EFFICIENT BAYESIAN DETECTION OF MULTIPLE EVENTS WITH A MINIMUM-DISTANCE CONSTRAINT

Georg Kail, Franz Hlawatsch, and Clemens Novak

Institute of Communications and Radio-Frequency Engineering, Vienna University of Technology
Gusshausstrasse 25/389, A-1040 Vienna, Austria; e-mail: gkail@nt.tuwien.ac.at

ABSTRACT

We propose a Bayesian method for detecting multiple events in signals under the practically relevant assumption that successive events may not be arbitrarily close and distant events are effectively independent. Our detector has low complexity since it involves only the (Monte Carlo approximation to the) one-dimensional marginal posteriors. However, its performance is good since the metric it minimizes depends on the entire event sequence. We also describe an efficient sequential implementation of our detector that is based on a tree representation and a recursive metric computation.

Index Terms— Event detection, pulse detection, Bayesian analysis, Monte Carlo method.

1. INTRODUCTION

In many applications, the signal of interest is a noisy version of some signal of interest that depends on certain event positions \( k_i \) and their number \( I \) are unknown. We will describe the event positions \( k_i \) by the binary “event indicator” sequence \( b_k, k = 1, \ldots, K \), which is 1 at all event positions (i.e., if \( k = k_i \) for some \( i \)) and 0 for all other \( k \). Let \( x \) and \( b \) denote the length-\( K \) vectors corresponding to the sequences \( x_k \) and \( b_k \), respectively.

Given an observed realization of \( x \), we would like to know how many events there are and at what positions they occur, for a prescribed minimum distance \( d_{\text{min}} \) in (1). We formulate this problem as the minimum-distance constrained detection of the event indicator vector \( b \) from \( x \). Adopting a Bayesian setting, we model \( b \) as a random variable that represents the prior probability mass function (pmf) \( p(b) \) and \( p(b|x) \), respectively. The detection of \( b \) from \( x \) will then be based in some way on the posterior \( p(b|x) \). We do not assume any specific structure of \( p(b|x) \) beyond (1) and the condition that temporarily distant \( b_k \) are only weakly dependent.

From Bayes’ theorem, \( p(b|x) \approx p(x|b) p(b) \). The likelihood function \( p(x|b) \) describes how the event positions \( k_i \) (characterized by \( b \)) affect the observed signal \( x \); it depends on the specific signal model considered and some statistical assumptions. The prior \( p(b) \) provides a way to incorporate the minimum-distance constraint (1). Let \( C \subset \{0,1\}^K \) be the set of all \( b \) satisfying (1). Then, we define \( p(b) \propto p(b) \) for \( b \in C \) and \( p(b) = 0 \) otherwise, where \( p(b) \) is some auxiliary pmf (e.g., independent and identically distributed). Note that setting the prior \( p(b) \) to zero for \( b \not\in C \) also forces the posterior \( p(b|x) \) to be zero for \( b \not\in C \). Note also that the \( b_k \) are strongly dependent for times on the order of \( d_{\text{min}} \). In fact, the minimum-distance constraint (1) postulates that if \( b_k = 1 \) for a given \( k \), then \( b_{k+d} = 0 \) for all \( d \) such that \( |d| \leq d_{\text{min}} - 1 \).

Classical MAP detectors. An optimal method for detecting the event indicator sequence \( b \) is the MAP sequence detector \[ b_{\text{MAP}}(x) \triangleq \arg \max_{b \in C} p(b|x), \] which minimizes the sequence error probability \( P(b \neq b) \). Unfortunately, this detector is of limited practical usefulness since finding the maximum of the typically high-dimensional posterior \( p(b|x) \) over the typically large set of hypotheses \( C \) is often infeasible.

An alternative optimal method is the MAP component detector, also known as “maximum posterior marginal/mode (MPM) detector” (e.g., [7]), which is defined as \[ b_{\text{MAP}}(x) \triangleq \arg \max_{b_k \in \{0,1\}} p(b_k|x), \quad k = 1, \ldots, K, \] This detector minimizes the component error probability \( P(b_k \neq b_k) \). However, in regarding only one marginal at a time, the MAP component detector ignores a significant part of the information contained in the joint posterior \( p(b|x) \), which may result in counterintuitive results. In particular, consider a fixed time interval \( K \) of length \( |K| \leq d_{\text{min}} \), and suppose that for a given observed \( x \), the probability that there is exactly one event in \( K \) is 1. Equivalently, \( \sum_{b \in C} p(b_k = 1|x) = 1 \) since the events \( b_k = 1 \) are mutually exclusive for \( k \in K \) (there may not be more than one event in \( K \) because of the minimum-distance constraint). Suppose further that none of the possible event positions \( k \in K \) is much more likely than the

This work was supported by the FWF project “Statistical Inference” (S10603-N13) within the National Research Network SISE.
others, so \( p(b_k = 1|x) < 1/2 \) for all \( k \in K \). It then follows that \( b_{MAP,k} = 0 \) for all \( k \in K \), i.e., the MAP component detector does not detect any event in \( K \). This is clearly counterintuitive as the probability that there is no event in \( K \) is zero.

**Monte-Carlo approximations.** A Monte-Carlo (MC) approximation \([8]\) often renders difficult calculations feasible (e.g., \([1–5, 7]\)). However, we will now argue that the above MAP detectors have practical limitations even in that case. The MC approximation is based on a sample \( S \triangleq \{b^{(m)}\}_{m=1}^{M} \) consisting of \( M \) realizations \( b^{(m)} \) of \( b \) that are drawn from the joint posterior \( p(b|x) \) by means of some sampling method such as the Gibbs sampler \([8]\). Let \( q(b) \) denote the relative multiplicity of a given \( b \in C \) in \( S \), i.e., the number of occurrences of \( b \in S \) normalized by the sample size \( |S| = M \). Then \( q(b) \) converges to the joint posterior \( p(b|x) \) as \( M \) increases \([8]\). Similarly, let \( q(b_k) \) denote the relative multiplicity of \( b_k \), i.e., the number of realizations of \( b \) in \( S \) that have the given \( b_k \in \{0, 1\} \) at position \( k \), normalized by the sample size \( M \). Then \( q(b_k) \) converges to the marginal posterior \( p(b_k|x) \) as \( M \) increases.

The MC approximation to the MAP sequence detector \( b_{MAP}(x) \), denoted \( b_{MC}(x) \), is obtained by replacing \( p(b|x) \) with \( q(b) \) in (2). In the practically relevant case of moderate sample size \( M \ll |C| \), the maximization in (2) may now be feasible, since the detector chooses only among the \( b \) contained in \( S \). However, replacing \( p(b|x) \) by \( q(b) \) is problematic, since the number of \( b \) different \( b \) in \( S \) (which is at most equal to \( |S| = M \)) is much smaller than the number \( |C| \) of hypotheses. This means that most \( b \in C \) must be expected to occur only once or not at all in \( S \), and thus \( q(b) \) will be quite different from the true posterior \( p(b|x) \). In particular, a small sample size may introduce artificial dependencies of components that are actually independent. This is undesired, especially since we assumed weak dependencies between distant events.

The MC version of the MAP component detector \( b_{MAP,k}(x) \), denoted \( b_{MC,k}(x) \), is obtained by replacing \( p(b_k|x) \) with \( q(b_k) \) in (3). Here, the limited sample size is no problem since there are only two possible hypotheses for \( b_k \). Hence, \( q(b_k) \) will be a good approximation to \( p(b_k|x) \). However, \( b_{MC,k}(x) \) inherits the basic problem of \( b_{MAP,k}(x) \), namely, that it ignores a significant part of the information contained in the joint posterior.

In \([3]\), we considered a MAP block detector as a compromise between the MAP sequence and component detectors. This detector detects nonoverlapping blocks of \( b \) separately. This is justified if \( b \) consists of blocks of moderate length that are effectively independent of each other. However, if this is not the case, the MAP block detector will perform poorly since, similarly to the MAP component detector, it ignores statistical dependencies.

### 3. THE PROPOSED DETECTOR

We now present a detector that does not suffer from the problems described above. The new detector is based on the marginal pmf \( p(b_k|x) \). However, if calculation of \( p(b_k|x) \) by marginalization of \( p(b|x) \) is infeasible, the MC approximation \( q(b_k) \) may be used instead of \( p(b_k|x) \) (as noted above, this approximation is accurate even for a small sample size \( M \)). If \( q(b_k) \) is being used, the detector’s complexity is quite low. At the same time, the detector’s decision metric is based on the entire sequence, which avoids the information loss associated with separate optimization of the components \( b_k \).

Hereafter, for notational simplicity, we write \( p_k \triangleq p(b_k = 1|x) \) for the posterior probability that \( b_k = 1 \). We note that due to \( b_k \in \{0, 1\} \), the posterior expectation of \( b_k \) is \( E(b_k|x) = p_k \), and the number of events in an interval \([k_1, k_2]\) is \( \sum_{k=k_1}^{k_2} b_k \). We will also consider the expected number of events in \([k_1, k_2] \) given \( x \):

\[
Q_{k_1,k_2} \triangleq E \left( \sum_{k=k_1}^{k_2} b_k | x \right) = \sum_{k=k_1}^{k_2} E \left( b_k | x \right) = \sum_{k=k_1}^{k_2} p_k = \sum_{k=k_1}^{k_2} p_k.
\]

Note that \( p_k \) and \( Q_{k_1,k_2} \) depend on \( x \). To avoid double counting at interval borders, we will use the modified quantity

\[
Q_{k_1,k_2}^2 \triangleq Q_{k_1,k_2} - \frac{p_{k_1}}{2} - \frac{p_{k_2}}{2}, \quad 1 \leq k_1 \leq k_2 \leq K,
\]

complemented by the left-end and right-end definitions

\[
Q_{0,k} \triangleq Q_{1,k} - \frac{p_0}{2}, \quad Q_{k,K+1} \triangleq Q_{k,K} - \frac{p_k}{2}.
\]

The new detector is motivated by the following four desiderata.

**D1.** Within any interval \([k_1, k_2]\) for which \( Q_{k_1,k_2} \) is close to 1, one event should be detected.

**D2.** The detected events should be at positions \( k \) where \( p_k \) is large.

**D3.** The detected sequence of events should comply with the minimum-distance constraint (1).

**D4.** The detected events should be at positions \( k \) where \( p_k \) is not smaller than at the adjacent positions \( k-1 \) and \( k+1 \).

These desiderata are reflected either in the decision metric or by a constraint, as explained next.

**D1.** Suppose that a given detector detects \( N \) events at positions \( l_n, n = 1, \ldots, N \) (hereafter briefly called “detected positions”). Then, based on the cumulative structure of \( Q_{l_n-1, l_n} \), the following fact can be verified: If there is roughly one detected event within any interval \([k_1, k_2]\) such that \( Q_{k_1,k_2} \approx 1 \), then \( Q_{l_n-1, l_n} \approx 1 \) for all \( n \in [2, N] \), and vice versa. Thus, **D1** can be rephrased as follows: Between any two successive detected positions \( l_{n-1}, l_n \), the expected number of events \( Q_{l_{n-1}, l_n} \) should be close to 1. The deviation of \( Q_{l_{n-1}, l_n} \) from 1 will be measured by \( \sum_k |Q_{l_{n-1}, l_n} - 1| \). Thus, **D1** corresponds to choosing the detected \( b \) (or, equivalently, \( N \) and the \( l_n \)) such that the following metric component is minimized:

\[
w_1(b) \triangleq \sum_{n=2}^N \sum_k |Q_{l_{n-1}, l_n} - 1| + w_1^{(1)}(l_1) + w_1^{(R)}(l_N).
\]

Here, the left-end term \( w_1^{(L)}(l_1) \) and the right-end term \( w_1^{(R)}(l_N) \) express our desire that the expected numbers of events to the left of \( l_1 \) and to the right of \( l_N \) are both close to or less than 1:

\[
w_1^{(L)}(l_1) \triangleq |Q_{0,l_1} - 1|, \quad w_1^{(R)}(l_N) \triangleq |Q_{l_N,K+1} - 1|, \quad (4)
\]

where \([\cdot]^{+}\) means that a negative value is replaced by zero. We note that **D1**—i.e., minimization of \( w_1(b) \) alone—would essentially define the relative positions of the detected events with respect to each other. Fixing one detected event position, e.g. \( l_1 \), determines all the others. However, the value of \( l_1 \) obtained by minimizing \( w_1(b) \) is rather arbitrary within some set of almost equivalent solutions.

**D2.** The absolute position of the detected events is strongly influenced by **D2**. We will formulate **D2** as the maximization of \( \sum_{n=1}^N p_{l_n} \) or, equivalently, minimization of the metric component

\[
w_2(b) \triangleq \sum_{n=1}^N (1 - p_{l_n}).
\]

**D3.** Rather than enforcing **D3**, i.e., \( b \in C \) by explicitly restricting the set of hypotheses, we will use the metric component (cf. \([9]\))

\[
w_3(b) \triangleq \begin{cases} 0, & b \in C, \\ \infty, & \text{otherwise.} \end{cases}
\]

\(1\)Our formulation of this equivalence is fuzzy only because of the discrete nature of \( k \). Replacing \( k \) by a continuous variable and, consequently, \( Q' \) by an integral would make the equivalence exact.
These path segments constitute a subtree that originates from a path node \( b \). The tree can be pruned, based on a left-right order. All hypotheses \( b \) such that all detected positions \( l_b \) are contained in \( K \). This reduces the number of hypotheses from \( 2^K \) to \( 2^J \). However, the actual constraint set (also taking into account D3) is \( W \cap C \), and thus still smaller. This set is nonempty, since it always contains the all-zero sequence \( b = 0 \).

**Definition of the detector.** Putting together D1 through D4, the proposed detector is finally defined via the constrained minimization

\[
\hat{b}(x) \triangleq \arg \min_{b \in \mathcal{W}} \{ w(b) \},
\]

with

\[
w(b) \triangleq w_1(b) + \gamma w_2(b) + w_3(b)
\]

\[
= \sum_{n=2}^{N} (|Q_{n-1}^l| - 1) + w_1^L(l_1) + w_1^R(l_N) + \gamma \sum_{n=1}^{N} (1 - p_{n+1}) + \sum_{n=2}^{N} \tilde{w}_3(l_{n-1}, l_n)
\]

where \( \tilde{w}_3(k_1, k_2) \triangleq 0 \) for \( |k_1 - k_2| \geq d_{\min} \) and \( \tilde{w}_3(k_1, k_2) \triangleq \infty \) otherwise.

**D4.** To reduce the number of hypotheses to be compared, D4 is formulated as a hard constraint \( b \in \mathcal{W} \) rather than via the metric. Let \( K = \{ k_j \}_{j=1}^{J} \) denote the set of all \( k \) such that \( p_k \geq p_{k-1} \) and \( p_k \geq p_{k+1} \). Then, \( W \) is defined as the set of all hypotheses \( b \) such that all detected positions \( l_b \) are contained in \( K \). This reduces the number of hypotheses from \( 2^K \) to \( 2^J \). However, the actual constraint set (also taking into account D3) is \( W \cap C \), and thus still smaller. This set is nonempty, since it always contains the all-zero sequence \( b = 0 \).

**4. EFFICIENT SEQUENTIAL IMPLEMENTATION**

In (5), \( w(b) \) is minimized over \( W \) or, equivalently, with respect to \( N \) and the position set \( \{ l_n \}_{n=1}^{N} \subseteq \mathcal{K} = \{ k_j \}_{j=1}^{J} \). At each of the \( J \) admissible positions \( k_j \), \( k_j \), \( b_k \) may be 0 or 1. Thus, the constraint set \( W \) consists of \( 2^J \) sequences \( b \), which, typically, is still too much. Fortunately, the structure of the decision metric \( w(b) \) in (6) allows for an efficient implementation that is sequential in \( k_j \).

**Tree representation.** In Fig. 1, \( W \) is represented by a binary tree whose levels (except the root level) correspond to the admissible positions \( k_j \). Thus, there are \( J + 1 \) levels including the root. We assume that the \( k_j \) are indexed in increasing order, i.e., \( k_{j+1} > k_j \), and displayed from left to right. The root and the leaves correspond to levels \( j = 0 \) and \( j = J \), respectively. The two branches leaving a node at level \( j \) correspond to the binary decision \( b_{k_{j+1}} = 0 \) (upward branch) or \( b_{k_{j+1}} = 1 \) (downward branch). At level \( j \), half of the \( 2^j \) nodes correspond to \( b_{k_{j+1}} = 0 \), the other half to \( b_{k_{j+1}} = 1 \); we will refer to them as 0-nodes and 1-nodes, respectively. Each of the \( 2^J \) leaves (equivalently, paths from root to leaf) corresponds to one \( b \in \mathcal{W} \).

**Pruning the tree.** The tree can be pruned, based on a left-right decomposition of the metric \( w(b) \) at 1-nodes. Each 1-node corresponds to a detected position. Consider a fixed 1-node \( N \) at level \( j \), corresponding to the detected position \( l_b = k_j \). Let \( b \) with \( b_{k_{j+1}} = 1 \) correspond to a path that passes through \( N \). This path is uniquely defined to the left of \( N \) (from root to \( N \), but to the right of \( N \) it may be any of the path segments connecting \( N \) to one of the leaves. These path segments constitute a subtree that originates from \( N \). From (6), it follows that the metric of \( b \) can be decomposed as

\[
w(b) = w_1(l_1, \ldots, l_n) + w_3(l_n, \ldots, l_N),
\]

with the “left partial metric” \( w_1(l_1, \ldots, l_n) \triangleq \sum_{n=2}^{N} |Q_{n-1}^l(l_n) - 1| + w_1^L(l_1) + w_1^R(l_N) + \gamma \sum_{n=1}^{N} (1 - p_{n+1}) + \sum_{n=2}^{N} \tilde{w}_3(l_{n-1}, l_n) \) and the “right partial metric” \( w_3(l_n, \ldots, l_N) \triangleq \sum_{n=0}^{N} |Q_{n-1}^r(l_n) - 1| + w_3^R(l_N) + \gamma \sum_{n=1}^{N} (1 - p_{n+1}) + \sum_{n=1}^{N} \tilde{w}_3(l_{n-1}, l_n) \).

Different 1-nodes \( N_m \) at the same level \( j \) correspond to the same detected position \( l_m = k_j \) (note that the index of \( l \) depends on the 1-node), but the paths passing through these 1-nodes are different. The left partial metric (LPM) \( w_1(l_1, \ldots, l_n) \) is generally different for these 1-nodes. However, the set of possible values for the right partial metric (RPM) \( w_3(l_n, \ldots, l_N) \) is the same, since the subtrees originating from the 1-nodes are equivalent. Because of (7), this means that only the 1-node with the smallest LPM \( w_1(l_1, \ldots, l_n) \) can be part of the minimum-metric path, and thus all the other 1-nodes at level \( j \) and the entire subtrees originating from them can be discarded. (A similar consideration does not apply to the 0-nodes because these do not correspond to a detected position \( l_m \), and hence no LPMs and RPMs can be associated with them.) The resulting pruning procedure yields a dramatic reduction of the number of paths (see Fig. 1). Before pruning, level \( j \) contains twice as many nodes as level \( j - 1 \). Out of the 1-nodes, which are half of them, only one is retained after pruning. Thus, passing to the next level, the number of left path segments to be retained is increased by 1, rather than doubled. There are two nodes at level \( j = 1 \), and therefore, at level \( j \), there are \( j + 1 \) surviving nodes (or left path segments), namely, \( j \) 0-nodes and one 1-node. At the final level \( j = J \), the number of surviving leaves (or paths, or hypotheses \( b \)) is only \( J + 1 \).

**Sequential operation.** At each level, we have to compute the LPM \( w_1(l_1, \ldots, l_n) \) for each 1-node. A sequential processing is possible because \( w_1(l_1, \ldots, l_n) \) can be computed recursively according to

\[
w_1(l_1, \ldots, l_n) = w_1(l_1, \ldots, l_{n-1}) + \Delta w_n,
\]

with \( \Delta w_n \triangleq |Q_{n-1}^l(l_n) - 1| + \gamma (1 - p_{n+1}) + \tilde{w}_3(l_{n-1}, l_n) \). This recursion is initialized by \( w_1(l_1) = w_1^L(l_1) + \gamma (1 - p_{n+1}) \).

At level \( j \), we update the LPMs of the \( j \) 1-nodes according to (8) and identify the surviving 1-node with minimum LPM. Processing the entire tree requires \( \sum_{j=1}^{J} J(J+1)/2 \) such updates. At the final level \( J \), we calculate the RPMs for the \( J + 1 \) leaves and add them to the LPMs. Since there are no detected positions to the right of \( k_J \), these RPMs reduce to the respective values of \( w_3^R(l_N) \).

We have thus calculated the final metrics \( w(b) \) of all \( J + 1 \) paths (or sequences \( b \)) that still qualify for the minimum-metric path; these paths constitute a significantly reduced subset of the constraint set \( W \). Finally, we choose the detection result \( b(x) \) as the \( b \) with min-
imimum metric. The total number of hypotheses we processed (the surviving leaves as well as the nodes we compared and discarded along the way) is only $J(J+1)/2 + 1$, rather than $2^J$.

**Complexity-reducing approximation.** Since $j$ metric updates are performed at the $j$th tree level, the computational complexity per level grows linearly with the level index $j$. We now propose a simple approximation that yields a roughly constant complexity per level.

Whereas most 1-nodes are discarded in our algorithm, 0-nodes are retained without checking. However, paths with mostly 0-nodes are likely to have a large metric, because $w_l(b)$ penalizes large zero intervals. Consider a 0-node at level $j$, corresponding to admissible position $k_j$. Since we cannot compute an LPM at 0-nodes, we will define a substitute based on (8), using the same expression up to the last 1-node on the path (corresponding to detected position $l_{n-1}$) but a modified update term because there is no $\Delta_0$ at a 0-node:

$$w_L^{(0)}(l_1, \ldots, l_{n-1}, k_j) \triangleq w_L(l_1, \ldots, l_{n-1}) + [Q'_n, k_j - 1]^+. $$

Compared with $\Delta w_L$ in (8), the new update term does not contain the terms due to $w_L(b)$ and $w_L(\emptyset)$ because there is no event at $k_j$. In the term due to $w_L(b)$, we replaced $l_i$ by $k_j$. However, since instead of an interval between two detected positions we have only part of such an interval, the expected number of events within $[l_{n-1}, k_j]$ should be close to or less than 1, which is expressed by the $\lfloor \cdot \rceil$ operation (note the analogy to (4)).

We now discard all 0-nodes at level $j$ whose LPM substitutes $w_L^{(0)}(l_1, \ldots, l_{n-1}, k_j)$ surpass the LPM of the single surviving 1-node by more than a predefined margin $\Delta_{\text{max}}$. If $\Delta_{\text{max}}$ is large enough (our experiments suggest that $\Delta_{\text{max}} = 1$ is a good choice), this additional pruning is unlikely to affect the detection result. However, the computational complexity per tree level is now roughly constant.

5. NUMERICAL RESULTS

Next, we demonstrate an example where the classical detectors fail whereas the proposed detector yields plausible results. The problem considered is to detect pulse-like events in an OCT (i.e., optical coherence tomography) signal [3] of length $K = 1024$. The observed signal is modeled as [3]

$$x_k = \sum_{i=1}^{l} a_k f_{x-k} + v_k = \sum_{i=1}^{K} (a_i b_i) f_{x-k} + v_k,$$

with a random pulse $f_x$, random weights $a_k$, and white Gaussian noise $v_k$. The priors of the various random variables were chosen as in [3]. The pulse locations $k_i$ (described by $b = (b_i)$) were assumed to satisfy (1) with $d_{\text{max}} = 7$. Fig. 2(a) shows a measured OCT signal $x_k$. Based on $x_k$, a sample of size $M = 450$ was drawn from the posterior $p(b|x)$ by means of a Gibbs sampler (with sufficient burn-in) [8]. The sequence of empirical one-dimensional marginal posteriors $\hat{p}_k \triangleq \hat{p}(b_k = 1)$ resulting from this sample (the MC approximation to $p_k = p(b_k = 1|x_k)$) is shown in Fig. 2(b).

The MC version of the MAP sequence detector chooses the $b$ appearing most often in the sample. In our example, 406, 16, and 4 different realizations appear, respectively, once, twice, and three times in the sample. Hence, there is a tie of four different “optimum” hypotheses $b$, at low empirical probability $q(b) = 3/450$. Clearly, this detector cannot produce reliable results. Also note that our sample comprises only 450 realizations whereas the number of possible realizations $|B|$ can be lower bounded very loosely by $|B|^\lfloor \text{max} \rceil \approx 10^{14}$.

For each $k$, the MC version of the MAP component detector chooses $b_k = 1$ if $q(b_k = 1) > q(b_k = 0)$ or equivalently $\hat{p}_k > 1/2$, and $b_k = 0$ otherwise. Here, $q(b_k)$ is a good approximation to the true marginal posterior $p(b_k|x)$. Since there are 450 realizations in our sample but only two hypotheses for each $b_k$, however, only one event is detected (cf. Fig. 2(b)). This is quite different from the expected number of pulses in the entire sequence, which is $Q_{1, K} = 80.72$. Hence, the detection result is grossly incorrect.

The result obtained with the proposed detector (using $\gamma = 1$) is illustrated by the crosses in Fig. 2(b). For the given sequence $\hat{p}_k$, the number of D4-admissible positions $\hat{k}_1, \ldots, \hat{k}_J$ is $J = 229$, so the number of hypotheses from which the detector chooses is $|W| = 2^{229} \approx 10^{89}$. Pruning the tree reduces the number of hypotheses to $J(J+1)/2 + 1 = 26536$, with no loss in performance. Discarding paths with long zero intervals (using $\Delta_{\text{max}} = 1$) reduces the number of processed hypotheses to 6049, while in our example the detection result is unchanged. The sequence detected by our method contains 62 pulses, which is fairly close to $Q_{1, K} = 80.72$.

6. CONCLUSION

We proposed a Bayesian method and algorithm for detecting sequences of events that are characterized by a minimum distance between neighboring events and weak dependencies between distant events. The detector uses the one-dimensional marginal posterior at each position, while its metric takes into account the entire sequence. The structure of the decision problem and metric allows for an efficient sequential implementation. Numerical results demonstrate the effectiveness of the proposed detector.

7. REFERENCES


