



## Review

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

Tabea Rebařka<sup>a,b,\*</sup>, François Roueff<sup>a</sup>, Antoine Souloumiac<sup>b</sup><sup>a</sup> Télécom Paristech, Département TSI/CNRS LTCI, 46 rue Barrault, 75634 Paris, France<sup>b</sup> CEA, LIST, Laboratoire Processus Stochastiques et Spectres, 91191 Gif-sur-Yvette, France

## ARTICLE INFO

## Article history:

Received 24 June 2008

Received in revised form

25 May 2010

Accepted 31 May 2010

Available online 8 June 2010

## Keywords:

Cramér–Rao bound

Exponential mixture

Fluorescence lifetime measurements

Gibbs sampler

Hellinger differentiability

TCSPC

## 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.

© 2010 Elsevier B.V. All rights reserved.

## Contents

|  |    |
|--|----|
| 1. Introduction . . . . .  | 2  |
| 2. General setting . . . . .   | 2  |
| 2.1. The pile-up model . . . . .   | 2  |
| 2.2. Hellinger differentiability . . . . .                                   | 3  |
| 3. The effect of pile-up on information bounds . . . . .                     | 6  |
| 3.1. Fisher information of $\lambda$ . . . . .                               | 6  |
| 3.2. Scale family . . . . .  | 7  |
| 3.3. Asymptotic behaviour in the Poisson case . . . . .                      | 8  |
| 3.4. Exponential case . . . . .  | 10 |
| 3.5. Multi-exponential case . . . . .  | 11 |
| 4. Gibbs sampler for the pile-up model with an exponential mixture . . . . . | 12 |
| 4.1. Algorithm . . . . .   | 13 |
| 4.2. Experimental results . . . . .  | 14 |
| Acknowledgment . . . . .   | 16 |
| References . . . . .   | 16 |

\* Corresponding author at: Télécom Paristech, Département TSI/CNRS LTCI, 46 rue Barrault, 75634 Paris, France. Tel.: +33 145817959; fax: +33 169087819.

E-mail address: [tabea.rebafka@telecom-paristech.fr](mailto:tabea.rebafka@telecom-paristech.fr) (T. Rebařka).

## 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 and 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 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 modelled by a Poisson distribution with parameter  $\lambda$ . The literature advises a low laser intensity such that the average number of photons  $\lambda$  hitting the detector is  $< 0.05$ , see Valeur (2002). However, a small  $\lambda$  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  $\lambda$  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

$$Z = \begin{cases} \min\{Y_1, \dots, Y_N\} & \text{if } N > 0, \\ \infty & \text{if } N = 0. \end{cases} \quad (1)$$

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

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

where  $M$  is the moment generating function associated with  $N$ , defined as  $M(u) = E(u^N)$  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  $\dot{M}(u) = E[Nu^{N-1}]$  for all  $[0, 1]$ , the density  $g$  is given by

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

**Proof.** Relation (2) is obtained by computing  $P(Z > z) = E[P(Z > z|N)]$ .

Relation (3) follows immediately, if for instance  $f$  is continuous. 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,  $\int |f_n - f| d\mathcal{L}_+ \rightarrow 0$  as  $n \rightarrow \infty$ . 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,  $\sup_y |F_n(y) - F(y)| \rightarrow 0$  as  $n \rightarrow \infty$  and, using the properties of  $M$  and  $\dot{M}$ ,

that  $\sup_z |G_n(z) - G(z)| \rightarrow 0$  and  $\int_z^\infty |g_n - g| d\mathcal{L}_+ \rightarrow 0$  as  $n \rightarrow \infty$  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  $\lambda$ ,  $p_\lambda(k) = P(N = k)$  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  $\mathcal{L}_+$ , 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  $\infty$ , 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  $f_\theta(t) = \theta f_1(t/\theta)$  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 and 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  $\nu$  and parameter set  $\Theta \in \mathbb{R}^d$  is said to be Hellinger differentiable with respect to  $\nu$  at a point  $\theta_0$ , if the map  $\theta \mapsto \xi_\theta(x) := \sqrt{f_\theta(x)}$  is differentiable in  $L^2(\nu)$ -norm at  $\theta_0$ . 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  $r_\theta$  satisfies  $\|r_\theta\|_2 = o(|\theta - \theta_0|)$  as  $\theta \rightarrow \theta_0$ .

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  $\theta_0$  of  $\Theta$  with derivative  $\dot{\xi}_{\theta_0}$ . Suppose that an estimator  $T(X)$  satisfies

$$\sup_{\theta \in U_\theta} E_\theta \{T(X)^2\} < \infty,$$

for some neighbourhood  $U_\theta$  of  $\theta_0$ . Then  $\theta \mapsto \gamma_\theta := E_\theta\{T(X)\}$  has a derivative at  $\theta_0$  given by

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

Define the Fisher information matrix by  $I_\theta = 4 \int \dot{\zeta}_\theta \dot{\zeta}_\theta^T dv$ . If, in addition,  $I_{\theta_0}$  is nonsingular, then

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

We refer to Pollard (2001, Chapter 3.5) or Bickel et al. (1998, Chapter 2.3) 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  $\theta_0$  with Hellinger derivative  $\dot{\zeta}_{\theta_0}$ ,
- (ii)  $E_{\lambda_0}(N^3) = \sum_{k \geq 0} k^3 p_{\lambda_0}(k) < \infty$ ,
- (iii)  $p_{\lambda_0}(1) > 0$ ,
- (iv) the function  $\lambda \mapsto \{p_\lambda(k)\}_{k \geq 1}$  is differentiable at  $\lambda_0$  as a function defined on  $\Lambda$  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  $\lambda_0$  or (b)  $p_\lambda(0) = 0$  for all  $\lambda$  in a neighbourhood of  $\lambda_0$ ,

then  $\{g_{\lambda,\theta}, (\lambda, \theta) \in \Lambda \times \Theta\}$  is Hellinger differentiable with respect to  $\overline{\mathcal{L}}_+$  at  $(\lambda_0, \theta_0)$  with Hellinger derivative

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

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

$$\partial_\theta \zeta_{\lambda_0, \theta_0}(z) = \begin{cases} \dot{\zeta}_{\theta_0}(z) \dot{M}_{\lambda_0}^{1/2}(1-F_{\theta_0}(z)) + \frac{\zeta_{\theta_0}(z) \ddot{M}_{\lambda_0}(1-F_{\theta_0}(z))}{\dot{M}_{\lambda_0}^{1/2}(1-F_{\theta_0}(z))} \int_{z_0}^{\infty} \dot{\zeta}_{\theta_0}(t) \dot{\zeta}_{\theta_0}(t) dt & \text{if } z \in \mathbb{R}_+, \\ 0 & \text{if } z = \infty, \end{cases} \quad (8)$$

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

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

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

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

**Remark 3.** The pile-up model  $\{g_{\lambda,\theta}, \theta \in \Theta\}$  parameterized by  $\theta$  with a given  $\lambda_0$  is Hellinger differentiable at a point  $\theta_0$  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.

**Remark 4.** Condition (iv) implies that, for all  $k \geq 1$ ,  $\lambda \mapsto p_\lambda(k)$  is differentiable at  $\lambda_0$ , and that the sequence made of their derivatives  $\{\check{p}_{\lambda_0}(k)\}_{k \geq 1}$  belongs to  $\ell_1$ . Hence  $\check{M}_{\lambda_0}(u)$  is well defined in (10) and is equal to the derivative of  $\lambda \mapsto \dot{M}_\lambda(u)$  at  $\lambda_0$ . Conversely, if  $\lambda \mapsto p_\lambda(k)$  is differentiable in a neighbourhood  $U_0$  of  $\lambda_0$  for all  $k \geq 1$  and there exist  $k_0 > 0, (h_k)_{k \geq k_0}$  and  $v_0 : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  such that, for all  $k \geq k_0$  and  $\lambda \in U_0$ ,

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

with  $v_0(r) \rightarrow 0$  as  $r \downarrow 0$  and  $\sum_{k > 0} k h_k < \infty$ , 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,  $p_\lambda(k) = e^{-\lambda} \lambda^k / k!$ , 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  $(\lambda_0, \theta_0)$  follows from that of the model  $\{f_\theta, \theta \in \Theta\}$  at  $\theta_0$ . Moreover, we have

$$\check{M}_\lambda(u) = e^{\lambda(u-1)} \{1 + \lambda(u-1)\}. \quad (12)$$

**Proof of 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  $\dot{\zeta}_{\lambda,\theta}(\infty)$  does not depend on  $\theta$  and in (9) as an application of Condition (v). Let us first show that the model  $\{g_{\lambda,\theta}, \theta \in \Theta\}$  is Hellinger differentiable at  $\theta_0$  for some given  $\lambda_0$ . 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  $G : L^2(\mathcal{L}_+) \rightarrow L^\infty(\mathcal{L}_+)$  defined by

$$G(\xi)(z) = \int_z^\infty \xi^2(t) dt,$$

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  $\theta_0$  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  $\dot{G}_\xi : L^2(\mathcal{L}_+) \rightarrow L^\infty(\mathcal{L}_+)$  given by

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

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  $\check{M}_{\lambda_0}$  and second derivative bounded by  $E_{\lambda_0}\{N(N-1)(N-2)\}$ . Hence, for all  $t, s \in [-1, 1]$ ,

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

where  $|r(x)| \leq E_{\lambda_0}\{N(N-1)(N-2)\}x^2/2$ . Applying (13), we obtain that  $H$  is differentiable at any  $\mathbf{u}$  in the unit ball of  $L^\infty(\mathcal{L}_+)$  with derivative  $\mathbf{v} \mapsto \mathbf{v} \times \dot{M}_{\lambda_0} \circ \mathbf{u}$ . 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  $\mathbf{v} \mapsto (\mathbf{v}/2)\dot{M}_{\lambda_0} \dot{M}_{\lambda_0}^{-1/2} \circ \mathbf{u}$ .

Let us now prove the Hellinger differentiability of  $\zeta_{\lambda, \theta}$  at  $(\lambda_0, \theta_0)$ . 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

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

where the  $o$ -term holds as  $\lambda \rightarrow \lambda_0$  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  $\lambda \rightarrow \lambda_0$ ,

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

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

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

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  $\theta_0$  as a map valued in  $L^2(\mathcal{L}_+)$  and  $\theta \mapsto 1 - F_\theta$  is continuous at  $\theta_0$  as a map valued in  $L^\infty(\mathcal{L}_+)$ . Hence, as  $\theta \rightarrow \theta_0$ ,

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

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

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

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  $\lambda$  on the information contained in the data, since in the context of fluorescence measurements  $\lambda$  presents the average number of photons hitting the detector and can be coarsely tuned in the experimental setup. At a low intensity  $\lambda$  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  $\theta_0$  is quite poor. The question of finding an optimal value for  $\lambda$  arises.

The main result of this section is Theorem 4: If the underlying density  $f_{\theta_0}$  belongs to a scale family, then the Cramér–Rao bound of the scale parameter  $\theta_0$  achieves a minimum at a point  $\lambda_{\text{opt}}$  which is the same for all  $\theta_0 > 0$ . Evidently, this is of great practical use for fluorescence measurements when  $F_{\theta_0}$  is considered to be the exponential distribution. Theorem 4 suggests tuning the parameter as close to  $\lambda_{\text{opt}}$  as possible independently of the unknown parameter  $\theta_0$ , to minimize the asymptotic variance of the maximum likelihood estimator of  $\theta_0$ . In Subsection 3.4 the optimal intensity is determined at  $\lambda_{\text{opt}} \approx 1.32$  using numerical evaluations of the information bounds. Carrying out the experiment at the intensity  $\lambda_{\text{opt}}$ , 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  $\theta_0$ . 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  $(\lambda_0, \theta_0)$  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  $\Theta$  and  $\Lambda$  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  $(\lambda_0, \theta_0)$ , with derivative  $\dot{\zeta}_{\lambda_0, \theta_0}$  given by (7)–(9). We denote the corresponding Fisher information matrix of the pile-up model by

$$I(\lambda_0, \theta_0) = 4 \int \dot{\zeta}_{\lambda_0, \theta_0} \dot{\zeta}_{\lambda_0, \theta_0}^T d\overline{\mathcal{L}}_+ = \begin{bmatrix} I_{\lambda}(\lambda_0, \theta_0) & I_{\lambda, \theta}(\lambda_0, \theta_0) \\ I_{\lambda, \theta}^T(\lambda_0, \theta_0) & I_{\theta}(\lambda_0, \theta_0) \end{bmatrix}, \quad (15)$$

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

$$CR(\lambda_0, \theta_0) = I^{-1}(\lambda_0, \theta_0) = \begin{bmatrix} CR_{\lambda}(\lambda_0, \theta_0) & CR_{\lambda, \theta}(\lambda_0, \theta_0) \\ CR_{\lambda, \theta}^T(\lambda_0, \theta_0) & CR_{\theta}(\lambda_0, \theta_0) \end{bmatrix}, \quad (16)$$

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 $\lambda$

In the pile-up model  $\{g_{\lambda, \theta}, (\lambda, \theta) \in \Lambda \times \Theta\}$  the Fisher information of  $\lambda$  is independent of the underlying distribution  $F_{\theta_0}$ , 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  $\theta_0$  and  $\{f_{\theta}, \theta \in \Theta\}$ . More precisely,

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

**Proof.** Using (9), we obtain

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

Let  $X$  be a random variable with density  $\zeta_{\theta_0}^2$ . As its distribution function  $F_{\theta_0}$  is continuous,  $F_{\theta_0}(X)$  has uniform distribution on  $[0,1]$  (see Devroye, 1986, Theorem 2.1). Consequently,

$$\int_0^{\infty} \frac{\zeta_{\theta_0}^2(z) \check{M}_{\lambda_0}^2(1-F_{\theta_0}(z))}{\check{M}_{\lambda_0}(1-F_{\theta_0}(z))} dz = E \left\{ \frac{\check{M}_{\lambda_0}^2(1-F_{\theta_0}(X))}{\check{M}_{\lambda_0}(1-F_{\theta_0}(X))} \right\} = \int_0^1 \frac{\check{M}_{\lambda_0}^2(u)}{\check{M}_{\lambda_0}(u)} du.$$

This completes the proof.  $\square$

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

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

The independence of  $I_\lambda$  of the underlying distribution  $F_{\theta_0}$  may appear surprising, but it indicates that the main part of information about  $\lambda$  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  $F_{\theta_0}$ . A comparison of the information  $I_\lambda$  in the Poisson case to the Fisher information of a Bernoulli model with parameter  $p_\lambda(0) = e^{-\lambda}$ , equal to  $\check{I}_\lambda = (e^\lambda - 1)^{-1}$ , indicates that for values of  $\lambda$  close to 0 the two information values almost coincide, see Fig. 1. Furthermore, observe that the Fisher information  $I_\lambda$  of the pile-up model slightly exceeds the information  $\check{I}_\lambda$  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  $f_{\theta_0}$  belongs to a scale family  $\{\theta f_1(\cdot/\theta), \theta > 0\}$  with  $d=1$ . We show that the Cramér–Rao bound  $\lambda \mapsto CR_\theta(\lambda, \theta_0)$  achieves a minimum at a point  $\lambda_{opt}$  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  $\check{\xi}_{\theta_0}(x) = 1/\sqrt{\theta_0} \check{\xi}_1(\theta_0 x)$ .

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

$$r_\theta(x) = \check{\xi}_\theta(x) - \check{\xi}_1(x) - (\theta - 1)\check{\xi}_1(x),$$

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

$$\left\| \check{\xi}_\theta(x) - \check{\xi}_{\theta_0}(x) - \frac{(\theta - \theta_0)}{\sqrt{\theta_0}} \check{\xi}_1(\theta_0 x) \right\|_2^2 = \int_0^\infty \left( \check{\xi}_{\theta/\theta_0}(u) - \check{\xi}_1(u) - \left( \frac{\theta}{\theta_0} - 1 \right) \check{\xi}_1(u) \right)^2 du = \|r_{\theta/\theta_0}\|_2^2.$$

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

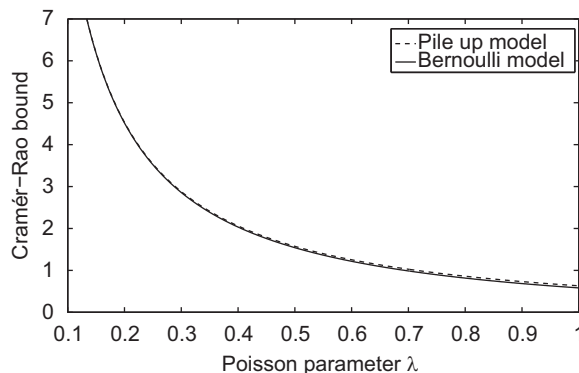
**Theorem 4.** Let  $\{f_\theta, \theta > 0\}$  be a scale family. If Assumption 2 holds for some  $(\lambda_0, \theta_0)$  and the Fisher information matrix  $I(\lambda_0, \theta_0)$  is nonsingular, then Assumption 2 and the invertibility of  $I(\lambda_0, \theta_0)$  hold for any positive value of  $\theta_0$  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,  $\operatorname{argmin}_{\lambda_0} CR_\theta(\lambda_0, \theta_0) = \operatorname{argmin}_{\lambda_0} CR_\theta(\lambda_0, 1)$ , and thus the value of  $\lambda_0$  minimizing the Cramér–Rao bound  $CR_\theta(\lambda_0, \theta_0)$  does not depend on  $\theta_0$ .

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

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



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

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  $F_{\theta_0}(z) = F_1(\theta_0 z)$ ,  $\zeta_{\theta_0}(x) = \sqrt{\theta_0} \zeta_1(\theta_0 x)$  and

$$\int_z^{\infty} \zeta_1(\theta_0 t) \dot{\zeta}_1(\theta_0 t) dt = \frac{1}{\theta_0} \int_{\theta_0 z}^{\infty} \zeta_1(t) \dot{\zeta}_1(t) dt, \quad z \in \mathbb{R}_+,$$

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,

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

and, likewise,  $I_{\lambda, \theta}(\lambda_0, \theta_0) = 1/\theta_0 I_{\lambda, \theta}(\lambda_0, 1)$ . 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  $2 \times 2$  Fisher information matrix

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

### 3.3. Asymptotic behaviour in the Poisson case

Our main interest is the question whether the currently used low intensity  $\lambda_0$  is optimal. Recall that in the fluorescence context  $N$  follows a Poisson distribution  $\text{Poi}(\lambda_0)$  and literature advises to use  $\lambda_0 < 0.05$ . In this section we investigate whether, from an information point of view, it is legitimate to collect fluorescence lifetime data at a very low intensity  $\lambda_0$ .

We provide a study of the behaviour of the Cramér–Rao bound in the Poisson case when  $\lambda_0 \rightarrow 0$  based on an approximation by a simpler model. Note that if  $\lambda_0$  is sufficiently small, the probability of detecting two or more photons becomes negligible and the Poisson distribution is close to a Bernoulli distribution with parameter  $1 - e^{-\lambda_0}$ . Hence, for small  $\lambda_0$  the pile-up model in the Poisson case can be approximated by the pile-up model with underlying Bernoulli distribution. To make this approximation precise we define the random variable  $X$  by

$$X = \begin{cases} Y & \text{with probability } 1 - e^{-\lambda_0}, \\ \infty & \text{with probability } e^{-\lambda_0}, \end{cases} \quad (21)$$

where  $Y$  has density  $f_{\theta_0}$ . The following proposition provides the asymptotic behaviour of the pile-up model in the Bernoulli case where the observations follow the same distribution as  $X$  given in (21). Thereafter we show that these results extend to the Poisson case. The advantage of the Bernoulli case relies on the fact that the corresponding information matrix is easier to calculate and to inverse than in the Poisson case.

**Proposition 1.** Denote the density of  $X$  defined in (21) by  $h_{\lambda_0, \theta_0}$  with  $(\lambda_0, \theta_0) \in \mathcal{A} \times \Theta$ . Suppose that  $\{f_{\theta}, \theta \in \Theta\}$  is Hellinger differentiable at  $\theta_0$ . If the Fisher information  $\mathcal{I}(\theta_0)$  associated with  $\{f_{\theta}, \theta \in \Theta\}$  is nonsingular, then the Cramér–Rao bound associated with the model  $\{h_{\lambda, \theta}, (\lambda, \theta) \in \mathcal{A} \times \Theta\}$  is given by

$$CR^h(\lambda_0, \theta_0) = \begin{bmatrix} CR_{\lambda}^h(\lambda_0, \theta_0) & 0 \\ 0 & CR_{\theta}^h(\lambda_0, \theta_0) \end{bmatrix} = \begin{bmatrix} e^{\lambda_0} - 1 & 0 \\ 0 & \frac{(\mathcal{I}^f)^{-1}(\theta_0)}{1 - e^{-\lambda_0}} \end{bmatrix}. \quad (22)$$

Consequently,  $CR_{\theta}^h(\lambda_0, \theta_0) \sim (\mathcal{I}^f)^{-1}(\theta_0) \lambda_0^{-1}$  as  $\lambda_0 \rightarrow 0$ .

In the proof of the proposition the following result (see, e.g. Pollard, 2001, Chapter 3.5 or Bickel et al., 1998, Chapter 2.1) is applied.

**Lemma 3.** If a collection of densities  $\mathcal{P} = \{f_{\theta}, \theta \in \Theta\}$  has a Hellinger derivative  $\dot{\zeta}_{\theta_0}$  in  $\theta_0$ , and if  $\theta_0$  is an interior point of  $\Theta$ , then  $\int \dot{\zeta}_{\theta_0}(z) \zeta_{\theta_0}(z) dz = 0$ .

**Proof of Proposition 1.** The density with respect to  $\overline{\mathcal{L}}_+$  of the random variable  $X$  defined in (21) is given by

$$h_{\lambda_0, \theta_0}(x) = \begin{cases} f_{\theta_0}(x)(1 - e^{-\lambda_0}) & \text{if } x \in \mathbb{R}_+, \\ e^{-\lambda_0} & \text{if } x = \infty. \end{cases}$$



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

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

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

We now show that the Fisher information matrix  $I^h(\lambda_0,\theta_0)$  associated with the model  $\{h_{\lambda,\theta}(\lambda,\theta) \in \mathcal{A} \times \Theta\}$  is given by

$$I^h(\lambda_0,\theta_0) = \begin{pmatrix} I_\lambda^h(\lambda_0,\theta_0) & I_{\lambda,\theta}^h(\lambda_0,\theta_0) \\ I_{\lambda,\theta}^h(\lambda_0,\theta_0) & I_\theta^h(\lambda_0,\theta_0) \end{pmatrix} = \begin{pmatrix} \frac{1}{e^{\lambda_0}-1} & 0 \\ 0 & (1-e^{-\lambda_0})I^f(\theta_0) \end{pmatrix}. \tag{23}$$

Indeed, we have

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

and

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

Furthermore, by applying Lemma 3, it follows that

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

This yields (23). Finally, (22) is obtained by  $CR^h(\lambda_0,\theta_0) = (I^h(\lambda_0,\theta_0))^{-1}$ .  $\square$

To extend the results of Proposition 1 to the Poisson case, the following lemma is used.

**Lemma 4.** *Let, for all  $\lambda > 0$ ,*

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

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

$$b_\lambda = O(1), \lambda a_\lambda \rightarrow 1 \quad \text{and} \quad d_\lambda \rightarrow d \quad \text{when } \lambda \rightarrow 0, \tag{24}$$

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

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

Further, denote the two diagonal blocks of the inverse of  $A_\lambda$  by  $m_\lambda$  and  $M_\lambda$ , where  $m_\lambda$  is a scalar and  $M_\lambda$  a  $p \times p$  matrix,

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

Then, when  $\lambda \rightarrow 0$ ,

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

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

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

Further,

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

where the  $(p-1) \times (p-1)$  matrix  $d_\lambda^{-(k,j)}$  results from  $d_\lambda$  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  $\lambda \rightarrow 0$ . 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,

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

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

We finally provide the asymptotic behaviour of the Cramér–Rao bound associated with the pile-up model in the Poisson case. We see that the Cramér–Rao bound of  $\theta_0$  tends to infinity as  $\lambda_0 \rightarrow 0$ . Hence, from an information viewpoint it is unfavourable to carry out the fluorescence experiment when  $\lambda_0$  is too small.

**Theorem 5.** Let  $p_{\lambda}$  be the Poisson distribution  $\text{Poi}(\lambda)$ . Suppose that  $\{f_{\theta}, \theta \in \Theta\}$  is Hellinger differentiable at  $\theta_0$ . 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  $\lambda_0$  small enough,  $I(\lambda_0, \theta_0)$  is nonsingular and the Cramér–Rao bound CR given by (16) satisfies the following asymptotic equivalences, as  $\lambda_0 \rightarrow 0$ ,

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

**Proof.** For clarity of notation, let  $I^{\mathcal{P}}$  be the Fisher information associated with the Poisson case and  $I^{\mathcal{B}}$  the Fisher information in the Bernoulli case given in (23). To show the theorem, we apply Lemma 4 with  $A_{\lambda_0} = I^{\mathcal{P}}(\theta_0, \lambda_0)$ . To verify the condition (24) of the lemma, we use the approximation of the pile-up model in the Poisson case by the pile-up model in the Bernoulli case. On the one hand, we clearly have by (23), as  $\lambda_0 \rightarrow 0$ ,

$$I_{\lambda}^{\mathcal{P}}(\lambda_0, \theta_0) \sim \lambda_0^{-1} \quad \text{and} \quad I_{\theta}^{\mathcal{P}}(\lambda_0, \theta_0) \sim \lambda_0 I^f(\theta_0). \tag{25}$$

On the other hand, we show the following asymptotic equivalences, as  $\lambda_0 \rightarrow 0$ ,

$$I_{\lambda}^{\mathcal{B}}(\lambda_0, \theta_0) - I_{\lambda}^{\mathcal{P}}(\lambda_0, \theta_0) = o(\lambda_0), \tag{26}$$

$$I_{\theta}^{\mathcal{B}}(\lambda_0, \theta_0) - I_{\theta}^{\mathcal{P}}(\lambda_0, \theta_0) = O(\lambda_0^2), \tag{27}$$

$$I_{\lambda, \theta}^{\mathcal{B}}(\lambda_0, \theta_0) = O(\lambda_0). \tag{28}$$

Combining (18) and (23) yields (26). To establish (27), applying Theorem 2 in the Poisson case (see (4) and (12)), and using the notation therein, we obtain, for all  $z \in \mathbb{R}_+$ ,

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

By the Cauchy–Schwarz inequality and since  $\|\dot{\zeta}_{\theta_0}\|_2 = 1$ , we have

$$\sup_{z \geq 0} \left| \int_z^{\infty} \dot{\zeta}_{\theta_0}(t) \dot{\zeta}_{\theta_0}(t) dt \right| \leq \|\dot{\zeta}_{\theta_0}\|_2.$$

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  $\lambda_0 \rightarrow 0$ ,  $\|\partial_{\theta} \zeta_{\lambda_0, \theta_0} - \sqrt{\lambda_0} \dot{\zeta}_{\theta_0}\|_2 = O(\lambda_0^{3/2})$ . This combined with (23) yields (27). Further applying Theorem 2 and using similar arguments as above and Lemma 3 yields (28).

Hence, Lemma 4 applies and together with (22) this completes the proof of the theorem.  $\square$

### 3.4. Exponential case

In view of the TCSPC application, we consider the special case where  $f_{\theta_0}$  belongs to the family of exponential distributions  $\{f_{\theta}(x) = \theta e^{-\theta x}, \theta > 0\}$  and  $p_{\lambda}$  is the Poisson distribution  $\text{Poi}(\lambda)$ . We will see that in this case it is relatively easy to determine the optimal value of  $\lambda_0$  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  $\lambda$  is small. The following theorem describes the asymptotic behaviour of the Fisher information  $I_{\theta}$  when  $\lambda$  tends to infinity. We state that the pile-up model with large  $\lambda$  contains (asymptotically) the same amount of information on the exponential parameter  $\theta$  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  $p_{\lambda}$  be the Poisson distribution  $\text{Poi}(\lambda)$ . Denote the Fisher information of the parameter  $\theta$  associated with the pile-up model by  $I_{\theta}$ . Then, for all  $\theta_0 > 0$ ,

$$\lim_{\lambda_0 \rightarrow \infty} I_{\theta}(\lambda_0, \theta_0) = \frac{1}{\theta_0^2}. \tag{29}$$

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  $\lambda \rightarrow \infty$ .

**Proof.** The result follows by applying  $n$  times Hospital's rule to

$$\begin{aligned} \lim_{\lambda \rightarrow \infty} \frac{\lambda^n b_\lambda(m,n)}{e^\lambda} &= \lim_{\lambda \rightarrow \infty} \frac{m! \sum_{k=0}^\infty \frac{\lambda^{k+n}}{k!(k+n)^{m+1}}}{e^\lambda} = \lim_{\lambda \rightarrow \infty} \frac{m! \sum_{k=0}^\infty \frac{\lambda^k (k+n) \cdots (k+1)}{k!(k+n)^{m+1}}}{e^\lambda} \\ &\leq m! \lim_{\lambda \rightarrow \infty} e^{-\lambda} \sum_{k=0}^\infty \frac{\lambda^k}{k!(k+n)^{m+1-n}} \leq m! \lim_{\lambda \rightarrow \infty} e^{-\lambda} \sum_{k=0}^\infty \frac{\lambda^k}{(k+1)!} = m! \lim_{\lambda \rightarrow \infty} \frac{1-e^{-\lambda}}{\lambda} = 0. \quad \square \end{aligned}$$

**Proof of Theorem 6.** Note that the Fisher information of  $\theta$  has the following explicit form:

$$\begin{aligned} I_\theta(\lambda_0, 1) &= 4 \int (\partial_\theta \check{\zeta}_{\lambda_0, 1}(z))^2 dz = \lambda_0 e^{\lambda_0} \int_0^\infty e^{-z + \lambda_0 e^{-z}} (1-z + \lambda_0 z e^{-z})^2 dz \\ &= 1 - e^{-\lambda_0} + \lambda_0 e^{-\lambda_0} \{b_{\lambda_0}(2,1) + \lambda_0^2 b_{\lambda_0}(2,3) - 2b_{\lambda_0}(1,1) - 2\lambda_0 b_{\lambda_0}(1,2) + 2\lambda_0 b_{\lambda_0}(2,2)\}. \end{aligned}$$

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  $\lambda_0 \rightarrow \infty$ . By applying Hospital's rule to the first term three times and to the second term two times, we obtain

$$\begin{aligned} \lim_{\lambda_0 \rightarrow \infty} |e^{-\lambda_0} \lambda_0^3 b_{\lambda_0}(2,3) - 2e^{-\lambda_0} \lambda_0^2 b_{\lambda_0}(1,2)| &= 2 \left| \lim_{\lambda_0 \rightarrow \infty} e^{-\lambda_0} \sum_{k=0}^\infty \frac{\lambda_0^k (k+1)(k+2)}{k!(k+3)^2} - 2 \lim_{\lambda_0 \rightarrow \infty} e^{-\lambda_0} \sum_{k=0}^\infty \frac{\lambda_0^k (k+1)}{k!(k+2)} \right| \\ &= 2 \lim_{\lambda_0 \rightarrow \infty} e^{-\lambda_0} \left| \sum_{k=0}^\infty \frac{\lambda_0^k}{k!} \frac{2k^2 + 7k + 5}{k^3 + 8k^2 + 21k + 18} \right| \leq 28 \lim_{\lambda_0 \rightarrow \infty} e^{-\lambda_0} \left| \sum_{k=0}^\infty \frac{\lambda_0^k}{(k+1)!} \right| = 0. \end{aligned}$$

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  $\lambda_0 \rightarrow \infty$ . This implies that  $I_\theta(\lambda_0, 1)$  tends to 1 as  $\lambda_0 \rightarrow \infty$ , and (20) gives (29).  $\square$

From Fig. 2 we see that the Cramér–Rao bound  $\lambda \mapsto CR_\theta(\lambda, 1)$  is convex in  $\lambda$ . A numerical minimization yields a minimum at  $\lambda_{\text{opt}} \approx 1.3275$  and Theorem 4 implies that  $\lambda_{\text{opt}}$  minimizes the Cramér–Rao bound  $\lambda \mapsto CR_\theta(\lambda, \theta_0)$  for all  $\theta_0 > 0$ . The value of  $\lambda_{\text{opt}}$  is much larger than the intensity usually used in fluorescence measurements ( $\lambda < 0.05$ ). We conclude that it is worthwhile to increase the intensity  $\lambda$  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.

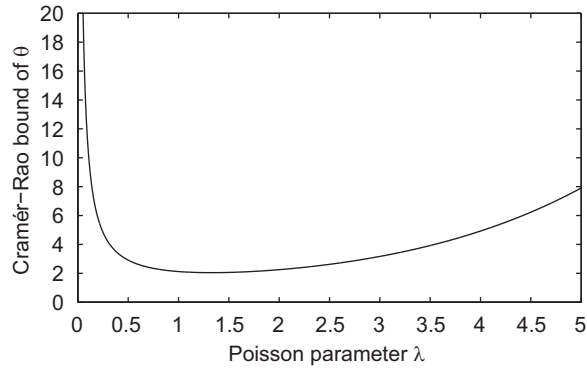
### 3.5. Multi-exponential case

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

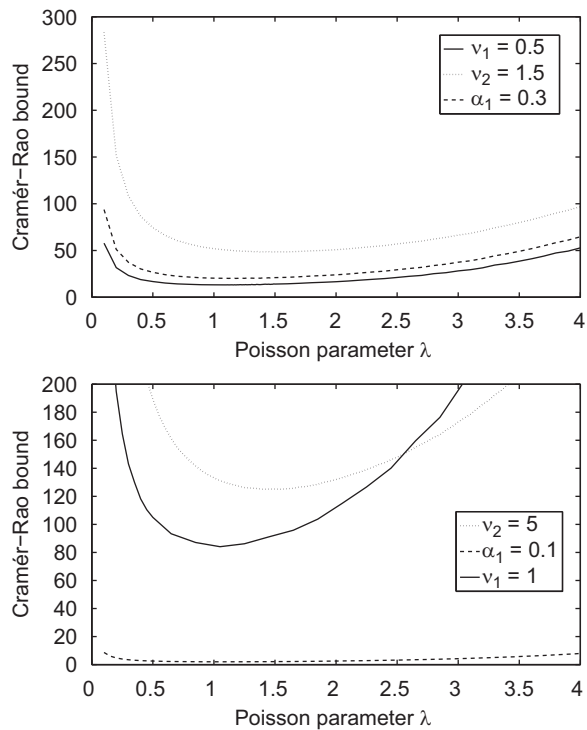
$$f_\theta(y) = \sum_{k=1}^K \alpha_k v_k e^{v_k y}, \tag{30}$$

where  $v_k > 0$  and  $0 < \alpha_k < 1$ . Note that  $\alpha_K$  is given indirectly by  $\alpha_K = 1 - \sum_{k=1}^{K-1} \alpha_k$ . Denote by  $\theta = (\alpha_1, \dots, \alpha_{K-1}, v_1, \dots, v_K)^T$  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  $\Theta$ . However, mixtures with two components or more ( $K \geq 2$ ) do not form a scale family. Consequently, Theorem 4 does not apply. That means, that the optimal value of  $\lambda_0$  that minimizes the Cramér–Rao bound will depend on the value of  $\theta_0$ . Hence, there is no global optimal value  $\lambda_{\text{opt}}$  as in the exponential case.

For a two-component mixture, Fig. 3 shows the graphs of the Cramér–Rao bounds of the associated pile-up model for different choices of parameters  $\theta_0$  in the Poisson case. Although all functions are convex in  $\lambda$ , note that no unique optimal value  $\lambda_{\text{opt}}$  minimizing all Cramér–Rao bounds simultaneously exists. The minimum values depend on  $\theta_0$ . Nevertheless, the graphs suggest choosing  $\lambda$  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.



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



**Fig. 3.** Cramér–Rao bounds  $\lambda \mapsto CR(\lambda, \theta_0)$  for two-component models.

**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  $\lambda_{\text{opt}}$  minimizing the Cramér–Rao bound exists. Simulations in the Poisson case reveal that the optimal value depends on the underlying distribution  $F_{\theta_0}$ , the larger the probability of censored observations, the larger the optimal value  $\lambda_{\text{opt}}$ . This phenomenon suggests that if the maximal observation time  $T$  is relatively short with respect to the tail of the distribution  $F_{\theta_0}$  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. The only estimation method known in fluorescence that takes into account the pile-up effect is a correction method for pile-up affected histograms, but this method does not provide parameter estimators (Coates, 1968). The pile-up likelihood is easy to compute by using (5), but even with standard models for the density  $f$ , it is difficult to maximize. While general

optimization techniques as simulated annealing (Robert and Casella, 2004, Section 5.2.3) may be applicable to maximize the likelihood function, here we propose a Bayesian approach based on a noninformative prior. It is known that under convenient regularity conditions, Bayes estimators as the posterior mean are asymptotically efficient, that is that the limit variance is independent from the prior distribution and coincides with the Cramér–Rao bound (see, e.g. Lehmann and Casella, 1998, Section 6.8 or Ibragimov and Has'minskii, 1981, Section 3.2).

In this section we focus on 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

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

shows that  $Z$  has an infinite exponential mixture distribution where the set of exponential parameters is given by  $v_1 \mathbb{N} + v_2 \mathbb{N} + \dots + v_K \mathbb{N}$ .

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 and 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).

The EM algorithm is based on the introduction of missing variables to obtain a conditional density of simple form. To our knowledge, there are no missing variables for the pile-up model such that the E-step becomes explicit. In Rebafka et al. (2009) a modified likelihood approach for the pile-up model is proposed which is based on a new contrast and that can be computed by an EM-type algorithm. The NPMLE methods seem to be inappropriate for the pile-up model 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, which is done in the following.

#### 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  $\delta_T$  the measure having mass one in  $T$  given by

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

Using mixture distributions we are confronted with the label switching problem as the mixture components are exchangeable. We may obtain identifiability of the model by ordering the exponential parameters

$$v_1 > v_2 > \dots > v_K.$$

To circumvent the inconveniences that go along with such an identifiability constraint (see, e.g. Marin et al., 2005, Section 1.3.4), we reparametrize the exponential parameters as it is done in Gruet et al. (1999). Hence, we consider the following parameters:

$$\varphi = v_1, \tau_k = \frac{v_k}{v_{k-1}}, \quad k = 2, \dots, K.$$

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  $\alpha_k$  are chosen in the entire set  $\{(\alpha_1, \dots, \alpha_K) \in [0, 1]^K : \sum_{k=1}^K \alpha_k = 1\}$ . We set

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

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

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

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  $\mathbf{b} = (B_1, \dots, B_K)$  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 are generated by first drawing the total number of photons  $N$  from the distribution

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

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

$$P(\mathbf{b} = (b_1, \dots, b_K) | N = n, Z, \theta) = \begin{cases} \frac{(n-1)! \prod_k \alpha_k^{b_k} \sum_k b_k v_k}{f(Z)(1-F(Z))^{n-1} \prod_k b_k!} e^{-Z \sum_k b_k v_k} & \text{for } n \geq 1, \text{ if } Z \in (0, T), \\ \frac{n! \prod_k \alpha_k^{b_k}}{(1-F(T))^n \prod_k b_k!} e^{-T \sum_k b_k v_k} & \text{for } n \geq 0, \text{ if } Z = T, \end{cases}$$

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

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

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

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

where  $\tau_1 = 1$  and  $q_K = 1$ . Furthermore, the posterior distribution of  $\tau_l$  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

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

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  $\alpha_k$  and any distinct values for the exponential parameters  $v_k$ . For the intensity  $\lambda$  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 Fig. 4. 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). 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  $\lambda = 1.5, v_1 = 2, v_2 = 0.5, \alpha_1 = 0.3$  and 5000 observations. From

**Table 1**  
Simulation results for  $K=1,2,3$  components and 3000, 3000, 10,000 observations, resp.

|                  | K=1       |       | K=2       |       |            |            |
|------------------|-----------|-------|-----------|-------|------------|------------|
|                  | $\lambda$ | $v$   | $\lambda$ | $v_1$ | $v_2$      | $\alpha_1$ |
| True values      | 3         | 0.1   | 1         | 6     | 0.5        | 0.75       |
| Estimated values | 3.05      | 0.101 | 0.994     | 6.03  | 0.518      | 0.742      |
|                  | K=3       |       | $v_2$     | $v_3$ | $\alpha_1$ | $\alpha_2$ |
|                  | $\lambda$ | $v_1$ |           |       |            |            |
| True values      | 1         | 5     | 1         | 0.1   | 0.33       | 0.33       |
| Estimated values | 1.01      | 4.88  | 0.970     | 0.101 | 0.347      | 0.329      |

50 simulations we obtain the following empirical covariance matrix for the estimates of  $(\lambda, v_1, v_2, \alpha_1)$ :

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

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

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

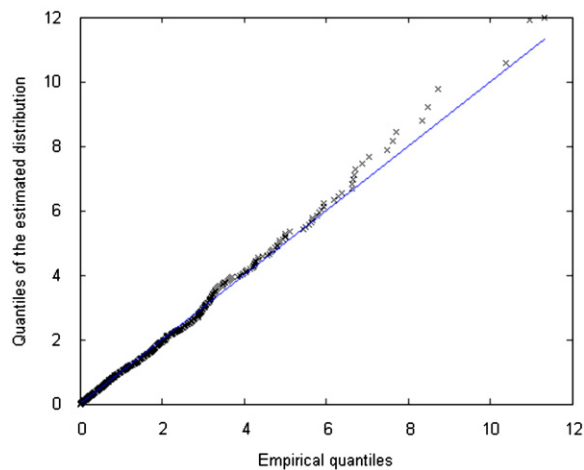


Fig. 4. Quantile–quantile plot for the two-component model with the true parameter values  $\lambda = 1, (v_1, v_2) = (6, 0.5), (\alpha_1, \alpha_2) = (0.75, 0.25)$ .

Table 2

Comparison of the standard estimation method and the MCMC method for a two-component model with true parameter values  $\alpha_1 = 0.33, \alpha_2 = 0.66, v_1 = 0.5, v_2 = 6$ .

| nb of obs   | 1000   | 3000   | 5000   | 10,000        |
|---|--------|--------|--------|---------------|
| <i>Standard estimation method (<math>\lambda = 0.05</math>)</i> |        |        |        |               |
| $\alpha_1$  |        |        |        |               |
| bias  | 0.0051 | 0.0040 | 0.0036 | 0.0028        |
| variance  | 0.0112 | 0.0035 | 0.0021 | 0.0010        |
| $v_1$   |        |        |        |               |
| bias  | 0.0488 | 0.0143 | 0.0098 | 0.0056        |
| variance  | 0.0447 | 0.0092 | 0.0054 | 0.0025        |
| $v_2$   |        |        |        |               |
| bias  | 0.6523 | 0.2198 | 0.1626 | 0.1080        |
| variance  | 6.6068 | 0.9788 | 0.5520 | 0.2606        |
| <i>MCMC method (<math>\lambda = 1.32</math>)</i>                |        |        |        |               |
| $\alpha_1$  |        |        |        |               |
| bias  | 0.0006 | 0.0001 | 0.0001 | $0.492e^{-4}$ |
| variance  | 0.0008 | 0.0003 | 0.0002 | 0.0001        |
| $v_1$   |        |        |        |               |
| bias  | 0.0049 | 0.0018 | 0.0010 | 0.0006        |
| variance  | 0.0023 | 0.0008 | 0.0005 | 0.0002        |
| $v_2$   |        |        |        |               |
| bias  | 0.0278 | 0.0069 | 0.0057 | 0.0023        |
| variance  | 0.1499 | 0.0491 | 0.0294 | 0.0147        |

Both matrices are quite close to each other, in particular, the variances of the posterior means obtained by the Gibbs sampler are close to the Cramér–Rao bound. We observe that some variances are even slightly smaller. This may be explained by the fact that a Bayesian approach is used and possibly prior information is used to reduce the variance. 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 two 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 (30) 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 presented in Table 2 we see that any bias and any variance of the standard method exceed the corresponding value of the MCMC method. Actually, comparing the results for 10,000 observations of the first method to the results for 1000 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 10 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 Rebařka (2007).

## Acknowledgment

We would like to thank the referee for the helpful comments.

## 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.
- Ibragimov, I.A., Has'minskii, R.Z., 1981. Statistical Estimation: Asymptotic Theory. Springer, New York.
- 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.
- Lehmann, E.L., Casella, G., 1998. Theory of Point Estimation. Springer, New York.
- Marin, J.-M., Mengersen, K., Robert, C.P., 2005. Bayesian modelling and inference on mixtures of distributions. In: Dey, D., C.R. (Eds.), *Handbook of Statistics*, vol. 25. Elsevier-Sciences (Chapter 16).
- 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>>.
- Rebařka, T., 2007. An MCMC approach for estimating a fluorescence lifetime with pile-up distortion. In: GRETSI 2007: 21<sup>ème</sup> colloque sur le traitement du signal et des images, Troyes, France.
- Rebařka, T., Roueff, F., Souloumiac, A., 2009. A corrected likelihood approach for the nonlinear transformation model with application to fluorescence lifetime measurements using exponential mixtures. Under revision. Available at <<http://hal.archives-ouvertes.fr/hal-00448724/fr>>.
- Robert, C.P., Casella, G., 2004. Monte Carlo Statistical Methods. Springer, New York.
- 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.