

# A BAYESIAN LOWER BOUND FOR PARAMETER ESTIMATION OF POISSON DATA INCLUDING MULTIPLE CHANGES

*Lucien Bacharach\**

*Mohammed Nabil El Korso<sup>†</sup>*

*Alexandre Renaux\**

*Jean-Yves Tournet<sup>‡</sup>*

<sup>\*</sup> Université Paris-Sud/L2S, 91192 Gif-sur-Yvette, France

<sup>†</sup> Université Paris-Ouest/LEME, 91140 Ville d'Avray, France

<sup>‡</sup> IRT/ENSEEIH/TéSA, 31071 Toulouse Cedex 7, France

## ABSTRACT

This paper derives lower bounds for the mean square errors of parameter estimators in the case of Poisson distributed data subjected to multiple abrupt changes. Since both change locations (discrete parameters) and parameters of the Poisson distribution (continuous parameters) are unknown, it is appropriate to consider a mixed Cramér-Rao/Weiss-Weinstein bound for which we derive closed-form expressions and illustrate its tightness by numerical simulations.

**Index Terms**— Bayesian Cramér-Rao bound, Weiss-Weinstein bound, change-point, Poisson time series.

## 1. INTRODUCTION

The problem of estimating change-point locations in time series has received much attention in the literature [1–3]. Specifically, discrete Poisson time series submitted to multiple abrupt changes are of specific interest, since Poisson distributions are involved in most problems dealing with counting processes [4, 5]. This type of models arise in signal processing, for applications such as astronomical time series segmentation [6–8]. In this context, the observed data (distributed according to a Poisson distribution) are subjected to abrupt changes whose locations are unknown. The values of the Poisson parameters associated with each interval are also unknown quantities that need to be estimated [9]. Several strategies have been investigated in the literature [10–14]. However, to the best of our knowledge, lower bounds for the mean square error of the resulting simulators have been derived only in a few specific cases. For instance, the case of a single change-point in the observation window was studied in [15]. The case of multiple changes was considered in [16]. The difficulty of deriving bounds for the parameters of piece-wise stationary data is mainly due to the discrete nature of change-point locations for which classical bounds such as the Cramér-Rao bound (CRB) are not appropriate anymore [17]. In [15, 16], the authors considered a lower bound for the mean square error (MSE) that does not require the differentiability of the log-likelihood function. Specifically, deterministic bounds, such as the Chapman-Robbins bound, have been derived for a single change-point in [15], and then extended to multiple changes in [16], with the strong assumptions that the Poisson parameters are known. On the other hand, and in order to improve the tightness of the resulting bound, we proposed in [18, 19] the use of the Weiss-Weinstein bound (WWB), which is known to be one of the tightest bound in the family of the Bayesian bounds [20, 21]. Nevertheless, these analyses were limited to the case of known Poisson parameters both in the single [18] and multiple [19] change-point scenarios.

In this paper, we fill this gap by proposing and deriving a new Bayesian lower bound for the global MSE (GMSE) of the parameters of Poisson distributed data subjected to multiple change-points.

The proposed bound is a mixed Bayesian Cramér-Rao (CR)/Weiss-Weinstein (WW) bound adapted to the fact that some unknown parameters are discrete (the WW part of the mixed bound is associated with the change locations) and that the other parameters are continuous (the CR part of the mixed bound is associated with the Poisson parameters). The idea of combining these two bounds was introduced in [22], with a recursive form. Of course, using a WWB for both discrete and continuous parameters would be theoretically possible. However, the WW bound is expressed as the supremum of a set of matrices whose computation is unfeasible in our scenario. Thus, the mixed Bayesian Cramér-Rao/Weiss-Weinstein bound (BCRWBB) is the appropriate alternative, whose computation can be achieved using a convex optimization procedure to compute this supremum based on the computation of the minimum volume ellipsoid covering a union of derived ellipsoids.

## 2. MULTIPLE CHANGE-POINTS IN POISSON TIME-SERIES: PROBLEM FORMULATION

We consider an independent discrete Poisson time series subjected to multiple changes. The resulting observation vector  $\mathbf{x} = [x_1, \dots, x_T]^T$  (of length  $T$ ) is defined as

$$\begin{cases} x_t \sim \mathcal{P}(\lambda_1), & \text{for } t = 1, \dots, t_1 \\ x_t \sim \mathcal{P}(\lambda_2), & \text{for } t = t_1 + 1, \dots, t_2 \\ \vdots & \vdots \\ x_t \sim \mathcal{P}(\lambda_{K+1}), & \text{for } t = t_K + 1, \dots, T, \end{cases} \quad (1)$$

where  $\mathcal{P}(\lambda_k)$ , for  $k = 1, \dots, K + 1$ , denotes the Poisson distribution of parameter  $\lambda_k$  on the  $k$ -th segment, i.e.,  $\Pr(x_t = \kappa_t) = \lambda_k^{\kappa_t} \exp\{-\lambda_k\} / (\kappa_t!)$ ,  $K$  denotes the total number of changes (assumed to be known), and  $t_k$  denotes the  $k$ -th change location, i.e., the sample point after which the parameter  $\lambda_k$  of the current segment switches to  $\lambda_{k+1}$ . The segmentation problem addressed in this work consists of i) segmenting the time series  $\mathbf{x}$ , i.e., estimating the locations of the changes  $t_k$ , and ii) estimating the Poisson parameters  $\lambda_k$  on each segment. The resulting unknown parameter vector is  $\boldsymbol{\theta} = [\boldsymbol{\lambda}^T, \mathbf{t}^T]^T$ , with  $\boldsymbol{\lambda} \triangleq [\lambda_1, \dots, \lambda_{K+1}]^T$  and  $\mathbf{t} \triangleq [t_1, \dots, t_K]^T$ . This unknown parameter vector lies in the parameter space  $\Theta = \mathbb{R}_+^{K+1} \times \{1, \dots, T\}^K$ , where  $\mathbb{R}_+$  denotes the set of real positive numbers. Using a Bayesian framework, we consider that both vectors  $\boldsymbol{\lambda}$  and  $\mathbf{t}$  are assigned a known prior. More precisely, the Poisson parameters  $\lambda_k$  are assumed to be independent and identically distributed (i.i.d.), and are assigned the conjugate gamma distributions with parameters  $\alpha_k$  and  $\beta$  (as in [8]), leading to the

following prior

$$f(\boldsymbol{\lambda}) = \prod_{k=1}^{K+1} \frac{\beta^{\alpha_k}}{\Gamma(\alpha_k)} \lambda_k^{\alpha_k-1} \exp(-\beta \lambda_k) \mathbb{I}_{\mathbb{R}_+}(\lambda_k) \quad (2)$$

in which  $\mathbb{I}_{\mathcal{E}}(\cdot)$  denotes the indicator function on the set  $\mathcal{E}$ , and  $\Gamma(\cdot)$  denotes the usual gamma function, i.e., for  $\alpha > 0$ ,  $\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} \exp(-x) dx$ . On the other hand, we assume that each change location  $t_k$ , for  $k = 1, \dots, K$ , is defined as the following random walk  $t_k = t_{k-1} + \varepsilon_k$  where  $\varepsilon_k$  are i.i.d. variables following a discrete uniform distribution on the set of integers  $\{1, \dots, \tau\}$ , and  $t_0 = 0$ . The value of  $\tau$  is chosen so that the final change  $t_K$  does not exceed  $T$ , i.e., the maximum possible value for  $\tau$  is  $\tau_{\max} = \lfloor (T-1)/K \rfloor$ , where  $\lfloor \cdot \rfloor$  denotes the floor function. This assumption might seem quite restrictive since there is generally no reason for all durations to be restricted to  $\tau$ . However, this hypothesis is necessary to make the derivations tractable. Consequently, using the probability chain rule, we obtain the following prior distribution for the unknown vector  $\mathbf{t}$

$$\Pr(\mathbf{t} = \boldsymbol{\ell}) = \frac{1}{\tau^K} \prod_{k=1}^K \mathbb{I}_{\{\ell_{k-1}+1, \dots, \ell_{k-1}+\tau\}}(\ell_k) \quad (3)$$

with  $\ell_0 = 0$ . Since vectors  $\boldsymbol{\lambda}$  and  $\mathbf{t}$  are independent, the joint prior for  $\boldsymbol{\lambda}$  and  $\mathbf{t}$  is expressed as  $f(\boldsymbol{\lambda}, \mathbf{t} = \boldsymbol{\ell}) = f(\boldsymbol{\lambda}) \Pr(\mathbf{t} = \boldsymbol{\ell})$ .

From the model (1), the likelihood of the observations can be written as

$$f(\mathbf{x} = \boldsymbol{\kappa} | \boldsymbol{\lambda}, \mathbf{t} = \boldsymbol{\ell}) = \prod_{k=1}^{K+1} \prod_{t=\ell_{k-1}+1}^{\ell_k} \frac{\lambda_k^{\kappa_t}}{\kappa_t!} \exp\{-\lambda_k\}. \quad (4)$$

The aim of the present paper is to study the estimation performance of the vector  $\boldsymbol{\theta}$  by deriving a lower bound on the mean square error (MSE) of any Bayesian estimator  $\hat{\boldsymbol{\theta}}(\mathbf{x})$  of  $\boldsymbol{\theta}$ . Both subvectors  $\boldsymbol{\lambda}$  and  $\mathbf{t}$  of  $\boldsymbol{\theta}$  have to be estimated simultaneously. However, as already mentioned in Section 1, the Cramér-Rao bound is not suited for changepoint analysis, since  $\boldsymbol{\ell}$  is a vector of discrete parameters. Thus, the idea is to use two different lower bounds w.r.t. each subvector of  $\boldsymbol{\theta}$ , resulting in a “mixed” Bayesian bound that corresponds to the Bayesian CR bound for the first subvector  $\boldsymbol{\lambda}$  of  $\boldsymbol{\theta}$ , and that corresponds to the so-called WW bound for the second subvector  $\mathbf{t}$  of  $\boldsymbol{\theta}$ . As already mentioned, the use of such a combined CR/WW lower bound was proposed in [22] in a target tracking context. Since our framework is different, the next section is devoted to the presentation of this bound referred to as BCRWWB. It is in fact a special case of a general family of lower bounds exposed in [23].

### 3. BAYESIAN CRAMÉR-RAO/WEISS-WEINSTEIN BOUND

We are interested in studying the estimation performance of a parameter vector  $\boldsymbol{\theta}$  that lies in a parameter space  $\Theta = \mathbb{R}^{K+1} \times \mathbb{N}^K$ . As explained in the previous section, this parameter vector can be split into two subvectors,  $\boldsymbol{\lambda} \in \Theta_{\boldsymbol{\lambda}} = \mathbb{R}_+^{K+1}$  and  $\mathbf{t} \in \Theta_{\mathbf{t}} = \mathbb{N}^K$ , so that  $\boldsymbol{\theta} = [\boldsymbol{\lambda}^T, \mathbf{t}^T]^T$  and  $\Theta = \Theta_{\boldsymbol{\lambda}} \times \Theta_{\mathbf{t}}$ . From a set of observations  $\mathbf{x} \in \Omega$ , the vector  $\boldsymbol{\theta}$  can be estimated by using any Bayesian estimation scheme, leading to an estimator  $\hat{\boldsymbol{\theta}}(\mathbf{x}) = [\hat{\boldsymbol{\lambda}}(\mathbf{x})^T, \hat{\mathbf{t}}(\mathbf{x})^T]^T$ . Let us recall that we aim at obtaining a lower bound on the global mean square error (GMSE) of this estimator, which corresponds to the Bayesian CR bound w.r.t.  $\boldsymbol{\lambda}$ , and which corresponds to the WW bound w.r.t.  $\mathbf{t}$ . The GMSE of  $\hat{\boldsymbol{\theta}}(\mathbf{x})$  is defined as the  $(2K+1) \times (2K+1)$  matrix

$$\text{GMSE}(\hat{\boldsymbol{\theta}}) = \mathbb{E}_{\mathbf{x}, \boldsymbol{\theta}} \left\{ [\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}(\mathbf{x})][\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}(\mathbf{x})]^T \right\} \quad (5)$$

in which  $\mathbb{E}_{\mathbf{x}, \boldsymbol{\theta}} \{\cdot\}$  denotes the expectation operation w.r.t. the joint distribution  $f(\mathbf{x}, \boldsymbol{\theta})$  which depends on both the observations and the parameters. Based on [23], by appropriately choosing some real-valued measurable functions  $\psi_k(\mathbf{x}, \boldsymbol{\theta})$ ,  $k = 1, \dots, 2K+1$ , defined on  $\Omega \times \Theta$  such that the following integrals exist and satisfy  $\int_{\Theta} \psi_k(\mathbf{x}, \boldsymbol{\theta}) f(\mathbf{x}, \boldsymbol{\theta}) d\boldsymbol{\theta} = 0$  for almost every (a.e.)  $\mathbf{x} \in \Omega$  and for  $k = 1, \dots, 2K+1$ , the following matrix inequality holds

$$\text{GMSE}(\hat{\boldsymbol{\theta}}) \succeq \mathbf{V} \mathbf{P}^{-1} \mathbf{V}^T \quad (6)$$

in which  $\mathbf{V}$  is a  $(2K+1) \times (2K+1)$  matrix whose elements are given by

$$[\mathbf{V}]_{k,l} = \mathbb{E}_{\mathbf{x}, \boldsymbol{\theta}} \{\theta_k \psi_l(\mathbf{x}, \boldsymbol{\theta})\} \quad (7)$$

and  $\mathbf{P}$  is a  $(2K+1) \times (2K+1)$  symmetric matrix, whose elements are given by

$$[\mathbf{P}]_{k,l} = \mathbb{E}_{\mathbf{x}, \boldsymbol{\theta}} \{\psi_k(\mathbf{x}, \boldsymbol{\theta}) \psi_l(\mathbf{x}, \boldsymbol{\theta})\}. \quad (8)$$

Note that the matrix inequality (6) means that the difference between its left and its right hand sides is a nonnegative definite matrix. One key point in the theory developed in [23] is the choice of the measurable functions  $\psi_k$ . For  $k$  restricted to  $\{1, \dots, K+1\}$  (continuous Poisson parameters), we define these functions as for the CR bound, i.e.,

$$\psi_k(\mathbf{x}, \boldsymbol{\theta}) = \begin{cases} \frac{\partial \ln f(\mathbf{x}, \boldsymbol{\theta})}{\partial \lambda_k}, & \text{if } \boldsymbol{\theta} \in \Theta' \\ 0, & \text{if } \boldsymbol{\theta} \notin \Theta' \end{cases} \quad (9)$$

where  $\Theta' = \{\boldsymbol{\theta} \in \Theta : f(\mathbf{x}, \boldsymbol{\theta}) > 0 \text{ a.e. } \mathbf{x} \in \Omega\}$ . Conversely, for  $k$  restricted to  $\{1, \dots, K\}$ , we define these measurable functions as for the WW bound, i.e.,

$$\psi_{K+1+k}(\mathbf{x}, \boldsymbol{\theta}) = \sqrt{\frac{f(\mathbf{x}, \boldsymbol{\theta} + \mathbf{h}_k)}{f(\mathbf{x}, \boldsymbol{\theta})}} - \sqrt{\frac{f(\mathbf{x}, \boldsymbol{\theta} - \mathbf{h}_k)}{f(\mathbf{x}, \boldsymbol{\theta})}} \quad (10)$$

where  $\mathbf{h}_k$  is any vector of size  $2K+1$  of the form  $\mathbf{h}_k = [\mathbf{0}_{K+1}^T, \mathbf{0}_{k-1}^T, h_k, \mathbf{0}_{K-k}^T]^T$ , for  $k = 1, \dots, K$ , in which  $\mathbf{0}_k$  denotes the zero vector of length  $k$ . Note that the value of  $h_k$  can be arbitrarily chosen by the user as far as it allows the invertibility of  $\mathbf{P}$ .

The next step in our analysis is to derive the matrix  $\mathbf{V}$ . Denote as  $\mathbf{V}_{22}$  the  $K \times K$  diagonal matrix whose elements are, for any  $k \in \{1, \dots, K\}$

$$[\mathbf{V}_{22}]_{k,k} = -h_k \mathbb{E}_{\mathbf{x}, \boldsymbol{\theta}} \left\{ \sqrt{\frac{f(\mathbf{x}, \boldsymbol{\theta} + \mathbf{h}_k)}{f(\mathbf{x}, \boldsymbol{\theta})}} \right\}. \quad (11)$$

Substituting (9) and (10) into (7), we obtain

$$\mathbf{V} = \begin{bmatrix} -\mathbf{I}_{K+1} & \mathbf{0}_{(K+1) \times K} \\ \mathbf{0}_{K \times (K+1)} & \mathbf{V}_{22} \end{bmatrix} \quad (12)$$

where  $\mathbf{I}_{K+1}$  denotes the  $(K+1) \times (K+1)$  identity matrix and  $\mathbf{0}_{(K+1) \times K}$  is the  $(K+1) \times K$  zero matrix, provided the following conditions are satisfied

1.  $f(\mathbf{x}, \boldsymbol{\theta})$  is absolutely continuous w.r.t.  $\lambda_k$ ,  $k = 1, \dots, K+1$ , a.e.  $\mathbf{x} \in \Omega$ ;
2.  $\lim_{\lambda_k \rightarrow 0} \lambda_k f(\mathbf{x}, \boldsymbol{\theta}) = \lim_{\lambda_k \rightarrow +\infty} \lambda_k f(\mathbf{x}, \boldsymbol{\theta}) = 0$ ,  $k = 1, \dots, K+1$ , a.e.  $\mathbf{x} \in \Omega$ .

Note that these two conditions correspond to the necessary and usual regularity conditions for the derivation of the Bayesian CR bound.

Similarly, by plugging (9) and (10) into (8), we obtain the expression of the matrix  $\mathbf{P}$ , which can be split into four blocks as follows

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} \\ \mathbf{P}_{12}^T & \mathbf{P}_{22} \end{bmatrix} \quad (13)$$

in which  $\mathbf{P}_{11}$  is the  $(K+1) \times (K+1)$  matrix whose elements are

$$[\mathbf{P}_{11}]_{k,l} = \mathbb{E}_{\mathbf{x},\boldsymbol{\theta}} \left\{ \frac{\partial \ln f(\mathbf{x}, \boldsymbol{\theta})}{\partial \lambda_k} \frac{\partial \ln f(\mathbf{x}, \boldsymbol{\theta})}{\partial \lambda_l} \right\} \quad (14)$$

$\mathbf{P}_{12}$  is the  $(K+1) \times K$  matrix whose elements are

$$[\mathbf{P}_{12}]_{k,l} = \mathbb{E}_{\mathbf{x},\boldsymbol{\theta}} \left\{ \frac{\partial \ln f(\mathbf{x}, \boldsymbol{\theta})}{\partial \lambda_k} \left( \sqrt{\frac{f(\mathbf{x}, \boldsymbol{\theta} + \mathbf{h}_l)}{f(\mathbf{x}, \boldsymbol{\theta})}} - \sqrt{\frac{f(\mathbf{x}, \boldsymbol{\theta} - \mathbf{h}_l)}{f(\mathbf{x}, \boldsymbol{\theta})}} \right) \right\} \quad (15)$$

and finally  $\mathbf{P}_{22}$  is the  $K \times K$  matrix whose elements are

$$[\mathbf{P}_{22}]_{k,l} = \mathbb{E}_{\mathbf{x},\boldsymbol{\theta}} \left\{ \left( \sqrt{\frac{f(\mathbf{x}, \boldsymbol{\theta} + \mathbf{h}_k)}{f(\mathbf{x}, \boldsymbol{\theta})}} - \sqrt{\frac{f(\mathbf{x}, \boldsymbol{\theta} - \mathbf{h}_k)}{f(\mathbf{x}, \boldsymbol{\theta})}} \right) \times \left( \sqrt{\frac{f(\mathbf{x}, \boldsymbol{\theta} + \mathbf{h}_l)}{f(\mathbf{x}, \boldsymbol{\theta})}} - \sqrt{\frac{f(\mathbf{x}, \boldsymbol{\theta} - \mathbf{h}_l)}{f(\mathbf{x}, \boldsymbol{\theta})}} \right) \right\}. \quad (16)$$

Note that the same bound can be obtained by defining all the functions  $\psi_k$  for  $k \in \{1, \dots, 2K+1\}$  as in (10), with  $\mathbf{h}_k = [\mathbf{0}_{k-1}^T, h_k, \mathbf{0}_{2K+1-k}^T]^T$ , by letting  $h_k$  tend to 0 for  $k = 1, \dots, K+1$ , and by using a Taylor expansion. Finally, the tightest lower bound is obtained by maximizing the right hand side of (6) w.r.t.  $h_1, \dots, h_K$ .

#### 4. APPLICATION TO MULTIPLE CHANGE-POINTS IN POISSON TIME-SERIES

This section presents the main results about the derivation of the lower bound presented in Section 3 for the problem formulated in Section 2. Due to space limitations, it is not possible to provide the full calculation details in this paper. However, all details can be found in the open-access technical report [24]. The joint distribution of the observation and parameter vectors can be expressed as

$$f(\mathbf{x} = \boldsymbol{\kappa}, \boldsymbol{\theta}) = f(\mathbf{x} = \boldsymbol{\kappa} | \boldsymbol{\lambda}, \mathbf{t} = \boldsymbol{\ell}) f(\boldsymbol{\lambda}, \mathbf{t} = \boldsymbol{\ell}). \quad (17)$$

After plugging (2), (3) and (4) into (17), we can deduce the expressions of  $f(\mathbf{x}, \boldsymbol{\theta} + \mathbf{h}_k)$  for  $k = 1, \dots, K$ , and  $\partial f(\mathbf{x}, \boldsymbol{\theta}) / \partial \lambda_k$  for  $k = 1, \dots, K+1$ . Let us first introduce some useful notations for the following mathematical functions. We first define the function  $\varphi_{h_k}(\mathbf{y})$  of the vector  $\mathbf{y} = [y_1, y_2]^T \in \mathbb{R}_+^2$  as

$$\varphi_{h_k}(\mathbf{y}) =$$

$$y_1^{\alpha_k-1} y_2^{\alpha_{k+1}-1} \exp \left\{ -\beta(y_1 + y_2) - |h_k| \frac{(\sqrt{y_2} - \sqrt{y_1})^2}{2} \right\} \quad (18)$$

and the following integral as

$$\Phi(h_k) = \frac{\beta^{\alpha_k + \alpha_{k+1}}}{\Gamma(\alpha_k) \Gamma(\alpha_{k+1})} \int_{\mathbb{R}_+^2} \varphi_{h_k}(\mathbf{y}) d\mathbf{y}. \quad (19)$$

We also define the function  $\phi_{h_k, h_{k+1}}(\mathbf{z})$  of the vector  $\mathbf{z} = [z_1, z_2, z_3]^T \in \mathbb{R}_+^3$  as the trivariate version of  $\varphi_{h_k}$ , i.e.,

$$\begin{aligned} \phi_{h_k, h_{k+1}}(\mathbf{z}) &= z_1^{\alpha_k-1} z_2^{\alpha_{k+1}-1} z_3^{\alpha_{k+2}-1} \\ &\times \exp \left\{ -\beta(z_1 + z_2 + z_3) - |h_k| \frac{(\sqrt{z_2} - \sqrt{z_1})^2}{2} \right. \\ &\quad \left. - |h_{k+1}| \frac{(\sqrt{z_3} - \sqrt{z_2})^2}{2} \right\}. \end{aligned} \quad (20)$$

We finally define the three functions  $u$ ,  $v$  and  $w$  as follows

$$u(\tau, h_k) = \begin{cases} \frac{(\tau - |h_k|)^2}{\tau^2} & \text{if } k \leq K-1 \text{ and } |h_k| \leq \tau \\ \frac{\tau - |h_K|}{\tau} & \text{if } k = K \text{ and } |h_K| \leq \tau \\ 0 & \text{if } |h_k| > \tau \end{cases} \quad (21)$$

$$v(\tau, h_k, h_{k+1}) = \begin{cases} \frac{(\tau - |h_k|)(\tau - |h_{k+1}|)}{\tau^3} & \text{if } k \leq K-1 \\ & \text{and } \max(|h_k|, |h_{k+1}|) \leq \tau; \\ \frac{\tau - |h_K|}{\tau^2} & \text{if } k = K \text{ and } |h_K| \leq \tau; \\ 0 & \text{if } \max(|h_k|, |h_{k+1}|) > \tau, \end{cases} \quad (22)$$

$$\begin{aligned} w(\mathbf{z}, \tau, h_k, h_{k+1}) &= 2 \max(\tau - |h_k| - |h_{k+1}|, 0) \\ &\quad - \max(\tau - \max(|h_k|, |h_{k+1}|), 0) \\ &\quad - \max(\tau - |h_k| - |h_{k+1}| + 1, 0) \\ &\quad + \frac{1 - r^{1 - \min(|h_k|, |h_{k+1}|)}(\mathbf{z})}{1 - r(\mathbf{z})} \end{aligned} \quad (23)$$

in which  $r(\mathbf{z}) = \exp \{-z_2 + \sqrt{z_1 z_2} + \sqrt{z_2 z_3} - \sqrt{z_1 z_3}\}$ . Using these functions, we now give the expressions of the matrix blocks composing  $\mathbf{V}$  and  $\mathbf{P}$ , i.e.,  $\mathbf{V}_{22}$ ,  $\mathbf{P}_{11}$ ,  $\mathbf{P}_{12}$ ,  $\mathbf{P}_{21}$  and  $\mathbf{P}_{22}$  which were introduced in Section 3.

After plugging (17) into (11) and computing the expectations, we obtain, for  $k = 1, \dots, K$

$$[\mathbf{V}_{22}]_{k,k} = -h_k u(\tau, h_k) \Phi(h_k). \quad (24)$$

The expression of  $\mathbf{P}_{11}$  is obtained by substituting (17) into (14), which leads to a diagonal matrix whose elements have the following form (for  $\alpha_k > 0$ )

$$[\mathbf{P}_{11}]_{k,k} = \frac{\beta(\tau + 1)}{2(\alpha_k - 1)} + \frac{\beta^2}{\alpha_k - 2}. \quad (25)$$

Similarly, the expressions of  $\mathbf{P}_{12}$  (of size  $(K+1) \times K$ ) and  $\mathbf{P}_{21}$  (of size  $K \times (K+1)$ ) can be obtained after plugging (17) into (15), which leads to

$$\mathbf{P}_{12} = \mathbf{P}_{21}^T = \begin{bmatrix} A_{1,1} & 0 & \cdots & 0 \\ A_{2,1} & A_{2,2} & \ddots & \vdots \\ 0 & A_{3,2} & \ddots & 0 \\ \vdots & \ddots & \ddots & A_{K,K} \\ 0 & \cdots & 0 & A_{K+1,K} \end{bmatrix} \quad (26)$$

where, for  $k = 1, \dots, K$ , and  $l = k$  or  $l = k-1$ ,

$$A_{k,l} = \pm h_l u(\tau, h_l) \frac{\beta^{\alpha_l + \alpha_{l+1}}}{\Gamma(\alpha_l) \Gamma(\alpha_{l+1})} \int_{\mathbb{R}_+^2} \left[ \sqrt{\frac{y_1}{y_2}}^{\pm 1} - 1 \right] \varphi_{h_l}(\mathbf{y}) d\mathbf{y} \quad (27)$$

in which both “ $\pm$ ” signs are “ $+$ ” signs if  $l = k-1$ , and they are “ $-$ ” signs if  $l = k$ .

Finally, by substituting (17) into (16), the matrix  $\mathbf{P}_{22}$  can be written as the following symmetric tridiagonal matrix

$$\mathbf{P}_{22} = \begin{bmatrix} B_1 & C_1 & 0 & \cdots & 0 \\ C_1 & B_2 & C_2 & \ddots & \vdots \\ 0 & C_2 & B_3 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & C_{K-1} \\ 0 & \cdots & 0 & C_{K-1} & B_K \end{bmatrix} \quad (28)$$

where, for  $k = 1, \dots, K$ ,  $B_k = 2(u(\tau, h_k) - u(\tau, 2h_k) \Phi(2h_k))$ , and for  $k = 1, \dots, K-1$ ,

$$C_k = v(\tau, h_k, h_{k+1}) \frac{\beta^{\alpha_k + \alpha_{k+1} + \alpha_{k+2}}}{\Gamma(\alpha_k) \Gamma(\alpha_{k+1}) \Gamma(\alpha_{k+2})} \times \int_{\mathbb{R}_+^3} \phi_{h_k, h_{k+1}}(\mathbf{z}) w(\mathbf{z}, \tau, h_k, h_{k+1}) d\mathbf{z}. \quad (29)$$

#### • Practical computation of the bound:

The lower bound given by the right-hand side of (6), that we will denote by  $\mathbf{R}$  can be computed using the previous formulas, i.e., from (24) to (29). It can be noticed that some integrals (in (24), (27) and (29)), do not have any closed-form expression requiring some numerical scheme for their computation. In this paper, we have used the adaptive quadrature method [25] that proved efficient for our computations.

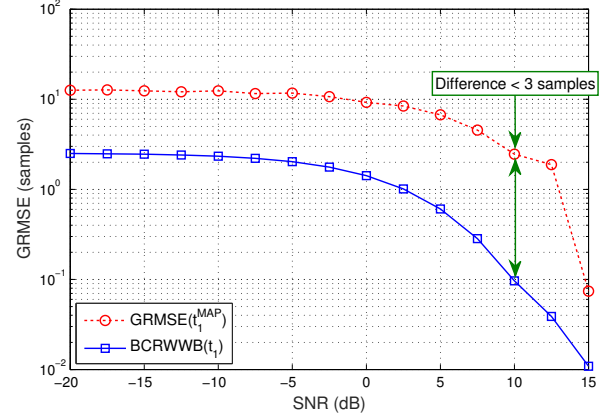
In addition, we would like to stress that, even if it does not appear explicitly with the adopted notations, the matrix  $\mathbf{R}$  actually depends on the parameters  $\alpha_1, \dots, \alpha_{K+1}, \beta, \tau, h_1, \dots, h_K$  (only the dependency on  $h_1, \dots, h_K$  has been mentioned from (18) to (23)). Since each vector  $\mathbf{h} = (h_1, \dots, h_K)$  leads to a lower bound  $\mathbf{R}(\mathbf{h})$ , one obtains a finite set of lower bounds  $\mathcal{W} = \{\mathbf{R}(\mathbf{h}) \mid \mathbf{h} \in \mathcal{H}\}$ , in which  $\mathcal{H}$  is the set of all possible values of  $\mathbf{h}$ . As already mentioned, the proper Cramér-Rao/Weiss-Weinstein lower bound is the tightest value of  $\mathbf{R}(\mathbf{h})$ , namely the supremum of  $\mathcal{W}$ , that we denote by  $\mathbf{B} = \sup(\mathcal{W}) = \sup_{h_1, \dots, h_K} \mathbf{R}(\mathbf{h})$ . The supremum operation has to be taken w.r.t. the Loewner partial ordering, denoted by “ $\preceq$ ” [26]. This ordering implies that a unique supremum in the finite set  $\mathcal{W}$  might not exist. However, it is possible to approximate this supremum by computing a minimal upper-bound  $\mathbf{B}^*$  of the set  $\mathcal{W}$ : this bound is such that, for all  $\mathbf{h} \in \mathcal{H}$ ,  $\mathbf{B}^* \succeq \mathbf{R}(\mathbf{h})$ , and there is no smaller matrix  $\mathbf{B}' \preceq \mathbf{B}^*$  that also verifies  $\mathbf{B}' \succeq \mathbf{R}(\mathbf{h}), \forall \mathbf{h} \in \mathcal{H}$ . It has been shown in [16, 27] that finding  $\mathbf{B}^*$  is equivalent to finding the minimum volume hyper-ellipsoid  $\varepsilon(\mathbf{B}^*) = \{\mathbf{x} \in \mathbb{R}^K \mid \mathbf{x}^T \mathbf{B}^* \mathbf{x} \leq 1\}$  that covers the union of hyper-ellipsoids  $\varepsilon(\mathbf{R}(\mathbf{h})) = \{\mathbf{x} \in \mathbb{R}^K \mid \mathbf{x}^T \mathbf{R}(\mathbf{h}) \mathbf{x} \leq 1\}$ . The search of this ellipsoid can actually be formulated as the following convex optimization problem [27]:

$$\begin{aligned} & \text{minimize} \quad \log(\det(\mathbf{B}^{1/2})) \\ & \text{subject to} \quad \begin{cases} b_1 \geq 0, b_2 \geq 0, \dots, b_{N_h} \geq 0, \\ \begin{bmatrix} \mathbf{B}^{-1} - \mathbf{b}_n(\mathbf{R}(\mathbf{h})_n)^{-1} & \mathbf{0}_{(2K+1) \times 1} \\ \mathbf{0}_{1 \times (2K+1)} & b_n - 1 \end{bmatrix} \preceq \mathbf{0}_{2K+2} \\ (n = 1, \dots, N_h) \end{cases} \end{aligned} \quad (30)$$

in which  $N_h$  denotes the number of elements of the set  $\mathcal{H}$ , and  $\mathbf{R}(\mathbf{h})_n \in \mathcal{W}$  is an indexed version of  $\mathbf{R}(\mathbf{h})$  (i.e., when  $n$  varies from 1 to  $N_h$ ,  $\mathbf{h}$  runs through all the possible combinations of  $h_1, \dots, h_K$ , and  $\mathbf{R}(\mathbf{h})_n$  runs through all the elements of  $\mathcal{W}$ ). The problem (30) can be solved efficiently using a semidefinite programming tool, such as the one provided in the CVX package [28].

### 5. NUMERICAL RESULTS

This section analyzes the evolution of the proposed bound as a function of a parameter that is classically used for changepoint estimation performance. This parameter is either referred to as “amount of change” [29], “magnitude of change” or “signal-to-noise ratio” (SNR): in [15, 16, 18], for Poisson distributed data, the SNR is defined as  $\nu = (\lambda_{k+1} - \lambda_k)^2 / \lambda_k^2$ , for  $k = 1, \dots, K+1$ . In our context, since each  $\lambda_k$  is a random variable with a gamma distribution of parameters  $\alpha_k$  and  $\beta$  (as stated in (2)), this leads to a lower



**Fig. 1.** Estimated GRMSE and proposed lower bound w.r.t. the change-point  $t_1$ , versus SNR, with  $T = 80$  snapshots and  $K = 1$  change in the mean rate of a Poisson time series.

bound  $\mathbf{B}$  that does not depend on  $\lambda_1, \dots, \lambda_{K+1}$ , and *a fortiori* on  $\nu$ . However, the bound depends upon the parameters  $\alpha_k$  and  $\beta$ , which can then be used to drive the average  $(\lambda_k)_{\text{mean}} = \alpha_k / \beta$  generated by the Gamma prior. Thus, by substituting  $(\lambda_k)_{\text{mean}}$  with  $\lambda_k$  in the definition of  $\nu$ , we obtain  $\bar{\nu} = (\alpha_{k+1} - \alpha_k)^2 / \alpha_k^2$ . Such a definition implies that the higher  $\bar{\nu}$ , the higher the amount of change between two consecutive segments, *on average*.

In this study, we present some simulation results obtained for  $T = 80$  observations,  $K = 1$  change, and with  $\bar{\nu}$  ranging from  $-20$  dB to  $15$  dB. Such a choice for the value of  $K$  is justified by the fact that it yields less complex expressions of the estimators given in (31) and (32). We chose  $\alpha_1 = 3$ , and the subsequent  $\alpha_2$  is given by  $\alpha_2 = \alpha_1(1 + \sqrt{\bar{\nu}})$ . We compare the proposed bound with the estimated global mean square error (GMSE) of the maximum a posteriori (MAP) estimator of  $\boldsymbol{\theta} = [\lambda_1, \lambda_2, t_1]^T$ . It is worth mentioning that, given the posterior density  $f(\boldsymbol{\lambda}, \mathbf{t} | \mathbf{x} = \boldsymbol{\kappa})$  (that is proportional to (17)), there is a closed-form expression of the MAP estimator of  $\boldsymbol{\lambda}$ , for a given  $\ell$ , that is, for  $k = 1, \dots, K+1$ :

$$\hat{\lambda}_k^{\text{MAP}}(\ell_{k-1}, \ell_k) = \frac{\alpha_k + (\sum_{t=\ell_{k-1}+1}^{\ell_k} \kappa_t) - 1}{\beta + (\ell_k - \ell_{k-1})}. \quad (31)$$

This closed form expression is then used to obtain the MAP estimator of  $\mathbf{t}$

$$\hat{\mathbf{t}}^{\text{MAP}} = \arg \max_{\ell} \ln f(\hat{\boldsymbol{\lambda}}^{\text{MAP}}(\ell), \mathbf{t} = \ell | \mathbf{x} = \boldsymbol{\kappa}). \quad (32)$$

The estimated global root mean square error (GRMSE) of  $\hat{\mathbf{t}}^{\text{MAP}}$  computed using 1000 Monte-Carlo runs and the associated lower bound are compared in Fig. 1. Even if there exists a gap between the GRMSE and the bound, the difference decreases as  $\bar{\nu}$  increases: at  $\bar{\nu} = 10$  dB, the difference in terms of number of samples is not more than 3 samples; at  $\bar{\nu} = 15$  dB, it is less than 0.1 samples. This gap is actually due to the discrete nature of the change-point location parameter  $t_1$ : when the estimator has an error of one sample, this corresponds to an error of  $T_s$  (that is the signal sampling period). Thus, it is usual to obtain such a gap in this context. The MAP behavior even seems to be closer to the bound for  $\bar{\nu} \geq 15$  dB. However, it could not be displayed for numerical reasons, the GRMSE tending steeply to zero. Finally, the derived bound provides a fair approximation of the change-point estimation behavior, in this context of Poisson data when the Poisson parameters  $\lambda_k$  are unknown.

## 6. REFERENCES

- [1] D. V. Hinkley, "Inference about the change-point in a sequence of random variables," *Biometrika*, vol. 57, no. 1, pp. 1–18, 1970.
- [2] D. Picard, "Testing and estimating change-points in time series," *Advances in Applied Probability*, vol. 17, no. 4, pp. 841–867, Dec. 1985.
- [3] J. Chen and A. K. Gupta, *Parametric Statistical Change Point Analysis*. Birkhäuser Basel, 2000.
- [4] A. E. Raftery, "Bayesian analysis of a poisson process with a change-point," vol. 73, no. 1, pp. 85–89, 1986.
- [5] M. Brown, "Detecting changes in a Poisson process monitored at random time instants," *Sequential Analysis*, vol. 35, no. 3, pp. 358–369, 2016.
- [6] J. D. Scargle, "Studies in astronomical time series analysis: V. bayesian blocks, a new method to analyze structure in photon counting data," *Astrophysical Journal*, vol. 504, pp. 405–418, Sep. 1998.
- [7] B. Jackson, J. S. Scargle, D. Barnes, S. Arabhi, A. Alt, P. Gioumousis, E. Gwin, P. Sangtrakulcharoen, L. Tan, and T. T. Tsai, "An algorithm for optimal partitioning of data on an interval," *IEEE Signal Process. Lett.*, vol. 12, pp. 105–108, Feb. 2005.
- [8] N. Dobigeon, J.-Y. Tournet, and J. D. Scargle, "Joint segmentation of multivariate astronomical time series: Bayesian sampling with a hierarchical model," *IEEE Trans. Signal Process.*, vol. 55, no. 2, pp. 414–423, Feb. 2007.
- [9] M. Brown, "Bayesian detection of changes of a Poisson process monitored at discrete time points where the arrival times are unknown," *Sequential Analysis*, vol. 27, no. 1, pp. 68–77, Feb. 2008.
- [10] P. M. Djurić, "A MAP solution to off-line segmentation of signals," in *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 4, Apr. 1994, pp. 505–508.
- [11] M. Lavielle, "Optimal segmentation of random processes," *IEEE Trans. Signal Process.*, vol. 46, no. 5, pp. 1365–1373, May 1998.
- [12] J.-Y. Tournet, M. Doisy, and M. Lavielle, "Bayesian off-line detection of multiple change-points corrupted by multiplicative noise: application to SAR image edge detection," *ELSEVIER Signal Processing*, vol. 83, no. 9, pp. 1871–1887, Sep. 2003.
- [13] E. Punskeya, C. Andrieu, A. Doucet, and W. Fitzgerald, "Bayesian curve fitting using MCMC with applications to signal segmentation," *IEEE Trans. Signal Process.*, vol. 50, no. 3, pp. 747–758, Mar. 2002.
- [14] P. Fearnhead, "Exact Bayesian curve fitting and signal segmentation," *IEEE Trans. Signal Process.*, vol. 53, no. 6, pp. 2160–2166, Jun. 2005.
- [15] A. Ferrari and J.-Y. Tournet, "Barankin lower bound for change points in independent sequences," in *Proc. of IEEE Workshop on Statistical Signal Processing (SSP)*, St. Louis, MO, USA, Sep. 2003, pp. 557–560.
- [16] P. S. La Rosa, A. Renaux, A. Nehorai, and C. H. Muravchik, "Barankin-type lower bound on multiple change-point estimation," *IEEE Trans. Signal Process.*, vol. 58, no. 11, pp. 5534–5549, Nov. 2010.
- [17] J.-Y. Tournet, A. Ferrari, and A. Swami, "Cramér-rao lower bounds for change-points in additive and multiplicative noise," *ELSEVIER Signal Processing*, vol. 84, no. 7, pp. 1071–1088, Jul. 2004.
- [18] L. Bacharach, A. Renaux, M. N. El Korso, and E. Chaumette, "Weiss-Weinstein bound for change-point estimation," in *Proc. of IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, Cancún, Mexico, Dec. 2015, pp. 477–480.
- [19] —, "Weiss-Weinstein bound on multiple change-points estimation," (submitted to *IEEE Trans. Signal Process.*), Aug. 2016.
- [20] H. L. Van Trees and K. L. Bell, Eds., *Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking*. New-York, NY, USA: Wiley/IEEE Press, Sep. 2007.
- [21] A. Renaux, P. Forster, P. Larzabal, C. D. Richmond, and A. Nehorai, "A fresh look at the Bayesian bounds of the Weiss-Weinstein family," *IEEE Trans. Signal Process.*, vol. 56, no. 11, pp. 5334–5352, Nov. 2008.
- [22] K. L. Bell and H. L. Van Trees, "Combined Cramér-Rao/Weiss-Weinstein bound for tracking target bearing," in *Proc. of IEEE Workshop on Sensor Array and Multi-channel Processing (SAM)*, Waltham, MA, USA, Jul. 2006, pp. 273–277.
- [23] E. Weinstein and A. J. Weiss, "A general class of lower bounds in parameter estimation," *IEEE Trans. Inf. Theory*, vol. 34, no. 2, pp. 338–342, Mar. 1988.
- [24] L. Bacharach, M. N. El Korso, A. Renaux, and J.-Y. Tournet, "A Bayesian Lower Bound for Parameter Estimation of Poisson Data Including Multiple Changes (extended)," *ArXiv e-prints*, Sep. 2016, arXiv:1609.03423.
- [25] L. F. Shampine, "MATLAB Program for Quadrature in 2D," *Applied Mathematics and Computation*, vol. 202, no. 1, pp. 266–274, Aug. 2008.
- [26] J. Borwein and A. Lewis, *Convex Analysis and Nonlinear Optimization*. 1st ed. New York: Springer-Verlag, 2000.
- [27] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [28] M. Grant and S. Boyd, *CVX: Matlab software for disciplined convex programming*. Dec. 2008 [Online]. Available: <http://stanford.edu/boyd/cvx>, Web page and software.
- [29] S. B. Fotopoulos, S. K. Jandhyala, and E. Khapalova, "Exact asymptotic distribution of change-point MLE for change in the mean of Gaussian sequences," *The Annals of Applied Statistics*, vol. 4, no. 2, pp. 1081–1104, Nov. 2010.