The family of multi-exponential distributions $\{f_\theta, \theta \in \Theta\}$ is Hellinger differentiable on Θ. However, mixtures with two components or more ($K \geq 2$) do not form a scale family. Consequently, Theorem 4 concerning the existence of a global optimal value λ_{opt} that minimizes the Cramér-Rao bound, does not apply.

For a two-component mixture, Figure 3 shows the graphs of the Cramér-Rao bounds of the associated pile-up model for different choices of parameters θ_0 in the Poisson case. Although all functions are convex in λ, note that no unique optimal value λ_{opt} minimizing all Cramér-Rao bounds simultaneously exists. The minimum values depend on θ_0. Nevertheless, the graphs suggest choosing λ within the interval $[1, 2]$ for a two-component model. All graphs are obtained by inverting an approximation of the Fisher information matrix obtained by Monte Carlo simulation.

Remark 7. If we introduce a finite observation time $T < \infty$ as in Remark 1, then even in the case of scale families no global optimal value λ_{opt} minimizing the Cramér-Rao bound exists. Simulations in the Poisson case reveal that the optimal value depends on the underlying distribution F_{θ_0}, the larger the probability of censored observations, the larger the optimal value λ_{opt}. This phenomenon suggests that if the maximal observation time T is relatively short with respect to the tail of the distribution F_{θ_0} then it is preferable to increase the average number of photons hitting the detector such that the observed

minimum falls within the interval $(0, T)$ with high probability.

4. Gibbs Sampler for the pile-up model with an exponential mixture

The goal of this section is to construct a parameter estimate adapted to pile-up affected fluorescence measurements. Coates (1968) proposes a correction method for pile-up affected histograms, but no parameter estimates. We consider the multi-exponential case where N has Poisson distribution $\text{Poi}(\lambda)$. Note that exponential mixtures are of general interest for survival or duration setups. Rewriting the distribution of a pile-up observation Z defined in (1) in the multi-exponential case

$$G_\theta(z) = 1 - \sum_{n=0}^{\infty} p_\lambda(n) n! \sum_{(m_1,\ldots,m_K):\sum_j m_j = n} \frac{\prod_{j=1}^{K} \alpha_j^{m_j}}{\prod_{j=1}^{K} m_j!} e^{-z \sum_{j=1}^{K} m_j \nu_j} \ ,$$

shows that Z has an infinite exponential mixture distribution where the set of exponential parameters is given by $\nu_1 \mathbb{N} + \nu_2 \mathbb{N} + \ldots + \nu_K \mathbb{N}$. Existing methods for estimating the mixing distribution of a finite or infinite exponential mixture are the EM algorithm as proposed in Jewell (1982), the penalized dual method (Pilla *et al.*, 2006) or the rotated EM algorithm (Pilla & Lindsay, 2001) both searching for the nonparametric maximum likelihood estimator (NPMLE). Another method is the Markov Chain Monte-Carlo (MCMC) approach proposed in Gruet *et al.* (1999). In our context the EM algorithm is hard to handle since the maximization step does not provide explicit solutions for all parameters. The NPMLE methods seem to be inappropriate unless one takes into account the specific structure of the parameter set and the weights. The MCMC approach, in contrast, can be adapted to the pile-up density despite the more involved structure of the model by defining a new set of missing variables.

4.1. Algorithm

Here a Gibbs sampler is presented to estimate the parameters of the pile-up model. In view of Remark 1, we consider a model with a finite maximal observation time T. Then the pile-up observation Z admits a density with respect to the measure $\mathcal{L}_{(0,T)} \otimes \delta_T$ where $\mathcal{L}_{(0,T)}$ denotes the Lebesgue measure on $(0, T)$ and δ_T the measure having mass one in T given by

$$g_{\lambda,\theta}(z) = \begin{cases} \lambda f_\theta(z) e^{-\lambda F_\theta(z)} & \text{if } z \in (0, T) \\ e^{-\lambda F_\theta(T)} & \text{if } z = T \ . \end{cases}$$

Following the approach of Gruet *et al.* (1999) identifiability of the model is obtained by ordering the exponential parameters: $\nu_1 > \nu_2 > \ldots > \nu_K$. In order to use improper priors as in Marin *et al.* (2005), the exponential parameters are reparametrized as follows

$$\varphi = \nu_1, \quad \tau_k = \frac{\nu_k}{\nu_{k-1}}, \quad k = 2, \ldots, K \ .$$

As the weights verify the constraint $\sum_{k=1}^{K} \alpha_k = 1$, a further reparametrization is necessary to assure that at every iteration of the Gibbs sampler the new values of α_k are chosen in the entire set $\{(\alpha_1, \ldots, \alpha_K) \in [0,1]^K : \sum_{k=1}^{K} \alpha_k = 1\}$. We set

$$q_k = \frac{\alpha_k}{1 - \sum_{j=1}^{k-1} \alpha_j}, \quad k = 1, \ldots, K-1 \ .$$

Using the following noninformative prior leads to proper posterior distributions as shown in Gruet *et al.* (1999)

$$\pi(\lambda, q_1, \ldots, q_{K-1}, \varphi, \tau_2, \ldots, \tau_K) = \frac{1}{\lambda \varphi} \prod_{k=1}^{K-1} \mathbb{1}\{q_k \in [0,1]\} \prod_{k=2}^{K} \mathbb{1}\{\tau_k \in [0,1]\} \ .$$

15

The Gibbs sampling algorithm consists of iterating the steps (a) data augmentation of the missing data and (b) generation of new parameter values drawn from the posterior distributions. Denote $\mathbf{b} = (B_1, \ldots, B_K)$ the missing data associated with the pile-up observation Z, where B_k is the number of photons generated from the k-th mixture component. Let X_z be a random variable with Poisson distribution $\text{Poi}(\lambda(1 - F_\theta(z)))$. The missing data is generated by first drawing the total number of photons N from the distribution

$$N|Z \overset{\mathcal{D}}{=} \begin{cases} X_Z + 1 & \text{if } Z \in (0, T) \\ X_T & \text{if } Z = T , \end{cases}$$

then by drawing a vector $\mathbf{b} = (B_1, \ldots, B_K)$ from the distribution

$$P(\mathbf{b} = (b_1, \ldots, b_K)|N = n, Z)$$

$$= \begin{cases} \frac{(n-1)! \prod_k \alpha_k^{b_k} \sum_k b_k \nu_k}{f(Z)(1-F(Z))^{n-1} \prod_k b_k!} e^{-Z \sum_k b_k \nu_k} & \text{for } n \geq 1, \text{ if } Z \in (0, T) \\ \frac{n! \prod_k \alpha_k^{b_k}}{(1-F(T))^n \prod_k b_k!} e^{-T \sum_k b_k \nu_k} & \text{for } n \geq 0, \text{ if } Z = T , \end{cases}$$

where $b_k \in \mathbb{N}$ and $\sum_k b_k = n$. We denote by $\theta = (\lambda, \phi, \tau_2, \ldots, \tau_K, q_1, \ldots, q_{K-1})$ the parameter vector and by $\theta^{(-\lambda)}$ the vector θ where the element λ is missing. The notations $\theta^{(-\phi)}, \theta^{(-\tau_l)}$ and $\theta^{(-q)}$ are defined likewise. Then the posterior distributions for the different parameters given the data $\mathbf{Z} = (Z_1, \ldots, Z_M)$, the missing data $\mathbf{B} = (B_{ik}, i = 1, \ldots, M, \ k = 1, \ldots, K)$ and the other parameters turn out to be

$$\lambda|(\mathbf{Z}, \mathbf{B}, \theta^{(-\lambda)}) \sim \Gamma\left(\sum_{i=1}^{M} \sum_{k=1}^{K} B_{ik}, M\right)$$

$$\phi|(\mathbf{Z}, \mathbf{B}, \theta^{(-\phi)}) \sim \Gamma\left(\sum_{i=1}^{M} \mathbb{1}\{\sum_{k=1}^{K} B_{ik} > 0\}, \sum_{k=1}^{K} \tau_1 \ldots \tau_k \sum_{i=1}^{M} B_{ik} Z_i\}\right)$$

$$q_l|(\mathbf{Z}, \mathbf{B}, \theta^{(-q_l)}) \sim \text{Beta}\left(\sum_{i=1}^{M} B_{il} + 1, \sum_{i=1}^{M} \sum_{k=l+1}^{K} B_{ik} + 1\right) ,$$

where $\tau_1 = 1$ and $q_K = 1$. Furthermore, the posterior distribution of τ_l is a mixture of Gamma distributions restricted to the interval $[0, 1]$ with $\sum_{i=1}^{M} \mathbb{1}\{\sum_{k=1}^{K} B_{ik} > 0\}$ components given by

$$p(\tau_l|\mathbf{Z}, \mathbf{B}, \theta^{(-\tau_l)}) \propto \exp\left\{-\phi \sum_{k=1}^{K} \tau_1 \ldots \tau_k \sum_{i=1}^{M} B_{ik} Z_i\right\} \prod_{i: \sum_{k=1}^{K} B_{ik} > 0} \left(\phi \sum_{k=1}^{K} \tau_1 \ldots \tau_k B_{ik}\right) \mathbb{1}_{[0,1]}(\tau_l) .$$

Parameter estimators are obtained by taking the mean of the generated values. The algorithm is robust to different starting values. We recommend to use uniformly distributed values for the weights α_k and any distinct values for the exponential parameters ν_k. For the intensity λ the initial value

$$-\log\left(\frac{M}{\sum_{i=1}^{M} \mathbb{1}\{Z_i = T\}}\right) ,$$

may be used since $p_\lambda(0) = e^{-\lambda}$.

4.2. Experimental Results

Simulations are carried out to evaluate the performance of the new Gibbs sampler. Table 1 presents results obtained for data simulated from models with different numbers of components K. Note that all estimated values are close to the true parameter values and, of course, by increasing the number of observations one may obtain even higher precision. Especially in the exponential case ($K = 1$) very

good estimates are provided for a quite high intensity ($\lambda = 3$). For illustration the quantile-quantile plot for the two-component model is presented in Figure 6. It shows that the estimated distribution is close to the true one, which is also confirmed by a high p-value of the Kolmogorov-Smirnov test (0.984).

Table 1: Simulation results for $K = 1, 2, 3$ components and 3,000, 3,000, 10,000 observations, resp.

	$K = 1$		$K = 2$			
	λ	ν	λ	ν_1	ν_2	α_1
true values	3	0.1	1	6	0.5	.75
estimated values	3.05	.101	.994	6.03	.518	.742
			$K = 3$			
	λ	ν_1	ν_2	ν_3	α_1	α_2
true values	1	5	1	0.1	.33	.33
estimated values	1.01	4.88	.970	.101	.347	.329

Let us compare the covariance matrix of the Gibbs sampler with the corresponding Cramér-Rao bound matrix in a specific case. We consider the two-component model with parameters $\lambda = 1.5, \nu_1 = 2, \nu_2 = 0.5, \alpha_1 = 0.3$ and 5,000 observations. From 50 simulations we obtain the following empirical covariance matrix for the estimates of $(\lambda, \nu_1, \nu_2, \alpha_1)$

$$10^{-3} \begin{pmatrix} 0.614 & -1.248 & 0.025 & -0.027 \\ -1.248 & 61.604 & 2.767 & -10.782 \\ 0.025 & 2.767 & 0.440 & -0.698 \\ -0.027 & -10.782 & -0.698 & 2.433 \end{pmatrix},$$

while the Cramér-Rao bound obtained by the inverse of a Monte-Carlo approximation of the Fisher information matrix is

$$10^{-3} \begin{pmatrix} 0.696 & 0.057 & -0.034 & -0.156 \\ 0.057 & 58.052 & 3.990 & -10.053 \\ -0.034 & 3.990 & 0.610 & -0.933 \\ -0.156 & -10.053 & -0.933 & 2.181 \end{pmatrix}.$$

We conclude that the Gibbs sampler is well adapted to pile-up affected data, and as it attains the Cramér-Rao bound it might lead to a significant reduction of the acquisition time. Therefore, we compare the MCMC method to the following estimation practice. Data from the pile-up model are obtained at a low laser intensity ($\lambda = 0.05$) such that the probability for 2 or more photons per laser pulse is negligible. Then the observed arrival times are considered as independent observations from the exponential mixture distribution given by (31) and the classical EM algorithm is applied. Repeated simulations provide estimates of the bias and the variance of the estimators for a two-component model and various numbers of observations. For the same two-component model we simulated data using the laser intensity $\lambda_{\text{opt}} = 1.32$ and applied the MCMC method. From the results shown in Table 2 we see that any bias and any variance of the standard method exceeds the corresponding value of the MCMC method. Actually, comparing the results for 10,000 observations of the first method to the results for 1,000 observations of the new method shows that all values from the MCMC method are smaller than the first. Thus, we obtain estimates of comparable statistical quality by using ten times less, but piled up observations instead of data without pile-up. Hence, there is a significant reduction of the acquisition time. For further simulation results we refer to Rebafka (2007).

Table 2: Comparison of the standard estimation method and the MCMC method for a two-component model with true parameter values $\alpha_1 = .33, \alpha_2 = .66, \nu_1 = .5, \nu_2 = 6$.

Standard estimation method ($\lambda = 0.05$)				
nb of obs		1000	3000	5000	10000
α_1	bias	0.0051	0.0040	0.0036	0.0028
	var	0.0112	0.0035	0.0021	0.0010
ν_1	bias	0.0488	0.0143	0.0098	0.0056
	var	0.0447	0.0092	0.0054	0.0025
ν_2	bias	0.6523	0.2198	0.1626	0.1080
	var	6.6068	0.9788	0.5520	0.2606

MCMC method ($\lambda = 1.32$)				
nb of obs		1000	3000	5000	10000
α_1	bias	0.0006	0.0001	0.0001	$.492e^{-4}$
	var	0.0008	0.0003	0.0002	0.0001
ν_1	bias	0.0049	0.0018	0.0010	0.0006
	var	0.0023	0.0008	0.0005	0.0002
ν_2	bias	0.0278	0.0069	0.0057	0.0023
	var	0.1499	0.0491	0.0294	0.0147

References

BICKEL, P., KLAASSEN, C., RITOV, Y. & WELLNER, J. (1998). *Efficient and Adaptive Estimation for Semiparametric Models*. Springer Verlag, New York.

BICKEL, P. J. & DOKSUM, K. A. (2001). *Mathematical Statistics, vol. I*. Prentice Hall.

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.

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.

DEVROYE, L. (1986). *Non-Uniform Random Variate Generation*. Springer-Verlag.

GRUET, M.-A., PHILIPPE, A. & ROBERT, C. P. (1999). MCMC control spreadsheets for exponential mixture estimation. *Journal of Computational and Graphical Statistics* **8**(2), 298–317.

JEWELL, N. P. (1982). Mixtures of exponential distributions. *Annals of Statistics* **10**(2), 479–484.

LAKOWICZ, J. R. (1999). *Principles of Fluorescence Spectroscopy*. Academic/Plenum, New York.

MARIN, J.-M., MENGERSEN, K. & ROBERT, C. P. (2005). Bayesian modelling and inference on mixtures of distributions. In: *Handbook of Statistics* (DEY, D. & (EDS), C. R., eds.), vol. 25, chap. 16. Elsevier-Sciences.

O'CONNOR, D. V. & PHILLIPS, D. (1984). *Time Correlated Single Photon Counting*. Academic Press, London.

PILLA, R. S., BARTOLUCCI, F. & LINDSAY, B. G. (2006). Model building for semiparametric mixtures. Available at http://arxiv.org/abs/math/0606077v1.

PILLA, R. S. & LINDSAY, B. G. (2001). Alternative EM methods for nonparametric finite mixture models. *Biometrika* **8**(2), 535–550.

POLLARD, D. (2001). Hellinger differentiability. Available at `http://www.stat.yale.edu/~pollard/Paris2001/Lectures/DQM.pdf`.

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.

VALEUR, B. (2002). *Molecular Fluorescence*. WILEY-VCH, Weinheim.

WALKER, J. G. (2002). Iterative correction for 'pile-up' in single-photon lifetime measurements. *Optics Communications* **201**, 271–277.