

ACCEPTED VERSION

Jason L. Williams

The best fitting multi-Bernoulli filter

Statistical Signal Processing (SSP), 2014 IEEE Workshop on, 2014 / pp.220-223

©2014 Crown. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Published version at: <http://dx.doi.org/10.1109/SSP.2014.6884615>

PERMISSIONS

<https://www.ieee.org/publications/rights/author-posting-policy.html>

Author Posting of IEEE Copyrighted Papers Online

The IEEE Publication Services & Products Board (PSPB) last revised its Operations Manual Section 8.1.9 on Electronic Information Dissemination (known familiarly as "author posting policy") on 7 December 2012.

PSPB accepted the recommendations of an ad hoc committee, which reviewed the policy that had previously been revised in November 2010. The highlights of the current policy are as follows:

- The policy reaffirms the principle that authors are free to post their own version of their IEEE periodical or conference articles on their personal Web sites, those of their employers, or their funding agencies for the purpose of meeting public availability requirements prescribed by their funding agencies. Authors may post their version of an article as accepted for publication in an IEEE periodical or conference proceedings. Posting of the final PDF, as published by IEEE *Xplore*[®], continues to be prohibited, except for open-access journal articles supported by payment of an article processing charge (APC), whose authors may freely post the final version.
- The policy provides that IEEE periodicals will make available to each author a preprint version of that person's article that includes the Digital Object Identifier, IEEE's copyright notice, and a notice showing the article has been accepted for publication.
- The policy states that authors are allowed to post versions of their articles on approved third-party servers that are operated by not-for-profit organizations. Because IEEE policy provides that authors are free to follow public access mandates of government funding agencies, IEEE authors may follow requirements to deposit their accepted manuscripts in those government repositories.

IEEE distributes accepted versions of journal articles for author posting through the Author Gateway, now used by all journals produced by IEEE Publishing Operations. (Some journals use services from external vendors, and these journals are encouraged to adopt similar services for the convenience of authors.) Authors' versions distributed through the Author Gateway include a live link to articles in IEEE *Xplore*. Most conferences do not use the Author Gateway; authors of conference articles should feel free to post their own version of their articles as accepted for publication by an IEEE conference, with the addition of a copyright notice and a Digital Object Identifier to the version of record in IEEE *Xplore*.

22 June 2020

<http://hdl.handle.net/2440/86055>

THE BEST FITTING MULTI-BERNOULLI FILTER

Jason L. Williams

National Security and ISR Division, Defence Science and Technology Organisation, Australia
and School of Electrical and Electronic Engineering, University of Adelaide, Australia

ABSTRACT

Recent derivations have shown that the full Bayes random finite set filter incorporates a linear combination of multi-Bernoulli distributions. The full filter is intractable as the number of terms in the linear combination grows exponentially with the number of targets; this is the problem of data association. A highly desirable approximation would be to find the multi-Bernoulli distribution that is closest to the full distribution in some sense, such as the set Kullback-Leibler divergence. This paper proposes an approximate method for achieving this, which can be interpreted as an application of the well-known expectation-maximisation (EM) algorithm.

1. INTRODUCTION

Recently, parallel derivations have found conjugate prior forms for target tracking using unlabelled random finite sets (RFSs) [1, 2] and labelled RFSs [3]. Both cases assume that each measurement is the consequence of at most one target and each target gives rise to at most one measurement, and [2] assumes that false alarms follow a Poisson point process. In each case, the form of the exact filter incorporates a linear combination of multi-Bernoulli (MB) distributions,¹ which can be expressed as:

$$f(X) = \sum_{a \in \mathcal{A}} w_a \sum_{\biguplus_{i=1}^N X_i = X} \prod_{i=1}^N f_{i|a}(X_i) \quad (1)$$

where the notation $\biguplus_{i=1}^N X_i = X$ denotes that the sum is over all disjoint subsets X_1, \dots, X_N whose union is X . We assume that $w_a \geq 0$ and $\sum_{a \in \mathcal{A}} w_a = 1$. The Bernoulli distribution $f_{i|a}(X_i)$ is given by:

$$f_{i|a}(X_i) = \begin{cases} 1 - r_{i|a}, & X_i = \emptyset \\ r_{i|a} f_{i|a}(x_i), & X_i = \{x_i\} \\ 0, & |X_i| > 1 \end{cases} \quad (2)$$

The complexity of exact methods is problematic as the number of terms in the sum over $a \in \mathcal{A}$ in (1) grows exponentially

¹The form in [2] is the union of a Poisson distribution (due to the use of a Poisson model of target birth) and a linear combination of MB distributions, but the latter component is the source of computational difficulty.

in both the number of targets and in time. This is the problem of data association, and the terms $a \in \mathcal{A}$ represent different hypotheses for the correspondence between measurements and Bernoulli components. This was addressed in [2] by seeking a MB approximation, which in effect collapses the sum in (1) back to $|\mathcal{A}| = 1$ after each time step. The two methods proposed directly approximate the probability distribution of data association, $p(a) = w_a$, arriving at two approaches that are closely related to joint probability data association (JPDA) and joint integrated PDA (JIPDA) [4, 5], and the MeMber filter [6, 7]. These approaches each have disadvantages: the former is near-optimal when targets are well-spaced² but fails when targets are closely spaced (referred to as coalescence, e.g., [8]), while the latter is more robust, but exhibits lower performance when targets are well-spaced.

A compelling alternative would be to find the MB distribution which minimises the set Kullback-Leibler (KL) divergence from the exact distribution: [6, p513]

$$D(f||g) = \int f(X) \log \frac{f(X)}{g(X)} \delta X \quad (3)$$

This paper presents an approximate method of finding the MB distribution that minimises the set KL divergence. The resulting algorithm is related to set JPDA (SJPDA) [9] (as explored further in [10]), but utilises RFS to accommodate an arbitrary and uncertain number of targets.

2. BEST-FITTING MULTI-BERNOULLI FILTER

The proposed filter can be viewed as consisting of three steps: prediction, update, and approximation. The prediction and update steps are described in detail in [2], and are omitted here due to space limitations. The form of the MB component of the distribution is preserved by prediction, but after update the distribution incorporates a component of the form (1), which we seek to approximate by a MB distribution. The approach proposed for performing this final step is described in problem 1. We refer to the resulting algorithm (incorporating prediction, update and approximation) as the best fitting multi-Bernoulli (BFMB) filter.

²i.e., cases in which each measurement could only feasibly result from (at most) one Bernoulli component.

Problem 1. Find the MB distribution $g(X)$ that minimises the KL divergence

$$\operatorname{argmin}_{[g_j]} \int f(X) \log \frac{f(X)}{g(X)} \delta X = \operatorname{argmax}_{[g_j]} \int f(X) \log g(X) \delta X \quad (4)$$

where $f(X)$ is of the form (1), and $g(X)$ is MB:

$$g(X) = \sum_{\uplus_{j=1}^N X_j = X} \prod_{j=1}^N g_j(X_j) \quad (5)$$

and the components $g_j(X_j)$ are similar in form to (2).

The RHS of (4) is obtained by separating the log of the quotient into the difference of the logs, and observing that the first term is constant WRT the variables of minimisation.

The first difficulty that we address is the form of the set integral: [6, p361]

$$\int f(X) \log g(X) \delta X \triangleq f(\emptyset) \log g(\emptyset) + \sum_{n=1}^{\infty} \frac{1}{n!} \int \cdots \int f(\{x_1, \dots, x_n\}) \log g(\{x_1, \dots, x_n\}) dx_1 \cdots dx_n \quad (6)$$

Lemma 1 shows that the multi-target set integral can be decomposed into a series of Bernoulli set integrals.

Lemma 1. Suppose $f(X)$ is as defined in (1), and $g(X)$ is an arbitrary set-valued function. Then

$$\int f(X) g(X) \delta X = \sum_{a \in \mathcal{A}} w_a \int \cdots \int \prod_{i=1}^N f_{i|a}(X_i) g\left(\bigcup_{i=1}^N X_i\right) \delta X_1 \cdots \delta X_N \quad (7)$$

The proof of the lemma is omitted due to space constraints, and can be found in [10].

Corollary 1. Let $[X_i] \triangleq (X_1, \dots, X_N)$. Suppose that an alternative definition of a set-valued function \tilde{g} satisfies $\tilde{g}([X_i]) = g(X)$ for any $[X_i]$ such that $\uplus_{i=1}^N X_i = X$. Then (7) can be equivalently evaluated as:

$$\int f(X) g(X) \delta X = \sum_{a \in \mathcal{A}} w_a \int \cdots \int \prod_{i=1}^N f_{i|a}(X_i) \tilde{g}([X_i]) \delta X_1 \cdots \delta X_N \quad (8)$$

Lemma 2. Suppose $\tilde{h}(X) \triangleq c_{|X|} h(X)$. Then

$$\int f(X) \log \tilde{h}(X) \delta X = \sum_n f(n) \log c_n + \int f(X) \log h(X) \delta X \quad (9)$$

where $f(n)$ is the cardinality distribution corresponding to $f(X)$.

The proof of lemma 2 simply separates the log of the product into the sum of logs and simplifies. With these preliminary results, we are now ready to state the first theorem, that the RFS KL divergence for the MB distributions (4) can be optimised by summing over assignments of Bernoulli components in the original distribution to Bernoulli components in the simplified distribution.

Definition 1. Denote by Π_N the set of complete permutation functions on $I_N \triangleq \{1, \dots, N\}$:

$$\Pi_N = \{\pi : I_N \rightarrow I_N | i \neq j \Rightarrow \pi(i) \neq \pi(j)\}$$

Theorem 1. The solution of the optimisation

$$\operatorname{argmax}_{[g_j]} \sum_{a \in \mathcal{A}} w_a \int \cdots \int \prod_{i=1}^N f_{i|a}(X_i) \cdot \log \tilde{g}([X_i]) \delta X_1 \cdots \delta X_N \quad (10)$$

where

$$\tilde{g}([X_i]) = \sum_{\pi \in \Pi_N} \prod_{i=1}^N g_{\pi(i)}(X_i) \quad (11)$$

is the same as the solution of problem 1.

The proof of the theorem is omitted due to space constraints, and can be found in [10].

2.1. Approximate solution of BFMB

We propose an approximate solution of (10) based on the view of expectation-maximisation (EM) presented in [11], effectively treating the correspondence between the underlying Bernoulli distribution $f_{i|a}(X)$ and the best-fitting Bernoulli distribution $g_j(X)$ as missing data. There is a separate missing data distribution (i.e., a distribution over correspondence) for each component a in the MB mixture; the distribution under the a -th component is $q_a(\pi)$. We constrain $q_a(\pi) \geq 0$; $\forall a, \pi$, and $\sum_{\pi \in \Pi_N} q_a(\pi) = 1 \forall a$. Accordingly, solution of (10) is equivalent to minimisation of $J([g_j])$, where

$$J([g_j]) = - \sum_{a \in \mathcal{A}} w_a \int \cdots \int \prod_{i=1}^N f_{i|a}(X_i) \cdot \log \sum_{\pi \in \Pi_N} \prod_{i=1}^N g_{\pi(i)}(X_i) \delta X_1 \cdots \delta X_N \quad (12)$$

$$\begin{aligned}
&= \sum_{a \in \mathcal{A}} w_a \left(\sum_{\pi \in \Pi_N} q_a(\pi) \right) \int \cdots \int \prod_{i=1}^N f_{i|a}(X_i) \cdot \\
&\quad \cdot \log \frac{\sum_{\pi \in \Pi_N} q_a(\pi)}{\sum_{\pi \in \Pi_N} \prod_{i=1}^N g_{\pi(i)}(X_i)} \delta X_1 \cdots \delta X_N \quad (13)
\end{aligned}$$

$$\begin{aligned}
&\leq \sum_{a \in \mathcal{A}, \pi \in \Pi_N} w_a q_a(\pi) \int \cdots \int \prod_{i=1}^N f_{i|a}(X_i) \cdot \\
&\quad \cdot \log \frac{q_a(\pi)}{\prod_{i=1}^N g_{\pi(i)}(X_i)} \delta X_1 \cdots \delta X_N \quad (14)
\end{aligned}$$

$$\begin{aligned}
&= \sum_{a \in \mathcal{A}, \pi \in \Pi_N} w_a q_a(\pi) \log q_a(\pi) \\
&\quad - \sum_{a \in \mathcal{A}, \pi \in \Pi_N} w_a q_a(\pi) \sum_{i=1}^N \int f_{i|a}(X_i) \log g_{\pi(i)}(X_i) \delta X_i \quad (15)
\end{aligned}$$

$$\begin{aligned}
&\leq T \cdot \sum_{a \in \mathcal{A}, \pi \in \Pi_N} w_a q_a(\pi) \log q_a(\pi) \\
&\quad - \sum_{a \in \mathcal{A}, \pi \in \Pi_N} w_a q_a(\pi) \sum_{i=1}^N \int f_{i|a}(X_i) \log g_{\pi(i)}(X_i) \delta X_i \quad (16)
\end{aligned}$$

$$\triangleq \tilde{J}_T([g_j], [q_a(\pi)])$$

where (13) simply multiplies by $\sum_{\pi \in \Pi_N} q_a(\pi) = 1$ and similarly adds $\log(1) = 0$, (14) invokes the log-sum inequality, and (15) replaces the log of a product with the sum of logs and simplifies. This process is commonly used to understand the behaviour of EM, e.g., [12, p363], [11]. In (16), we observe that the first term is negative (since $0 \leq q_a(\pi) \leq 1$), hence incorporating a multiplier $0 \leq T \leq 1$ loosens the bound. In statistical physics this corresponds to the temperature; the need for the change is discussed in the experimental results. Setting $T = 0$ yields the ‘‘winner-takes-all’’ variant of EM [11] which is widely used in pattern recognition, e.g., as the k -means algorithm.

Subsequently, $\tilde{J}_T([g_j], [q_a(\pi)])$ is minimised by block coordinate descent, alternating between minimisation with respect to $[g_j]$ (M-step), and $[q_a(\pi)]$ (E-step). These two steps can be solved as:

$$g_j(X) = \sum_{a \in \mathcal{A}, \pi \in \Pi_N} w_a q_a(\pi) f_{\pi^{-1}(j)|a}(X) \quad (17)$$

$$q_a(\pi) \propto \prod_{i=1}^N \exp \left\{ \frac{1}{T} \int f_{i|a}(X) \log g_{\pi(i)}(X) \delta X \right\} \quad (18)$$

If the distributions $[g_j]$ are to be constrained to be Bernoulli-Gaussian, (17) is replaced by expressions matching the probability of existence, mean and covariance to the expression in (17); details of this can be found in [10]. The Bernoulli-Gaussian form is convenient since it permits closed-form evaluation of (18).

This procedure is guaranteed to converge to a local minimum of \tilde{J} since it is strictly convex with respect to both $[g_j]$ and $[q_a(\cdot)]$ (but not jointly convex). In most applications of EM, the missing data is estimated for each of a finite number of training samples. In turn, this guarantees that log-sum inequality is tight at the optimum, and hence that the procedure will converge to a local minimum of the original likelihood function (in this case $J([g_j])$) [11]. In the present application, each point in the multi-target distribution $f(X)$ is effectively a training sample, so to maintain a similar guarantee, the missing data would need to be permitted to be a function of the joint target state $X = \bigcup_{i=1}^N X_i$. The above derivation could be easily modified to address this case, and the resulting method could be implemented by sampling the joint state space and estimating missing data for every joint sample. However, the computational complexity of this procedure is likely to be prohibitive, as observed in KLSJPDA [9]. The proposed method may be viewed as constraining the missing data to vary only with the association hypothesis a . Alternatively, we simply minimise an upper bound that is not tight at the optimum. SJPDA simplifies the complexity of KLSJPDA similarly, constraining the assignment to depend only on the association hypothesis.

3. EXPERIMENTAL RESULTS

We demonstrate the approach on a simple one-dimensional case involving two Bernoulli-Gaussian tracks. The prior distributions have a probability of existence of 0.8 and 0.9, means of 10 and 11, and a variance of 9. Two measurements are received, at coordinates 8 and 13. The probability of detection is 0.6, the false alarm intensity is 0.1, and the variance of the additive Gaussian measurement noise is 1.

The marginal posterior of each track is shown in figure 1(a) (i.e., that obtained using TOM-MeMBer/P described in [2], conceptually similar to JPDA); the initialisation used for BFMB is a series of Bernoulli-Gaussian distributions matching the probability of existence, mean and covariance of these posterior marginals. The solution obtained using BFMB with small T is shown in figure 1(b). While the Gaussian distribution fit to the marginals would attempt to cover both modes (leading to two very similar tracks with much higher variance), BFMB successfully separates the two peaks, eliminating coalescence. The RFS KL divergence (3) of the MSBJ solution is shown in figure 1(c). The dotted line is the KL divergence of the Bernoulli-Gaussian distribution fit to the marginal posterior (i.e., the initialisation used in the optimisation). With $T > 0.7$, the algorithm converges (very slowly) to two identical Gaussian distributions covering both modes. This occurs because the reduction in the bound \tilde{J}_T obtained by setting $q_a(\pi)$ to uniform (thereby maximising its entropy) outweighs the improvement that can be achieved in the mean log-likelihood of the true data under the simplified distribution. When the upper bound is loosened by reducing the tem-

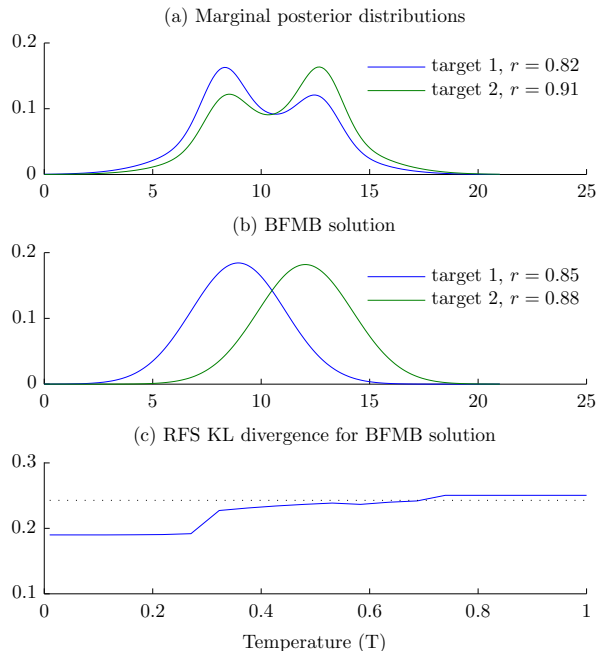


Fig. 1. Results of experiments: (a) shows the marginal posterior distribution, (b) shows the result of BFMB with $T = 0.01$, and (c) shows the RFS KL divergence (3) of the solution of BFMB as a function of the temperature T (evaluated numerically).

perature (reducing the influence of the entropy term), the algorithm concentrates more on improving the log-likelihood of the simplified distribution, and convergence occurs rapidly. The result in figure 1(c) demonstrates that the solutions obtained in this case have a lower RFS KL divergence (i.e., the desired but intractable objective).

With $T = 0$, the E-step reverts to finding the most likely assignment π_a for each association hypothesis $a \in \mathcal{A}$ (where the cost is the negative log of (18), omitting the $\frac{1}{T}$ factor); this can be solved efficiently using methods such as the auction algorithm. The difference between the cases with $T = 1$ and $T = 0$ is analogous to the difference between using EM and k -means to estimate parameters of a Gaussian mixture.

4. CONCLUSION AND EXTENSIONS

This paper has presented a principled, approximate method for finding the MB distribution that minimises the RFS KL divergence from the full RFS distribution. The disadvantage of the method is that it requires enumeration of all global association hypotheses $a \in \mathcal{A}$, and solution of an assignment problem for each hypothesis. The extension in the preprint [10] demonstrates that, using the preferred value $T = 0$, the optimisation can be expressed in a form that avoids explicit enumeration of global association hypotheses. Subsequently, with an approximation of the feasible set of the optimisation,

a polynomial time algorithm is obtained. Computational results show excellent performance in very challenging dense target scenarios.

5. REFERENCES

- [1] J. L. Williams, “Experiments with graphical model implementations of multiple target multiple Bernoulli filters,” in *Proc. 7th International Conference on Intelligent Sensors, Sensor Networks and Information Processing*, Adelaide, Australia, December 2011, pp. 532–537.
- [2] —, “Marginal multi-Bernoulli filters (extended version),” *Submitted to IEEE Trans. Aerosp. Electron. Syst.*, 2013. [Online]. Available: <http://arxiv.org/abs/1203.2995v2>
- [3] B.-T. Vo and B.-N. Vo, “Labeled random finite sets and multi-object conjugate priors,” *IEEE Trans. Signal Process.*, vol. 61, no. 13, pp. 3460–3475, 2013.
- [4] T. Fortmann, Y. Bar-Shalom, and M. Scheffe, “Sonar tracking of multiple targets using joint probabilistic data association,” *IEEE J. Ocean. Eng.*, vol. 8, no. 3, pp. 173–184, Jul 1983.
- [5] D. Musicki and R. J. Evans, “Joint integrated probabilistic data association: JIPDA,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 40, no. 3, pp. 1093–1099, July 2004.
- [6] R. P. S. Mahler, *Statistical Multisource-Multitarget Information Fusion*. Norwood, MA: Artech House, 2007.
- [7] B.-T. Vo, B.-N. Vo, and A. Cantoni, “The cardinality balanced multi-target multi-Bernoulli filter and its implementations,” *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 409–423, Feb. 2009.
- [8] H. A. Blom and E. A. Bloem, “Probabilistic data association avoiding track coalescence,” *IEEE Trans. Autom. Control*, vol. 45, no. 2, pp. 247–259, February 2000.
- [9] L. Svensson, D. Svensson, M. Guerriero, and P. Willett, “Set JPDA filter for multitarget tracking,” *IEEE Trans. Signal Process.*, vol. 59, no. 10, pp. 4677–4691, October 2011.
- [10] J. L. Williams, “An efficient, variational approximation of the best fitting multi-Bernoulli filter,” *Submitted to IEEE Trans. Sig. Proc.*, 2014. [Online]. Available: <http://arxiv.org/abs/1403.5022>
- [11] R. Neal and G. E. Hinton, “A view of the EM algorithm that justifies incremental, sparse, and other variants,” in *Learning in Graphical Models*. Kluwer Academic Publishers, 1998, pp. 355–368.
- [12] K. P. Murphy, *Machine learning: a probabilistic perspective*. Cambridge, MA: MIT Press, 2012.