

Efficiency and Sensitivity of Methods for Assessing Ambiguity in Data Association Decisions

Bret D. Kragel, Shawn M. Herman, and Nick J. Roseveare
Numerica Corporation, 4850 Hahns Peak Dr., Loveland, Colorado

ABSTRACT

The central problem in multitarget, multisensor surveillance is that of determining which reports from separate sensors arise from common objects. Due to stochastic errors in the source reports, there may be multiple data association hypotheses with similar likelihoods. Moreover, established methods for performing data association make fundamental modeling assumptions that hold only approximately in practice. For these reasons, it is beneficial to include some measure of uncertainty, or ambiguity, when reporting association decisions. In this paper, we perform an analysis of the benefits versus runtime performance of three methods of producing ambiguity estimates for data association: enumeration of the k -best data association hypotheses, importance sampling, and Markov Chain Monte Carlo estimation. In addition, we briefly examine the sensitivity of ambiguity estimates to violations of the stochastic model used in the data association procedure.

Keywords: Multitarget tracking, data association, importance sampling, Markov Chain Monte Carlo, Murty's method, k -best enumeration

1. PROBLEM OVERVIEW

The central problem in multitarget, multisensor surveillance is that of partitioning sensor reports into sets of observations that most likely originated from the same truth object, versus observations that most likely were produced by clutter. Let $Z_k = \{z_{i_k}^k\}_{i_k=1}^{m_k}$ denote a sequence of noise-contaminated measurements produced by a sensor at time t_k , and let $Z^M = \{Z_k\}_{k=1}^M$ denote the data from M sensor scans, potentially from multiple sensors. The central data association problem in multitarget, multisensor data fusion can be generally posed as (cf. Poore¹)

$$H^* = \operatorname{argmax}_{H \in \mathcal{H}_M} \{L_{H(M)}\}, \quad L_{H(M)} := \frac{\Pr(H | Z^M)}{\Pr(H_0 | Z^M)}, \quad (1)$$

where H denotes a partition of the data into tracks and false alarms, H_0 denotes a reference partition in which all reports are declared to be false alarms, \mathcal{H}_M denotes the set of all feasible partitions of the data Z^M , and H^* denotes the optimal partition. Since the number of partitions, or *data association hypotheses*, grows exponentially with the number of frames, online solution of the association problem typically requires some form of approximation. This leads to multiple-frame processing, in which data association hypotheses are maintained within a sliding window of $N \leq M$ data frames. To obtain the best data partition at the front of the window for the current frame, one needs to solve an $(N + 1)$ -dimensional assignment problem. In *single-frame* tracking systems, association decisions are based on the information from a single frame of data, i.e., $N = 1$; thus, the single-frame approach requires the solution of a *two-dimensional assignment* problem and maintains only a single data association hypothesis over time.

Typically, when evaluating the likelihood ratio (1), one uses a negative log-likelihood score instead: $c_{H(M)} := -\ln L_{H(M)}$. This implies

$$\Pr(H | Z^M) = e^{-c_{H(M)}} \Pr(H_0 | Z^M), \quad (2)$$

which gives the *probability of a hypothesis* as a function of its cost and the probability of the null hypothesis.

Further author information: (Send correspondence to B.D.K.)

B.D.K.: E-mail: bret.kragel@numerica.us, Telephone: 1 970 461 2000

S.M.H.: E-mail: shawn.herman@numerica.us, Telephone: 1 970 461 2000

2. AMBIGUITY ASSESSMENT

Although multiple hypothesis trackers can maintain many alternative data association hypotheses, track reports would typically be generated using the best hypothesis.² Just as each track report includes both a state estimate and a corresponding state covariance matrix, it is beneficial to provide an assessment of data association ambiguity along with the best association hypothesis.

We begin by defining the concept of ambiguity within the context of multiple hypothesis tracking as introduced by Gadaleta et al.³ Let $\mathcal{H}_N = \{H_1, H_2, \dots, H_\varpi\}$ denote the ranked set of all possible N -frame data association hypotheses at the M -th time frame, such that H_1 is the “best” solution returned from the assignment solver, H_2 the “second-best”, and so forth, where $\varpi = |\mathcal{H}_N|$ denotes the cardinality of the hypothesis space. Then, the negative log-likelihood costs of these hypotheses are ordered according to

$$c_{H_1}(M) \leq c_{H_2}(M) \leq \dots \leq c_{H_\varpi}(M), \quad (3)$$

and, within the N -frame hypothesis space,

$$\sum_{i=1}^{\varpi} \Pr(H_i | Z^M) = 1. \quad (4)$$

Therefore, we can use Equation (2) to obtain^{2,4}

$$\Pr(H_i | Z^M) = \frac{\Pr(H_i | Z^M)}{1} = \frac{\Pr(H_i | Z^M)}{\sum_{j=1}^{\varpi} \Pr(H_j | Z^M)} = \frac{e^{-c_{H_i}(M)}}{\sum_{j=1}^{\varpi} e^{-c_{H_j}(M)}}, \quad (5)$$

which gives the *probability of a hypothesis in terms of its cost* $c_H(M)$.

The probabilities $\Pr(H_i | Z^M)$ of the N -frame data association hypotheses are directly related to the N -frame probabilities of the data assignments as shown next. For example, for a measurement-to-track association problem, let h_{ijk} denote the *probabilistic data association event* that measurement \mathbf{z}_j associates with track \mathbf{x}_i within sensor data frame Z_k . Then, the exact N -frame* probability $\Pr(h_{ijk})$ of this event is given by

$$\Pr(h_{ijk}) = \sum_{H \in \mathcal{H}_N^{ijk}} \Pr(H | Z^M), \quad (6)$$

where $\mathcal{H}_N^{ijk} \subseteq \mathcal{H}_N$ denotes the subset of those data association hypotheses from the set \mathcal{H}_N that postulate the event h_{ijk} . The assignment probability $\Pr(h_{ijk})$ serves as a measure of *data association uncertainty* or *ambiguity*: If $\Pr(h_{ijk}) < 1$, the tracking system estimates that there is a nonzero probability that the corresponding association from the best solution H_1 is incorrect. However, to compute the assignment probability $\Pr(h_{ijk})$, we need to find all those data association hypotheses in which the event h_{ijk} is postulated. Although current combinatorial algorithms are capable of finding the optimal data association hypothesis with negligible computational effort, the size of \mathcal{H}_N^{ijk} can become enormous for complicated scenarios; thus, it may be computationally impractical for any algorithm to explicitly enumerate all “reasonable” hypotheses. For this reason, in the sequel we describe and assess several methods for estimating data association ambiguity.

Section 2.1 introduces the most straightforward approach, k -best enumeration, which approximates association probabilities using only the k hypotheses with the greatest probability mass. Since the computational effort required to generate each hypothesis using the k -best method is relatively high, we also consider two stochastic estimation methods. In Section 2.2, we formulate a general framework for using Monte Carlo methods to estimate association probabilities, then specialize this framework to Markov Chain Monte Carlo methods in Section 2.3 and importance sampling in Section 2.4. Section 3 is devoted to a comparative analysis of these three methods, while Section 4 summarizes the results.

2.1 Ambiguity Estimation Using k -Best Enumeration

Estimation of association probabilities using k -best enumeration is straightforward: To approximate the optimal solution to Equation (1), a multiple hypothesis tracking system directly maintains a set $\mathcal{H}_N^{kM} \subset \mathcal{H}_N$ of k_M “complete” data association hypotheses $H_i \in \mathcal{H}_N^{kM}$, $i = 1, \dots, k_M$ at time t_M within a sliding window of data frames.

*Note that the N -frame probabilities are approximations of the full M -frame probabilities for $N < M$.

In the following, let $\mathcal{H}_N^{k_M} = \{H_1, H_2, \dots, H_{k_M}\}$ denote the set of k_M ranked solutions at time t_M . Then, as long as k_M is sufficiently large, Equation (5) can be approximated by

$$\Pr(H_i | Z^M) \approx \frac{e^{-c_{H_i}(M)}}{\sum_{j=1}^{k_M} e^{-c_{H_j}(M)}}. \quad (7)$$

Note that (3) implies

$$\Pr(H_1 | Z^M) \geq \Pr(H_2 | Z^M) \geq \dots \geq \Pr(H_{k_M} | Z^M), \quad (8)$$

meaning that (7) includes the k_M hypotheses with the highest probability.

As stated by Popp,⁴ it may be necessary to “shift” the costs in (2) for numerical stability:

$$\Pr(H_i | Z^M) = \frac{e^{-c_{H_i}(M)+c_s}}{\sum_{j=1}^{k_M} e^{-c_{H_j}(M)+c_s}}, \quad (9)$$

where a possible choice for the cost shift is $c_s = c_{H_1}$. Using the approximation (7), the individual association probabilities (6) can be estimated using

$$\Pr(h_{ijk}) \approx \sum_{H \in \mathcal{H}_N^{ijk} \cap \mathcal{H}_N^{k_M}} \Pr(H | Z^M). \quad (10)$$

It is important to note that the k -best method is not “compelled” to compute k full hypotheses; indeed, if no hypothesis with nonnegligible probability remains after the $\bar{k} < k$ best hypotheses have been generated, the method can halt processing. This capability can sometimes result in significant reduction in computational effort.

2.2 Monte Carlo Ambiguity Estimation

As mentioned at the beginning of Section 2, for tracking scenarios with large numbers of densely-spaced objects, k -best enumeration typically cannot generate a number of hypotheses sufficient to compute accurate association probabilities in a reasonable amount of time. For these situations, we consider two stochastic algorithms: Markov Chain Monte Carlo sampling and importance sampling. Since both methods are special cases of Monte Carlo estimation, we precede their introduction with a short exposition on the use of Monte Carlo methods for ambiguity assessment.

Consider the problem of evaluating the expectation

$$E[g(X)] = \int_{x \in \mathcal{X}} g(x)p(x)dx, \quad (11)$$

where X is a random variable defined on the space \mathcal{X} and distributed according to the probability function p , $E[\cdot]$ denotes the expectation operator, and g is some function of interest. Then $E[g(X)]$ can be estimated by generating a random sample $\{X_1, X_2, \dots, X_n\}$ from p and computing the empirical average⁵

$$\bar{g}_n = \frac{1}{n} \sum_{i=1}^n g(x_i), \quad (12)$$

where the lower case letters denote a realization of the random sample. If $E[g(X)] < \infty$, the Strong Law of Large Numbers guarantees

$$\lim_{n \rightarrow \infty} \bar{g}_n \stackrel{a.e.}{=} E[g(X)]. \quad (13)$$

Ambiguity assessment can be accommodated within the framework of Equations (11) – (13) by identifying $\mathcal{X} \leftrightarrow \mathcal{H}$ and $X \leftrightarrow H$, choosing $\Pr(H|Z^M)$ as the probability function for H , and setting $g(H) = \mathbb{I}_{ijk}(H)$, where

$$\mathbb{I}_{ijk}(H) = \begin{cases} 1, & \text{if } h_{ijk} \in H, \\ 0, & \text{otherwise.} \end{cases}$$

Then,

$$\Pr(h_{ijk}) = E[\mathbb{I}_{ijk}(H)] \approx \frac{1}{n} \sum_{i'=1}^n \mathbb{I}_{ijk}(H_{i'}), \quad (14)$$

where the random sample $\{H_1, \dots, H_n\}$ is generated from $\Pr(H|Z^M)$.

The difficulty now becomes the problem of generating samples from

$$\Pr(H|Z^M) = \frac{p(H, Z^M)}{p(Z^M)}, \quad (15)$$

an issue complicated by two factors. First, though $p(H, Z^M)$ can be computed,⁶ $p(Z^M)$ cannot be computed directly with reasonable computational resources. Second, even if (15) could be computed, drawing samples from this distribution would remain nontrivial. Nevertheless, there are methods available for solving this problem. Two such methods, Markov Chain Monte Carlo (MCMC) sampling and importance sampling (IS), appear to be suitable for the problem of ambiguity estimation. In the following, we describe these methods and assess their performance within the context of ambiguity assessment as expressed in Equation (14).

2.3 Markov Chain Monte Carlo Ambiguity Estimation

Recently, some authors⁷⁻⁹ have begun experimenting with the use of Markov Chain Monte Carlo (MCMC) methods for multitarget tracking. As we shall see, the primary advantage of MCMC is the ability to generate samples from fairly arbitrary probability distributions; however, the samples generated are strongly correlated,^{10,11} which significantly complicates convergence analysis.⁵ Nevertheless, due to the inherent limitations of ambiguity estimation using k -best enumeration, we wish to evaluate the MCMC method's potential for enhancing ambiguity assessment in difficult tracking scenarios.

2.3.1 Markov Chains

Given a large (finite) set Ω of combinatorial structures (e.g., the space of all data association hypotheses), a sequence of random variables $\{X_0, X_1, X_2, \dots\}$ defined on Ω is called a *Markov chain* if

$$\Pr(X_{n+1} = x | X_n = x_n, \dots, X_0 = x_0) = \Pr(X_{n+1} = x | X_n = x_n), \quad x \in \Omega,$$

where $x_n \in \Omega$ denotes a realization of X_n , $n = 0, 1, 2, \dots$; thus, a Markov chain is characterized by the condition that, given the current state, all future states are conditionally independent of the prior states. Let $p_n(x) = \Pr(X_n = x)$ denote the state distribution at time (step) n , with initial distribution $p_0(x)$. If the *transition kernel* T_n at time n is defined by $T_n(x, y) = \Pr(X_{n+1} = y | X_n = x)$, $x, y \in \Omega$, then by the total probability theorem,

$$p_{n+1}(y) = \sum_{x \in \Omega} p_n(x) T_n(x, y).$$

A Markov chain is called *homogeneous* if the transition probabilities do not depend on time; that is, $T_n \equiv T$, $\forall n$. The distribution $\pi(x)$ is *stationary* with respect to the Markov chain with transition probabilities $T_n(x, y)$ if for all n ,

$$\pi(y) = \sum_{x \in \Omega} \pi(x) T_n(x, y).$$

A homogeneous Markov chain is said to be *irreducible* if there is a positive probability of moving from any state x to any other state y in a finite number of time steps, i.e., $\forall x, y \in \Omega$, $\exists n = n(x, y)$, such that $T^n(x, y) > 0$. The *period* of a state $x \in \Omega$ is defined as $d(x) = \gcd\{n \geq 1 \mid T^n(x, x) > 0\}$ [†]. The Markov chain is called *aperiodic* if $d(x) = 1$, $\forall x$.

The *detailed balance condition* holds for a homogeneous Markov chain if there is a distribution $\pi(x)$, such that for all states, $x, y \in \Omega$,

$$\pi(y)T(y, x) = \pi(x)T(x, y).$$

Note that the detailed balance condition implies $\pi(x)$ is a stationary distribution, since

$$\pi(y) = \pi(y) \sum_{x \in \Omega} T(y, x) = \sum_{x \in \Omega} \pi(y)T(y, x) = \sum_{x \in \Omega} \pi(x)T(x, y).$$

[†] $\gcd \triangleq$ greatest common divisor

Algorithm 1 Metropolis-Hastings Algorithm

Given: $\pi, x \in \Omega, N > 0$

- 1: **for** $n = 1$ to N **do**
 - 2: Select candidate state $y \in \Omega$ from the proposal distribution $q(y|x)$.
 - 3: Sample u from $U(0, 1]$.
 - 4: **if** $u < a(x, y) = \min \left\{ 1, \frac{\pi(y)q(x|y)}{\pi(x)q(y|x)} \right\}$ **then**
 - 5: Set $x = y$.
 - 6: **else**
 - 7: Set $x = x$.
 - 8: **end if**
 - 9: **end for**
-

2.3.2 The Metropolis-Hastings Algorithm

MCMC can be used when one desires to sample from a distribution π known only up to a constant factor, i.e., $\pi = \pi^*/C$, where π^* is known, but the constant C is unknown or difficult to compute. Note that this is the situation in the multitarget tracking problem, where one must evaluate probabilities of the form (15), and the unknown normalizing constant is $C = p(Z^M)$. MCMC operates by constructing an aperiodic and irreducible Markov chain that has the distribution of interest as its unique stationary distribution, then sampling from this chain as described below.

The most popular method in the tracking literature for generating a Markov chain with the desired characteristics is the Metropolis-Hastings algorithm,^{5,8,12} described in Algorithm 1. Given the current state $x \in \Omega$, a *candidate state* $y \in \Omega$ is chosen according to a *proposal distribution* $q(y|x)$. The proposed state y is accepted only if the value of the acceptance function

$$a(x, y) = \min \left\{ 1, \frac{\pi(y)q(x|y)}{\pi(x)q(y|x)} \right\} \quad (16)$$

is greater than or equal to a number u drawn uniformly from the interval $(0, 1]$. If y is rejected, then the chain remains in the current state x . In this manner, every proposed state with $\pi(y)q(x|y) \geq \pi(x)q(y|x)$ is accepted, while any other proposal with $\pi(y)q(x|y) < \pi(x)q(y|x)$ has probability $\frac{\pi(y)q(x|y)}{\pi(x)q(y|x)}$ of being accepted. This second condition is designed to keep the chain from getting “stuck” at a local optimum. For $y \neq x$, the resulting transition kernel, derived by multiplying the probability of proposal with the probability of acceptance, is given by

$$T(x, y) = q(y|x)a(x, y) = \min \left\{ q(y|x), \frac{\pi(y)}{\pi(x)}q(x|y) \right\}.$$

Thus, as can be seen by

$$\pi(y)T(y, x) = \pi(y) \min \left\{ q(x|y), \frac{\pi(x)}{\pi(y)}q(y|x) \right\} = \pi(x) \min \left\{ \frac{\pi(y)}{\pi(x)}q(x|y), q(y|x) \right\} = \pi(x)T(x, y),$$

the inclusion of $q(y|x)$ in (16) ensures that the detailed balance condition holds[‡], meaning that $\pi(x)$ is stationary for the Markov chain. Since aperiodicity is guaranteed by the acceptance function, if the proposal distribution can be constructed in such a way that the Metropolis-Hastings Markov chain is also irreducible, then as noted by Robert and Casella,⁵ the satisfaction of the aperiodicity and irreducibility conditions implies that a sequence produced by the Metropolis Hastings algorithm can be employed as an independent identically distributed sample from $\pi(x)$, with convergence of the empirical average of the chain guaranteed by the ergodic theorem;^{5,13} that is,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n g(X_i) \stackrel{a.e.}{=} E[g(X)], \quad (17)$$

where $\{X_1, X_1, X_2, \dots, X_n\}$ is a sequence of random variables produced by the Metropolis-Hastings Markov chain.

[‡]Note that the detailed balance condition always holds for the transition from x to itself.

Algorithm 2 Proposal Distribution for MCMC Testing

Given: Current hypothesis H , where $n_c = |\text{Dom}(H)|$ denotes the number of hypothesized common objects.

- 1: **if** $n_c = 0$ **then**
 - 2: Add a random association.
 - 3: **else if** $n_c = \min\{n_1, n_2\}$ **then**
 - 4: Select $u \in U(0, 1]$.
 - 5: **if** $u \leq \frac{1}{2}$ **then**
 - 6: Delete a random association.
 - 7: **else**
 - 8: Perform a random swap.
 - 9: **end if**
 - 10: **else**
 - 11: Select $u \in U(0, 1]$.
 - 12: **if** $u \leq \frac{1}{3}$ **then**
 - 13: Add a random association.
 - 14: **else if** $\frac{1}{3} < u \leq \frac{2}{3}$ **then**
 - 15: Delete a random association.
 - 16: **else**
 - 17: Perform a random swap.
 - 18: **end if**
 - 19: **end if**
-

2.3.3 Estimating Association Probabilities Using MCMC

If we can use the Metropolis-Hastings method to construct a Markov chain that converges to $\Pr(H|Z^M)$, then we can draw sample hypotheses $\{H_1, H_2, H_3, \dots\}$ from this chain and, according to (14), estimate the individual association ambiguities using

$$\Pr(h_{ijk}) \approx \frac{N_{ijk}}{N_{MC}}, \quad (18)$$

where N_{MC} is the total number of samples drawn, and N_{ijk} is the number of sampled hypotheses that postulate the event h_{ijk} . It is important to note that n indexes the *sampled* hypotheses here; thus, in contrast to the situation of Section 2.1, the hypotheses drawn from the simulated Markov chain are not ordered; that is, H_n is not necessarily “better” than H_{n+1} for any n .

To simplify our experimental analysis in Section 3, we considered a two-source track-to-track association problem, rather than the more typical measurement-to-track problem. For this scenario, our previous results can be applied by choosing $M = 2$, where Z_1 denotes a set of n_1 tracks from the first source, and Z_2 is a set of n_2 tracks from the second source. Hypotheses are considered to be bijective mappings from subsets of $\{1, \dots, n_1\}$ onto subsets of $\{1, \dots, n_2\}$. For this problem, we used an MCMC proposal distribution $q(\tilde{H}|H)$ that constructs a proposal hypothesis \tilde{H} from the current hypothesis H by altering the set of associations in H in one of three ways: addition, deletion, or swap. For an addition, we select uniformly at random one index $i \in \{1, \dots, n_1\} \setminus \text{Dom}(H)$ and one index $j \in \{1, \dots, n_2\} \setminus \text{Rng}(H)$, initialize a proposal hypothesis \tilde{H} with $\tilde{H} = H$, and set $\tilde{H}(i) = j$. For a deletion, we select uniformly at random one index $i \in \text{Dom}(H)$, initialize a proposal hypothesis \tilde{H} with $\tilde{H} = H$, and remove the chosen association by removing i from $\text{Dom}(\tilde{H})$ and $\tilde{H}(i)$ from $\text{Rng}(\tilde{H})$. For a swap, we select uniformly at random two indices $i_1, i_2 \in \text{Dom}(H)$, initialize a proposal hypothesis \tilde{H} with $\tilde{H} = H$, and swap the associations by setting $\tilde{H}(i_1) = H(i_2)$ and $\tilde{H}(i_2) = H(i_1)$.

The proposal distribution $q(\tilde{H}|H)$ is described in detail in Algorithm 2. Note that the algorithm must handle two special cases. In the case that the current hypothesis postulates no associations, a deletion or swap is not possible and an addition must be proposed. If all tracks from either source have been associated in the current hypothesis, an addition is not possible and either a deletion or swap must be proposed. Finally, in the typical case in which some, but not all, tracks from each source have been associated, the algorithm is free to propose at random an addition, deletion, or swap.

Hypothesis	Column Assignments	Cost	Probability
H_1	(3, 2, 1)	-70	~ 0.5
H_2	(0, 2, 3)	-70	~ 0.5
H_3	(2, 0, 3)	-55	1.53×10^{-7}
H_4	(0, 2, 1)	-55	1.53×10^{-7}
H_5	(3, 2, 0)	-50	1.03×10^{-9}
H_6	(2, 0, 1)	-40	4.70×10^{-14}

Table 1. Column assignments for the 6-best hypotheses corresponding to the cost matrix of Equation (19). For example, the assignment (3, 2, 1) indicates that Column 1 was assigned to Row 3, Column 2 to Row 2 and Column 3 to Row 1. A ‘0’ indicates an unassigned column.

2.3.4 On the Difficulty of Constructing Appropriate Proposal Distributions

We note that although (17) guarantees convergence in theory for irreducible, aperiodic Markov chains, this does not guarantee efficient sampling in practice. Consider the cost matrix

$$\mathbb{C} = \begin{bmatrix} \infty & \infty & -20 \\ -20 & -35 & -15 \\ -15 & \infty & -35 \end{bmatrix}, \quad (19)$$

which might occur after gating for two sets of reports in a typical tracking scenario, where \mathbb{C}_{ij} is the cost of assigning report i from the first set with report j from the second set, and the cost of not assigning a report is zero (not shown in matrix). There are only 15 hypotheses with cost less than infinity for this problem, with the best two solutions consuming nearly all of the probability mass. The hypotheses with nonnegligible probability are listed in Table 1. The k -best algorithm can easily enumerate all solutions, resulting in the exact ambiguity matrix[§]

$$\mathbb{A} = \begin{bmatrix} 0.0 & 0.0 & \sim 0.5 \\ \sim 0.0 & \sim 1.0 & 0.0 \\ \sim 0.5 & 0.0 & \sim 0.5 \end{bmatrix},$$

where $\mathbb{A}_{ij} = \Pr(h_{ij})$ is the probability of assigning report i from the first set of reports with report j from the second set.

In contrast, regardless of initial hypothesis and despite using 10^6 Monte Carlo samples, the ambiguity matrix returned by the MCMC sampler of Section 2.3.3 reflects exactly one of the two best hypotheses depending on initial seed:

$$\mathbb{A}_1 = \begin{bmatrix} 0.0 & 0.0 & 1.0 \\ 0.0 & 1.0 & 0.0 \\ 1.0 & 0.0 & 0.0 \end{bmatrix} \quad \text{or} \quad \mathbb{A}_2 = \begin{bmatrix} 0.0 & 0.0 & 0.0 \\ 0.0 & 1.0 & 0.0 \\ 0.0 & 0.0 & 1.0 \end{bmatrix},$$

where \mathbb{A}_1 corresponds to H_1 , and \mathbb{A}_2 corresponds to H_2 . Upon examination of the probable hypotheses and the proposal distribution from Algorithm 2, it becomes clear why this is so. Once the chain has reached either H_1 or H_2 , a situation that always occurs during the MCMC burn-in period, the chance of leaving that hypothesis becomes remote. For example, if the chain is in H_1 , the shortest sequence to H_2 requires a transition through H_5 . Since the H_1 to H_5 has an acceptance probability on the order of 10^{-9} , we would on average need about 10^9 Monte Carlo samples before accepting this deletion.

Clearly, the problem described in this simple example can be circumvented by employing a more sophisticated proposal distribution;^{10,12} however, additional complexity generally implies more computational effort and, since both the proposal distribution $q(\tilde{H}|H)$ and its inverse $q(H|\tilde{H})$ must be computed, such an approach would be more difficult to design and implement. It should also be noted that as the complexity of the assignment problem increases, one can expect more ‘traps’ like the one in our simple example, requiring more sophisticated (and slower) proposal distributions.

[§]Note that the ambiguities do not sum to unity across some of the rows and columns due to positive probability that the row or column remains unassigned.

2.4 Ambiguity Estimation Using Importance Sampling

While Markov Chain Monte Carlo methods provide a general framework for sampling from a (nearly) arbitrary probability distribution, the estimation of quantities computed from the sample may suffer from slow convergence due to the inherent correlation between MCMC samples. A somewhat less general method that utilizes independent samples is importance sampling.^{5,14}

Consider again the problem (11) of evaluating the expectation

$$E[g(X)] = \int_{x \in \mathcal{X}} g(x)p(x)dx, \quad (20)$$

where the probability distribution $p(x)$ is too complex to be sampled from directly but can be computed to within a multiplicative constant; that is, $p(x) = p^*(x)/C_p$, where $p^*(x)$ can be computed. Assume now the existence of another probability function, $q(x)$, from which independent samples can be generated easily, and which can also be evaluated to within a multiplicative constant; that is $q(x) = q^*(x)/C_q$, where $q^*(x)$ can be computed. If the support of p is contained in the support of q , then

$$E[g(X)] = \int_{x \in \mathcal{X}} g(x) \frac{p(x)}{q(x)} q(x) dx, \quad (21)$$

and $E[g(X)]$ can be estimated by drawing a random sample from $q(X)$ and computing

$$\hat{g}_n = \frac{1}{\bar{w}_n} \sum_{i=1}^n g(x_i) w_i, \quad (22)$$

with weights $w_i = p^*(x_i)/q^*(x_i)$, and $\bar{w}_n = \sum_{i=1}^n w_i$. Similar to (12), the empirical average (22) converges to $E[g(X)]$ by the Strong Law of Large Numbers.⁵ In principle, any distribution that meets the above criteria can be used as a sampling distribution, but not all distributions will be equal in terms of convergence rate. Generally, distributions that are similar to the target distribution but have “fatter tails” are preferable.⁵

For the two-source track-to-track association problem analyzed in Section 3, we used an importance sampling distribution based on the assignment cost matrix; this favored hypotheses with lower costs, but allowed any feasible hypothesis to be generated with positive probability. As was the case with the MCMC proposal distribution of Section 2.3.3, the importance sampling approach can encounter “traps” similar to the one described in Section 2.3.4. In such cases the convergence rate can be improved by implementing a more sophisticated, albeit slower, sampling technique.

3. COMPARATIVE ANALYSIS OF AMBIGUITY ESTIMATION METHODS

In order to perform a comparative analysis of the three methods of ambiguity assessment presented in the preceding sections, namely k -best enumeration, importance sampling, and MCMC, we considered a simple data association scenario consisting of two sensors acquiring noisy measurements of targets distributed randomly throughout a common field of view. Missed detections were considered, but false targets were not simulated.

3.1 Simulation Procedure

Let $Z_1 := \{(\hat{x}_i, P_i), i = 1, \dots, n_1\}$ and $Z_2 := \{(\hat{y}_j, Q_j), j = 1, \dots, n_2\}$ denote the set of track state reports from Sensors 1 and 2, respectively. For Sensor 1, \hat{x}_i denotes the estimated position of Track i , and P_i is the corresponding error covariance of the estimate. Likewise, for Sensor 2, \hat{y}_j denotes the estimated position of Track j , and Q_j is the error covariance of the estimate. The data association hypothesis space \mathcal{H} comprises all bijective mappings from subsets of the Sensor 1 indices $\{1, \dots, n_1\}$ onto subsets of the Sensor 2 indices $\{1, \dots, n_2\}$. Because we wish to determine which, if any, of the tracks from Sensor 1 and Sensor 2 correspond to the same object, we seek the assignment (or hypothesis) $H^* \in \mathcal{H}$ for which

$$\Pr(H^* | Z_1, Z_2) = \max_H \Pr(H | Z_1, Z_2); \quad (23)$$

Algorithm 3 Metrics Generation for Monte Carlo Testing

```

1: for  $n \in \mathcal{T}$  do
2:   for  $V \in \mathcal{S}$  do
3:     for  $m = 1$  to  $N_m$  do
4:       Generate a set of  $n$  targets uniformly at random within a surveillance region of volume  $V$ .
5:       Generate noisy estimates by Sensor 1 and Sensor 2 of the target states.
6:       Detect and generate track state reports for  $n_1 \leq n$  Sensor 1 targets and  $n_2 \leq n$  Sensor 2 targets.
7:       Determine the optimal track-to-track assignment hypothesis  $H_1$  using the  $k$ -best algorithm.
8:       Compute  $P_E^{m,V} = \frac{n_c}{n_1}$ , where  $n_c$  denotes the number of correct Sensor 1 associations in the optimal hypothesis.
9:       for  $A \in \mathcal{A}$  do
10:        Compute the assignment probabilities  $\Pr(h_{ij})$  using algorithm  $A$ .
11:        Compute  $P_A^{m,V} = \frac{1}{n_1} \sum_{i=1}^{n_1} \Pr(h_{i,H_1(i)})$  for algorithm  $A$ .
12:       end for
13:     end for
14:     Compute  $P_E^V = \frac{1}{N_m} \sum_{m=1}^{N_m} P_E^{m,V}$  (empirical probability of correct association).
15:     for  $A \in \mathcal{A}$  do
16:       Compute  $P_A^V = \frac{1}{N_m} \sum_{m=1}^{N_m} P_A^{m,V}$  (predicted probability of correct association).
17:     end for
18:   end for
19: end for

```

that is, we seek the association hypothesis with the maximum *a posteriori* probability (MAP assignment). The hypothesis probability function is given by^{6,15}

$$\Pr(H | Z_1, Z_2) \propto \prod_{i \in \text{Dom}(H)} \frac{\exp\left(-\frac{1}{2} (\hat{x}_i - \hat{y}_{H(i)})^T (P_i + Q_{H(i)})^{-1} (\hat{x}_i - \hat{y}_{H(i)})\right)}{\beta_T (1 - P_D^1)(1 - P_D^2) |2\pi (P_i + Q_{H(i)})|^{1/2}}, \quad (24)$$

where $\text{Dom}(H)$ is the set of Sensor 1 track indices postulated to originate from objects tracked by both sensors, β_T is the object density in the common surveillance region, and P_D^k is the probability that Sensor k , $k = 1, 2$, detects an object in its field of view.

We simulated scenarios of various densities by generating targets uniformly at random in 3-dimensional surveillance regions of various sizes according to Algorithm 3, where $\mathcal{A} = \{k\text{-best, IS, MCMC}\}$ denotes the set of ambiguity assessment algorithms, $\mathcal{T} = \{5, 20\}$ is the set of target counts used, \mathcal{S} denotes 30 uniformly distributed surveillance region sizes, and N_m is the number of Monte Carlo simulations for each configuration. The settings $N_m = 2000$ and $N_m = 500$ were used for the 5-target and 20-target scenarios, respectively, with a smaller number of samples used in the second case to keep total runtime manageable. In order to simulate a large range of possible operating conditions, the smallest surveillance region measured 100 meters on a side, while the largest measured 20000 meters on a side. The reported error covariances were assumed to be normally distributed with a standard deviation of 100 meters in each dimension, and the probability of detection was set at $P_D = 0.90$ for both sensors. *These values do not reflect any specific scenario, sensor characteristics, or multitarget tracker performance, but were chosen for comparison of the proposed ambiguity estimation methods under controlled conditions.*

Since the number of nonnegligible hypotheses can be expected to increase exponentially as the target density increases, it is necessary to devise a measure of scenario density that is comparable across tests. We chose to use the average Mahalanobis distance (normalized by the dimension of the state space) between each truth object and its nearest neighbor. This normalized target spacing was computed according to

$$\frac{1}{3n} \sum_{i=1}^n \min_{j \neq i} \{(x_i - x_j)^T (P + Q)^{-1} (x_i - x_j)\},$$

where P is the Sensor 1 covariance, Q is the Sensor 2 covariance, and x_i denotes the true position of Target i , $i = 1, \dots, n$. Note that the dimensions of the smallest surveillance regions considered were equal to the sensor noise level, representing very dense scenarios.

We also varied the number of samples used by our estimation algorithms for each of the test configurations presented above. Since the computational effort required by each method to generate one candidate hypothesis varies considerably, we ensured comparable runtime results by adjusting the number of samples used by each method for a given test configuration. Generally, for a given setting of $k = \{100, 1000, 10000\}$ in the k -best approach, $10k$ importance samples and $100k$ MCMC proposals were generated.

The main goal of our analysis is to assess the relative performance versus runtime of the algorithms under consideration. This comparison was based on two metrics: average maximum distance from the baseline ambiguity matrix versus runtime, and predicted probability of correct association versus scenario density. The first of these metrics, distance from the baseline ambiguity matrix, was computed simply as[¶]

$$\frac{1}{N_m} \sum_{m=1}^{N_m} \max_{i,j} \{|\mathbb{A}_*^m - \mathbb{A}_{est}^m|\}, \quad (25)$$

where for Monte Carlo run $m \in \{1, \dots, N_m\}$, \mathbb{A}_*^m denotes the exact, or *baseline*, ambiguity matrix, \mathbb{A}_{est}^m is the *estimated* ambiguity matrix, and the maximum is taken over all entries of the absolute difference of the matrices. For the second metric, the predicted probability of correct association (computed in Line 16 of Algorithm 3) was compared with the empirical probability of correct association (computed in Line 14 of Algorithm 3). For a sufficient number of Monte Carlo runs, the empirical probability of correct association should yield a reasonably accurate estimate of the true average probability that an association is correct, while the predicted probability will be a function of the accuracy of the association probability estimates derived from the k -best, importance sampling, and MCMC methods.

It is important to note here that the likelihood equation of (24) is derived^{6,15} by integrating an equation of the form

$$\Pr(H | Z_1, Z_2) = \prod_{i \in \text{Dom}(H)} \frac{\int_{\mathcal{S}} p(\hat{x}_i | x) p(\hat{y}_{H(i)} | x) dx}{\beta_T (1 - P_D^1) (1 - P_D^2) \int_{\mathcal{S}} p(\hat{x}_i | x) dx \int_{\mathcal{S}} p(\hat{y}_{H(i)} | x) dx}, \quad (26)$$

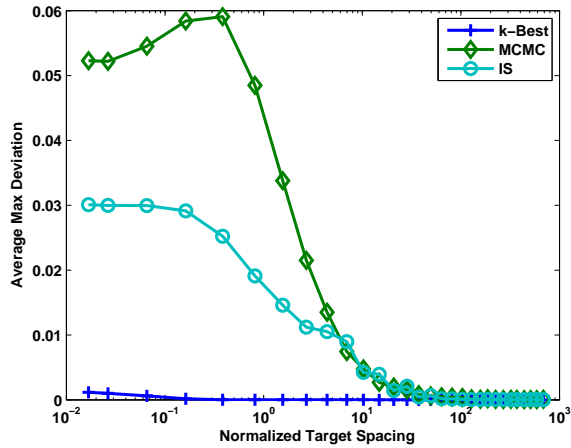
where \mathcal{S} is the surveillance region, and $p(\hat{x}_i | x)$ and $p(\hat{y}_{H(i)} | x)$ are Gaussian densities. If appreciable mass of $p(\hat{x}_i | x)$ or $p(\hat{y}_j | x)$ exists outside of \mathcal{S} , Equation (24) is no longer a valid expression for (26). In this case, the estimate of $\Pr(H | Z_1, Z_2)$ for the best hypothesis can be too large (optimistic) or too small (pessimistic). However, if the probabilities of the most likely hypotheses are nearly equal, the best hypothesis is more likely to have an overestimate of its true posterior probability than an underestimate of it. Therefore, because a dense scenario has (i) many targets near the boundary of \mathcal{S} , and (ii) many hypotheses with nearly equal probabilities, its probability of correct association P_A^V is likely to be overconfident when compared with the empirical results P_E^V , as will be seen in Section 3.2.

3.2 Simulation Results

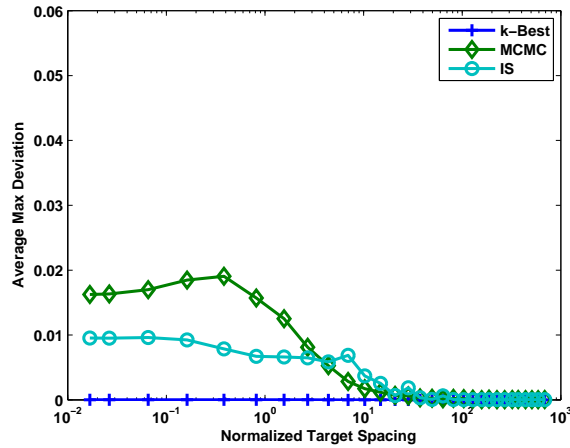
Using the simulation procedure of Section 3.1, we compared the performance of k -best enumeration, importance sampling, and MCMC for estimation of association probabilities (i.e., ambiguity). Figure 1 displays the average maximum deviation from the baseline ambiguity matrix for the 5-target case. The cardinality of the hypothesis space for this case was 1546, so the baseline ambiguity matrix could be computed exactly using the k -best approach with $k = 1546$. The k -best enumeration method (with $k \in \{100, 1000\}$) outperforms the Monte Carlo methods for this problem because the number of hypotheses with significant probability seldom exceeds k . The importance sampling and MCMC methods appear competitive with each other for this scenario, with the importance sampler enjoying an advantage in accuracy.

The average maximum deviation for the 20-target case is depicted in Figure 2. The cardinality of the hypothesis space for this case was well beyond the capacity of the k -best method, so the baseline ambiguity matrix was generated by MCMC using 10 million samples. This plot is much more interesting with respect to the k -best method. For both values of k , k -best enumeration performs better than the Monte Carlo methods until the scenario density is well below unity. Note also that

[¶]As in Section 2.3.4, the ambiguity matrix \mathbb{A} is defined by $\mathbb{A}_{ij} = \Pr(h_{ij})$, $i = 1, \dots, n_1$, $j = 1, \dots, n_2$, where n_k denotes the number of reports from Sensor k , $k = 1, 2$.

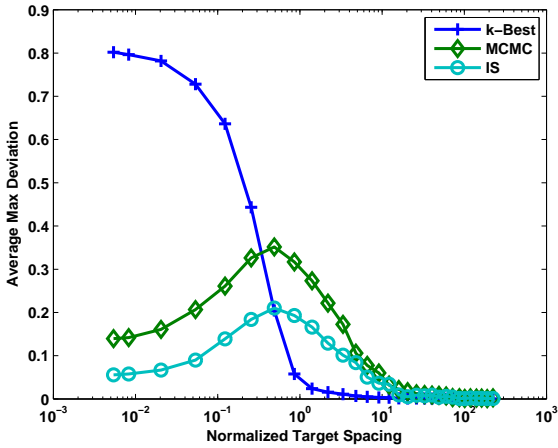


(a) Samples: k -best = 100, IS = 1000, MCMC = 10000.

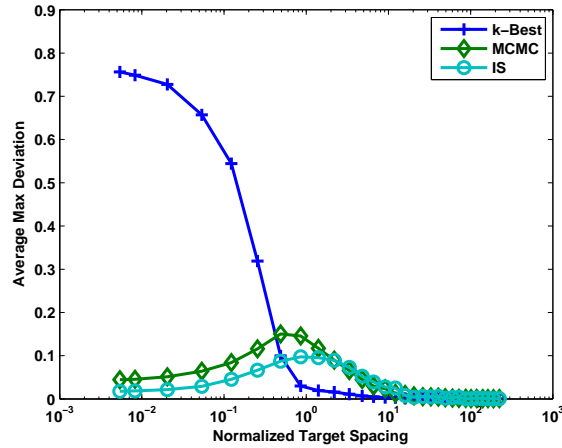


(b) Samples: k -best = 1000, IS = 10000, MCMC = 100000.

Figure 1. Average maximum deviation: 5 targets.



(a) Samples: k -best = 100, IS = 1000, MCMC = 10000.



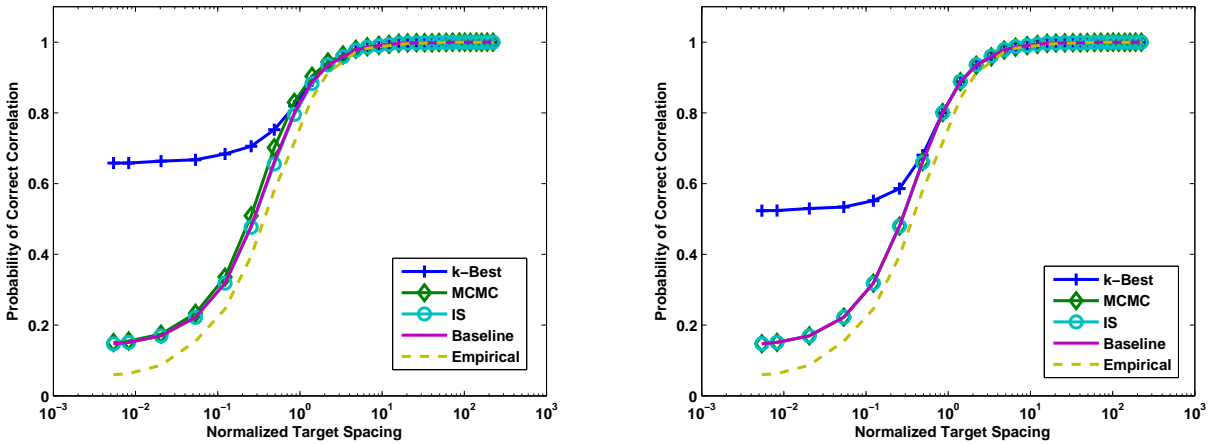
(b) Samples: k -best = 1000, IS = 10000, MCMC = 100000.

Figure 2. Average maximum deviation: 20 targets.

a ten-fold increase in k does not significantly change the curve for the k -best method; thus, while a costly increase in k is not likely to improve the performance of the k -best method for dense scenarios, the k -best method is more accurate down to problem densities slightly below unity. For the Monte Carlo methods, the importance sampler and MCMC are again competitive, with the importance sampler again having an advantage in accuracy.

Finally, it is noteworthy that for both the 5-target and 20-target scenarios, the Monte Carlo methods show apparent improvement in the convergence of the estimated ambiguity matrix to the baseline ambiguity matrix as the size of the surveillance region approaches the noise level. Since the noise level at this density is much higher than the average target separation, no hypothesis is any more likely than another, and the hypothesis probabilities approach a uniform distribution, as can be seen by examining the empirical curves of Figure 3. Since the importance sampling distribution and MCMC acceptance function are driven by relative likelihoods, these sampling methods will in effect mimic a uniform sampling distribution, which is appropriate for this situation.

The probability of correct association is displayed in Figure 3 for the 20-target case. Again, the baseline curve was computed using 10 million MCMC proposals. As was the case for Figures 1(a) and 2(a), Figure 3(a) corresponds to a setting of $k = 100$; however, Figure 3(b) uses $k = 10000$ rather than $k = 1000$. Note also that this metric aggregates the data more than the average maximum deviation metric does. In the left-hand plot, we see that both the importance sampling and MCMC methods are very close to the baseline curve, while the baseline, importance sampling, and MCMC



(a) Samples: k -best = 100, IS = 1000, MCMC = 10000. (b) Samples: k -best = 10000, IS = 100000, MCMC = 1000000.

Figure 3. Probability of correct association: 20 targets.

curves are superimposed in the right-hand plot. The k -best curve also shows good agreement with the baseline curve until the problem density overwhelms this method. This indicates that these three methods are indeed converging to the curve corresponding to the exact ambiguity. It is quite interesting to note that that failure of the k -best method does not occur until after the empirical curve has begun to deviate from the baseline curve, indicating that the k -best method is viable except for densities where the typical likelihood formulation is invalid.

Object Count	Per-Sample Runtime	
	IS	MCMC
5	2.21×10^{-6}	0.85×10^{-6}
20	11.12×10^{-6}	1.03×10^{-6}

Table 2. Average per-sample runtimes (seconds) for Monte Carlo ambiguity estimation methods.

We now turn our attention to runtime analysis of the three ambiguity estimation algorithms. Table 2 displays the average time per-sample required to generate importance samples and MCMC proposals for both the 5-target and 20-target cases. Note that while the average time required by the importance sampler increased noticeably for the 20-target case, the average time required to generate an MCMC proposal remained nearly constant. This occurs because our MCMC proposal generator alters at most a single association for each new hypothesis, whereas the independent samples generated by the importance sampler require reevaluation of all associations for each new hypothesis.

Figures 4 and 5 plot average runtime against average maximum deviation for a fixed density. In both figures, the left-hand plot corresponds to a normalized target spacing of 0.1, while the spacing for the right-hand plot is unity. Different runtime and error combinations were generated by sweeping the k -best settings over $k \in \{10, 30, 100, 300, 1000, 3000, 10000\}$. The importance samples sweep was $10k$, and MCMC proposals sweep was $100k$. The results are clear for the 5-target case in Figure 4: Regardless of target spacing, k -best enumeration exhibits superior performance, while the importance sampler is slightly better than MCMC. The 20-target results in Figure 5 are more interesting. In the right-hand plot, we again see that k -best enumeration maintains a solid advantage, while the importance sampler and MCMC are competitive, though the importance sampler exhibits better performance at lower runtimes. In the left-hand plot, the k -best method can no longer enumerate a sufficient number of hypotheses, resulting in significant performance degradation. In contrast, the importance sampler and MCMC remain effective.

In summary, it would appear that for small numbers of targets, the k -best method is uniformly superior, an advantage that is maintained for larger numbers of targets down to a normalized target spacing of slightly less than unity. Outside of these conditions, either importance sampling or MCMC can be used, though our results indicate somewhat better performance for the importance sampler.

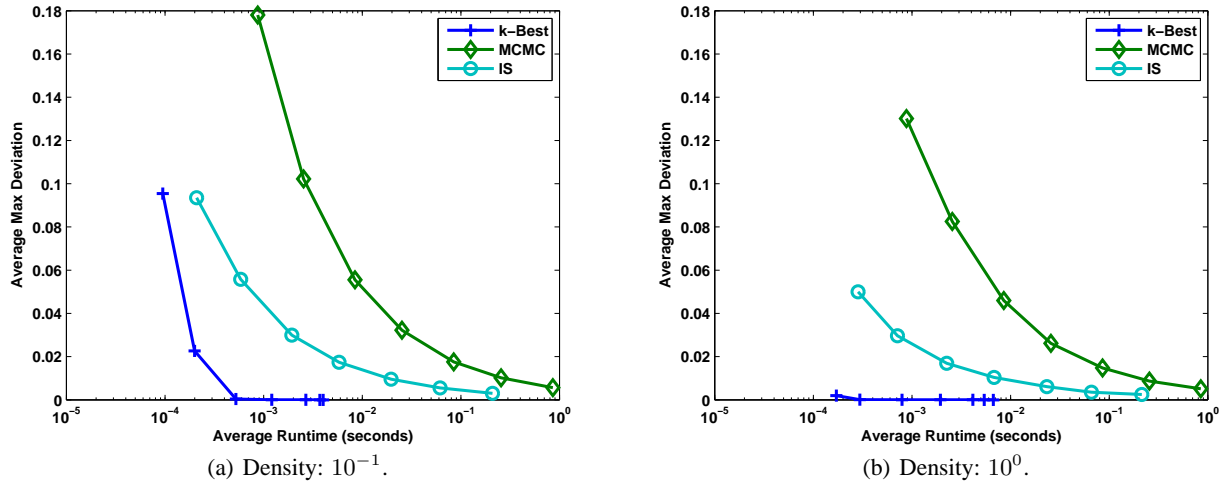


Figure 4. Average maximum deviation vs. runtime: 5 targets.

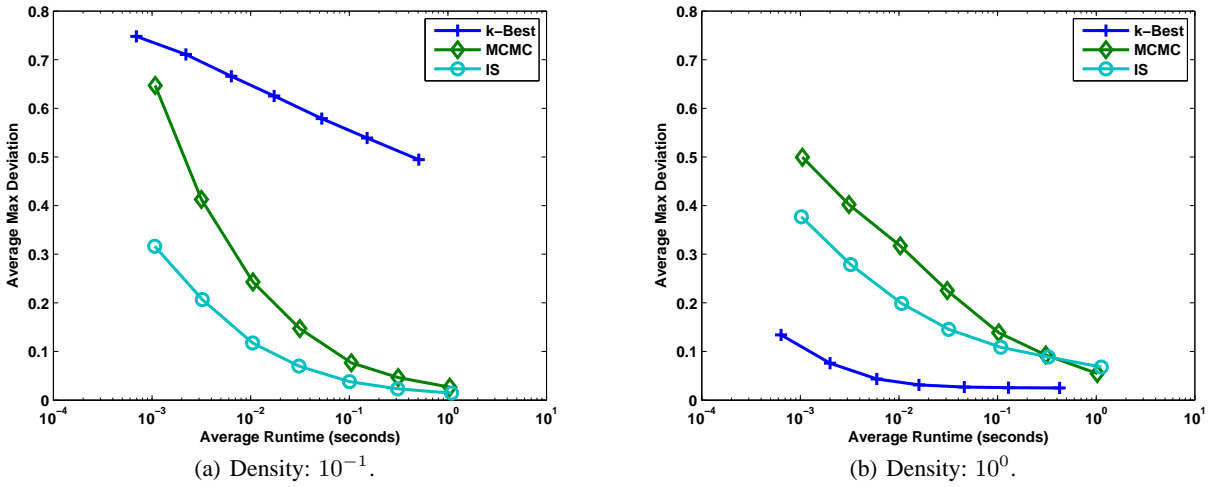


Figure 5. Average maximum deviation vs. runtime: 20 targets.

4. CONCLUSIONS

We evaluated three different techniques for online estimation of association probabilities in multitarget tracking: k -best enumeration, Markov Chain Monte Carlo, and importance sampling. We used a variety of object count and density configurations ranging from those that satisfied the assumptions of the multitarget likelihood model typically used in data association, to some that were very dense, with the more difficult configurations allowing us to evaluate the sensitivity of the algorithms to various operating conditions. Despite the inherent combinatorial limitations, k -best enumeration had the best runtime and accuracy performance throughout the range of densities commensurate with the typical multitarget likelihood formulation. There was significant, but steady, performance degradation only at target densities that violate the assumptions of the data association model. For very difficult problems, the Monte Carlo techniques were superior to k -best enumeration; both the importance sampler and MCMC were able to achieve arbitrarily accurate association probabilities with reasonable computational effort. In our tests, the importance sampler required generation of fewer samples than the MCMC approach, although the per-sample runtime was smaller for MCMC and did not increase appreciably with problem size. As the number of samples required for a given task increases, it appears that MCMC performance may eventually exceed that of the importance sampler. These results suggest that, depending on desired accuracy, MCMC is appropriate for problems involving large numbers of densely-spaced targets, while importance sampling is preferable for moderate numbers of densely-spaced targets, or for larger numbers of targets if they are not as densely spaced.

ACKNOWLEDGMENTS

This work was supported in part by SMDC Contract W9113M-06-C-0156.

REFERENCES

1. A. B. Poore, "Multidimensional assignment formulation of data association problems arising from multitarget tracking and multisensor data fusion," *Computational Optimization and Applications* **3**, pp. 27–57, 1994.
2. S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems*, Artech House, 1999.
3. S. Gadaleta, S. Herman, S. Miller, F. Obermeyer, B. J. Slocumb, A. B. Poore, and M. Levedahl, "Short-term ambiguity assessment to augment tracking data association information," in *Eighth International Conference on Information Fusion, Philadelphia, PA*, 2005.
4. R. Popp, K. Pattipati, and Y. Bar-Shalom, " m -best S -D assignment algorithm with application to multitarget tracking," *IEEE Transactions on Aerospace and Electronic Systems* **37**(1), pp. 22–39, 2001.
5. C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, Springer, 2004.
6. B. Kragel, S. Danford, and A. Poore, "MAP probabilities for data association and absolute bias estimation with an arbitrary number of sensors," Technical Report TR-07-001, Numerica Corporation, August 2007.
7. S. Oh, S. Russell, and S. Sastry, "Markov chain Monte Carlo data association for general multiple-target tracking problems," in *Proc. of the 43rd IEEE Conference on Decision and Control, Paradise Island, Bahamas, Dec. 2004*, **1**, pp. 735–742, 2004.
8. H. Pasula, S. J. Russel, M. Ostland, and Y. Ritov, "Tracking many objects with many sensors," in *Proceedings of the International Joint Conferences on Artificial Intelligence 1999 (IJCAI-99)*, pp. 1160–1161, 1999.
9. S. Mori and C. Chong, "Evaluation of a posteriori probabilities of multi-frame data association hypotheses," in *Signal and Data Processing of Small Targets*, O. E. Drummond, ed., **6699**, SPIE, (San Diego, CA), August 2007.
10. M. Jerrum and A. Sinclair, *Approximations for NP-Hard Problems*, ch. The Markov Chain Monte Carlo Method: An Approach to Approximate Counting and Integration, pp. 482–520. PWS Publishing, 1996.
11. R. M. Neal, "Probabilistic inference using Markov chain Monte Carlo methods," Technical Report CRG-TR-93-1, Dept. of Computer Science, University of Toronto, September 1993.
12. S. Oh and S. Sastry, "A polynomial-time approximation algorithm for joint probabilistic data association," in *American Control Conference (ACC)*, (Portland, OR), June 2005.
13. L. Tierney, *Markov Chain Monte Carlo in Practice*, ch. Introduction to General State-Space Markov Theory, pp. 59–74. Chapman & Hall / CRC, 1996.
14. D. J. MacKay, *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press, 2003.
15. S. Danford, B. Kragel, and A. Poore, "Joint MAP bias estimation and data association: Algorithms," in *Signal and Data Processing of Small Targets*, O. E. Drummond, ed., **6699**, SPIE, (San Diego, CA), August 2007.