Estimation of Information-Theoretic Quantities for Particle Clouds

Shankar Kulumani*
George Washington University, Washington, District of Columbia 20052
Islam I. Hussein†
Christopher W. T. Roscoe‡ and Matthew P. Wilkins§
Applied Defense Solutions, Columbia, Maryland 21044
Paul W. Schumacher, Jr.¶
Air Force Research Laboratory, Kihei, Hawaii 96753

When compared to alternative approaches such as Gaussian Mixture Models (GMMs), particle clouds more faithfully represent uncertainty. A concern about particle clouds, however, is their inability to provide the analyst with closed form expressions for many standard information theoretic quantities such as entropy and divergence. Recent advances in information theory have provided techniques that can approximately estimate such quantities. One approach in the literature is the use of the k-th nearest neighbor (k-NN) algorithm to, firstly, estimate the probability density function that the particle cloud represents. Given these density estimates, one can then compute various information theoretic quantities. In this paper we review the k-NN algorithm and then discuss two applications. The first application is the estimation of the entropy of a particle cloud. Specifically, we show that the entropy of a nonlinear Hamiltonian system is conserved if canonical coordinates are used as a coordinate frame. The second application is to estimate the divergence between two particle clouds. Specifically, we use the estimated Bhattacharyya divergence to solve an uncorrelated track (UCT) correlation problem.

I. Introduction

INFORMATION theoretic quantities are crucial to many space situational awareness (SSA) problems. Examples include the use of entropy to detect nonlinearity[1] and the use of the Bhattacharyya divergence to solve the uncorrelated track association problem.[2,3] In the case of continuous random variables, information theoretic quantities are typically expressed in terms of integrals, for which a closed form solution exists in only a handful of cases. One such special case is when the underlying probability density functions are normally distributed. In that case, both entropy and divergence are computable in closed form. We note here that for the case of Gaussian Mixture Models (GMMs), a closed form expression for entropy and divergence do not exist, but can only be obtained approximately.[4,5]

The problem is made more challenging when uncertainty is represented, not by an analytical density function or an approximation thereof, but by a particle cloud. In this case, one of the earliest methods is to use a histogram to approximate the probability density function (PDF) based on the particles.[6] Histograms, however, suffer from the fact that they are not smooth, and that the PDF estimate is sensitive to the bin parameter choice (bin end points and bin width). To avoid some of the pitfalls of histograms, kernel methods have been developed. In kernel methods, a set of non-negative functions that integrate to one and

*Graduate Student. Student Member AIAA.
†Aerospace Engineer; IHussein@applieddefense.com. Member AIAA.
‡Aerospace Engineer; CRoscoe@applieddefense.com. Member AIAA.
§Aerospace Engineer; MWilkins@applieddefense.com.
¶Principal Investigator, Astrodynamics for SSA. Associate Fellow AIAA.

have zero mean are used as basis functions to approximate the PDF based on independent and identically distributed samples drawn from the underlying unknown distribution.[7] Ref. 8 is one of the earliest papers that investigated the use of histograms and kernel density estimators to estimate the entropy of a particle sample, and a performance comparison was performed. One type of kernel density estimators is the k-NN algorithm.[9] We utilize the k-NN algorithm in this paper to estimate the density. Once the density estimate is obtained, one can then use the Monte-Carlo (MC) framework to compute the desired information theoretic quantity. We will provide a brief overview of the algorithm in this paper. Alternatively, one can use, for example, the expectation maximization (EM) algorithm[10] to convert a particle cloud into a GMM and then use approximate algorithms[4,5] to estimate information quantities from the generated GMM. We do not implement this approach in this paper. However, we will investigate the use of an EM-based GMM approach in future work and compared its performance to the k-NN algorithm.

Returning to information theoretic applications, in this paper we will apply the k-NN density estimator to estimate several information theoretic quantities that are of importance in SSA. Three main quantities are of interest. The first is information entropy. Entropy has been proposed in Ref. 1 as a means to detect non-linearity in the context of nonlinear filtering using GMMs. If one were to use particle techniques instead, one would then need a way to estimate the entropy of a particle cloud. In fact, we use a particle representation of the uncerlying PDF of a Hamiltonian system to show that in some very specific cases (e.g., when a system is Hamiltonian and canonical coordinates are used) entropy is conserved despite the nonlinear nature of the system. Another application of the k-NN algorithm is to estimate the divergence between two particle clouds. In Refs. 2, 3, the authors suggest the use of the Bhattacharyya divergence to solve the UCT problem. In that work, however, the authors operated entirely within the framework of the Unscented Kalman Filter (UKF), where the underlying PDFs were all assumed Gaussian. For long-duration UCT correlation problems, Gaussianity may be a very assumption and one has to use particle clouds to better represent uncertainty. In that case, one can use any of the methods described above to estimate the divergence between two particles. We will address this application in this paper.

One final application of density estimation, which is currently under investigation but that is not explored in this paper, is the estimation of mutual information. One can express mutual information in terms of entropy[11] and, hence, one can directly use entropy estimators to estimate mutual information. Mutual information was used in Refs. 12, 13 to solve the observation-to-observation data association problem, where it is desired to determine the associations given a set of observations, without any postulated track information to test against. In that case, the analyst is expected to determine the associations between observations, and also produce a statistical track for the object that generated them. In other words, the approach also solves the initial orbit determination (IOD).

The rest of the paper is organized as follows. We first summarize some of the basic information theoretic quantities of interest. We then briefly describe the k-NN algorithm and demonstrate its basic capabilities and caveats. We then describe its use to estimate information theoretic quantities, specifically: entropy and the Bhattacharyya divergence. We then present to SSA applications: (1) conservation of entropy for Hamiltonian systems and (2) the use of Bhattacharyya divergence to solve the UCT problem. We conclude the paper with a summary of our results, as well as current and future research directions.

II. Information-Theoretic Quantities

A. Entropy

For a given probability density function $p(\boldsymbol{x})$ in the random variable $\boldsymbol{x} \in \mathbb{R}^n$, the Shannon information entropy is given by (see Ref. 11):

$$H(p) = -\int p(\boldsymbol{x}) \log(p(\boldsymbol{x}))\mathrm{d}\boldsymbol{x} = -E[\log(p(\boldsymbol{x}))], \tag{1}$$

where $E[g(\boldsymbol{x})]$ is the expected value of the function $g(\boldsymbol{x})$.

A more general case is the continuous variable Rényе entropy given by (see Ref. 14, 15):

$$H_\alpha(p) = \frac{1}{1-\alpha} \log \int p^\alpha(\boldsymbol{x})\mathrm{d}\boldsymbol{x} = \frac{1}{1-\alpha} \log\left(E[p^{\alpha-1}(\boldsymbol{x})]\right), \tag{2}$$

for some parameter α. Note that in the limit as $\alpha \to 1$, the Rényi entropy is the Shannon entropy.

The utility of expressing the above integrals in terms of an expectation derives from the *Monte Carlo Principle*, which states that for a collection of independent and identically distributed random variables $\boldsymbol{x}_i$, $i = 1, \ldots, N$, drawn from the density $p(\boldsymbol{x})$ we have the Monte Carlo integration (see Ref. 16):

$$E[g] = \int g(\boldsymbol{x})p(\boldsymbol{x})\mathrm{d}\boldsymbol{x} \simeq \frac{1}{N}\sum_{i=1}^{N} g(\boldsymbol{x}_i) \tag{3}$$

Thus, we have

$$H(p) = -E[\log(p(\boldsymbol{x}))] \simeq -\frac{1}{N}\sum_{i=1}^{N} \log(p(\boldsymbol{x}_i)), \tag{4}$$

and

$$H_\alpha(p) = \frac{1}{1-\alpha}\log\left(E[p^{\alpha-1}(\boldsymbol{x})]\right) \simeq \frac{1}{1-\alpha}\log\left(\frac{1}{N}\sum_{i=1}^{N} p^{\alpha-1}(\boldsymbol{x}_i)\right). \tag{5}$$

These expressions are useful only if we have analytical expressions for the underlying PDFs $p(\boldsymbol{x})$. For the case when uncertainty is represented by a particle cloud, one has to approximate the underlying probability density function, and then computing the entropy from the above expressions.

B. Divergence

In general, one can consider the general class of Rényi information divergences. Analysis of other divergences, such as the Cauchy Schwartz divergence, can be handled similarly. Information divergence between two PDFs p_1 and p_2 in the random variable $\boldsymbol{x}$ is given by:[14]

$$D_\alpha(p_1||sp_2) = \frac{1}{\alpha-1}\log\int p_1(\boldsymbol{x})^\alpha p_2(\boldsymbol{x})^{1-\alpha}\mathrm{d}\boldsymbol{x}. \tag{6}$$

The Kullback-Leibler divergence is given by:

$$D_{\mathrm{KL}}(p_1||p_2) = \int p_1(\boldsymbol{x})\log\frac{p_1(\boldsymbol{x})}{p_2(\boldsymbol{x})}\mathrm{d}\boldsymbol{x} = E_{p_1}\left[\log\frac{p_1(\boldsymbol{x})}{p_2(\boldsymbol{x})}\right], \tag{7}$$

which we note is not symmetric in p_1 and p_2. We note that $D_{\mathrm{KL}} = \lim_{\alpha\to 1} D_\alpha$.

Another distance metric is given by the Bhattacharyya divergence:

$$D_{\mathrm{B}}(p_1||p_2) = -\log \mathrm{B}_C(p_1,p_2) = -\log\left[\int\sqrt{p_1(\boldsymbol{x})p_2(\boldsymbol{x})}\mathrm{d}\boldsymbol{x}\right]. \tag{8}$$

We can formulate this metric in terms of the expectation by reformulating Eq. (8) as follows,

$$\begin{aligned} D_{\mathrm{B}}(p_1||p_2) &= -\log\left[\int p_1(\boldsymbol{x})\sqrt{\frac{p_2(\boldsymbol{x})}{p_1(\boldsymbol{x})}}\mathrm{d}\boldsymbol{x}\right], \\ &= -\log\left[E_{p_1}\left[\sqrt{\frac{p_2(\boldsymbol{x})}{p_1(\boldsymbol{x})}}\right]\right]. \end{aligned} \tag{9}$$

As the Bhattacharyya divergence is symmetric, due to the product of p_1 and p_2, a similar result holds in the case of the expectation taken with respect to p_2. Expressing the divergence in terms of the expectation again allows us to apply Monte Carlo integration to approximate the value of the integral. Applying the Monte carlo princple to Eq. (7) and Eq. (8) results in

$$D_{\mathrm{KL}}(p_1||p_2) \approx \frac{1}{N}\sum_{i=1}^{N}\log\frac{p_1(\boldsymbol{x}_i)}{p_2(\boldsymbol{x}_i)}, \tag{10}$$

$$D_{\mathrm{B}}(p_1||p_2) \approx -\log\left[\frac{1}{N}\sum_{i=1}^{N}\sqrt{\frac{p_1(\boldsymbol{x}_i)}{p_2(\boldsymbol{x}_i)}}\right]. \tag{11}$$

If we can sample from either $p_1(\boldsymbol{x}_i)$ or $p_2(\boldsymbol{x}_i)$ (not necessarily both), we can then easily employ the above expression to obtain an estimate of the divergence. Otherwise, if we have a method of estimating the PDFs at any arbitrary value. We will discuss the k-NN algorithm next to estimate the PDF $\hat{p}_i(\boldsymbol{x})$, given a sample of particles drawn from the unknown distribution..

III. Nearest Neighbor Estimate

The nearest neighbor estimator[9] is a non-parametric method to estimate a PDF. We assume that we are given samples from some unknown distribution $p(\boldsymbol{x})$ and our objective is to generate an estimate $\hat{p}(\boldsymbol{x})$ without assuming a form of the unknown distribution. Unlike histogram based methods this approach is applicable and tractable for higher dimensional systems.

The probability that $\boldsymbol{x}$, drawn randomly from $p(\boldsymbol{x})$, lies in the set Ω is given by

$$P\left[\boldsymbol{x} \in \Omega\right] = \int_{\Omega} p(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}.$$

We now assume that we can draw N samples from the distribution, $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$. The probability that k samples from the total N lie in the set Ω is given by the binomial distribution

$$\begin{aligned} Q(k) &= \binom{N}{k} P^k \left(1-P\right)^{N-k}, \\ &= \frac{N!}{k!\left(N-k\right)!} P^k \left(1-P\right)^{N-k}. \end{aligned}$$

It is easy to show that the expected value and variance of $\frac{k}{N}$ are given by

$$E\left[\frac{k}{N}\right] = P, \quad Var\left[\frac{k}{N}\right] = \frac{P(1-P)}{N}.$$

In this manner, we estimate the probability as $P \approx \frac{k}{N}$, that a sample $\boldsymbol{x}$ lies in the set Ω. If we further assume that the set Ω is sufficiently small and that the PDF $p(\boldsymbol{x})$ does not vary much over the set Ω, we can define the following,

$$\int_{\Omega} p(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \approx P(\boldsymbol{x}) V,$$

where V is the volume of the set Ω. If we assume a sufficiently small volume and larger number of samples, and we can define an estimate of the PDF $\hat{p}(\boldsymbol{x})$ as

$$\hat{p}(\boldsymbol{x}) = \frac{k}{NV},$$

where V is the volume surrounding the test point $\boldsymbol{x}$, N is the total number of samples, and k the number of samples that lie within the volume V. For Euclidean spaces, $\boldsymbol{x} \in \mathbb{R}^d$, the kth nearest neighbor estimate is given by

$$\hat{p}(\boldsymbol{x}) = \frac{k}{N \cdot c_d \cdot \rho_k^d(\boldsymbol{x})}, \tag{12}$$

where $\rho_k(\boldsymbol{x}) = \|\boldsymbol{x} - \boldsymbol{x}_k\|$, the distance from the test point to the kth nearest neighbor, and c_d is the volume of the unit sphere in $\mathbb{R}^d$ given by

$$c_d = \frac{\pi^{\frac{d}{2}}}{\left(\frac{d}{2}\right)!}.$$

We demonstrate the application of the kth nearest neighbor estimator by generating an estimate of the normal distribution. We generate $N = 20000$ random samples from $p(x) = \mathcal{N}\left\{0, 1\right\}$ and $k = 200$, where $\mathcal{N}\left\{m, \sigma\right\}$ denotes a normal distribution with mean m and standard deviation σ. We compute $\hat{p}(x)$ over the entire domain which allows us to visualize the estimate as seen in Figure 1.

A. Information-Theoretic Quantity Estimation

With the framework to estimate a PDF in place, we are able to compute various information quantities for arbitrary point clouds. In general, it is not possible to have an exact analytical representation of the PDF of

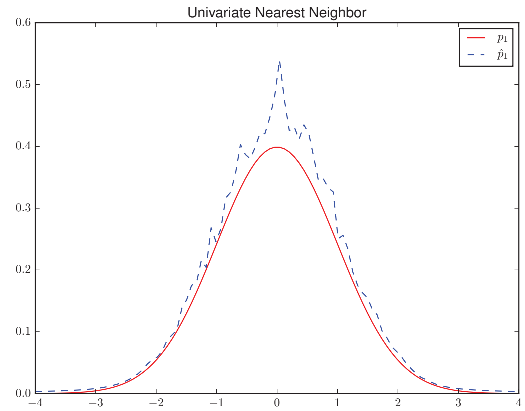


Figure 1. Univariate nearest neighbor estimate

a random system. Furthermore, any analytical approximation may not remain valid through the evolution of the state dynamics. As a result, it is more accurate to represent the PDF with a series of particles, forming a point cloud representation of any distribution function. From this point cloud, we seek to estimate several quantities, namely entropy and divergence between two arbitrary point clouds.

Using the tools laid out previously, we demonstrate the accuracy of computing the entropy, Kullback-Leibler divergence, and Bhattacharyya divergence for several sample scenarios. We first compute these information quantities for a univariate scenario. We define two unknown univariate Gaussian distributions:

$$\begin{aligned} p_1(x) &= \mathcal{N}\left\{1, 1\right\}, \\ p_2(y) &= \mathcal{N}\left\{-1, 1\right\}. \end{aligned} \tag{13}$$

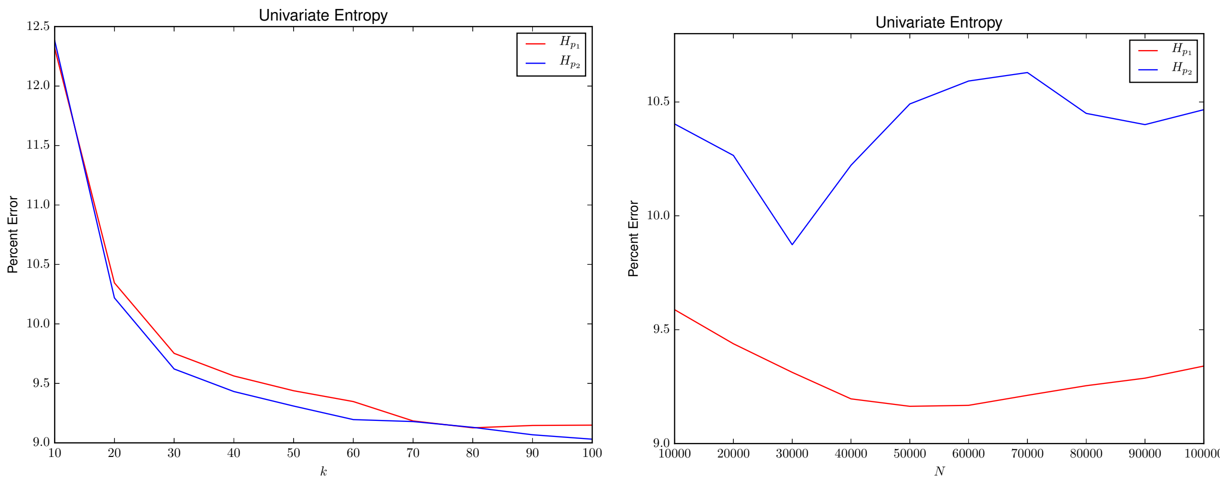
We compute the entropy of each distribution by applying Eq. (4) where we estimate the value of the PDF using Eq. (12). As the unknown distributions are Gaussian, there is a straightforward analytical expression for the entropy, which is given by

$$H(p_g) = \frac{1}{2} \log\left(2\sigma^2 \pi e\right). \tag{14}$$

We compute the percent error between the estimated entropy and the analytical result for a variety of values of k and N. Figure 2(a) shows that for a fixed sample size $N = 20000$, an increase in the value of k lowers the error of the entropy estimation. However, care must be taken with the selection of k as arbitrarily large values may have a detrimental effect on the estimation. In fact, depending on the problem parameters (such as dimension, sample size, etc.), there is an optimal k value. In this paper we do not implement an optimal k value. This will be implemented in a future publication. For a detailed discussion on the optimal choice of k, we refer the reader to Ref. 17. Figure 2(b) shows the effect of varying the sample size for a fixed value of $k = 50$. The univariate example is not sensitive to variations in the number of particles. However, for higher dimensional problems the number of particles plays a critical role in the estimation accuracy.

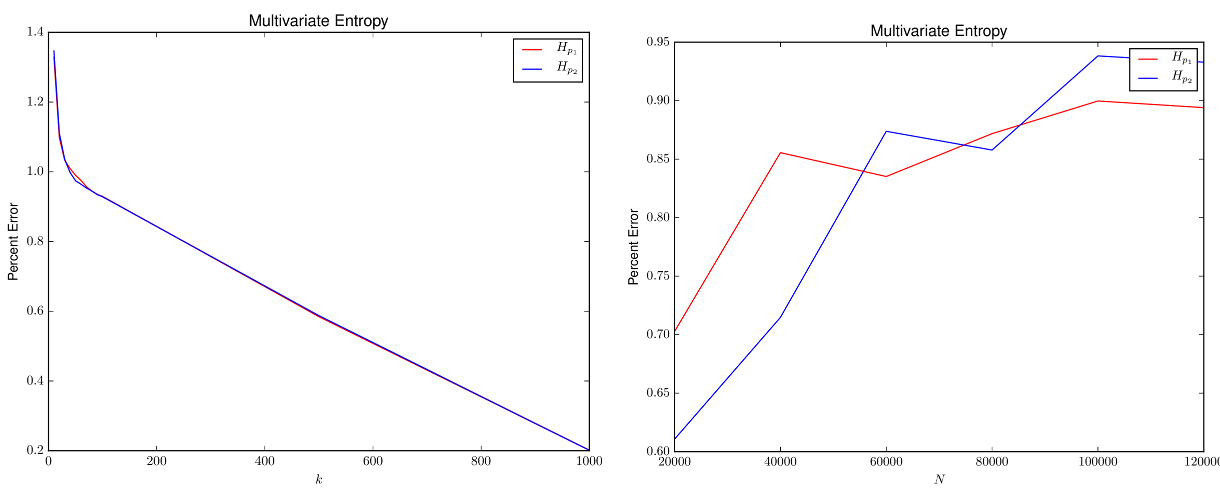
We perform a similar test with two six-dimensional multivariate Gaussian PDFs. We again compare the accuracy of the entropy estimation for a variety of values of k nearest neighbor and N sample size. Figure 3(a) shows the error given a fixed number of particles $N = 50000$. Figure 3(b) shows the error given a fixed nearest neighbor $k = 200$. In both cases, the estimate closely matches the true analytical value. The value of k should be chosen as a function of the system dimension and the total number of samples.

Next, we compute the Kullback-Leibler and Bhattacharyya divergence for both a univariate and multivariate Gaussians. The analytical expressions of the Kullback-Leibler and Bhattacharyya divergence between



(a) Sensitivity to kth nearest neighbor. **(b) Sensitivity to sample size N.**

Figure 2. Univariate Entropy Estimation.



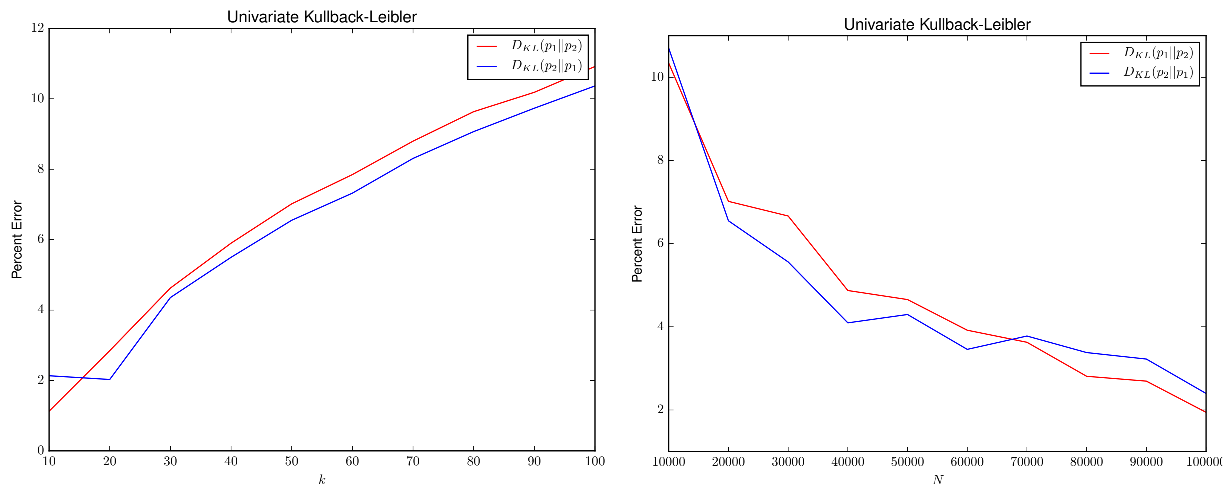
(a) Sensitivity to kth nearest neighbor. **(b) Sensitivity to sample size N.**

Figure 3. Multivariate Entropy Estimation.

$p_1 = \mathcal{N}\{\mu_1, \Sigma_1\}$ and $p_2 = \mathcal{N}\{\mu_2, \Sigma_2\}$ are given by

$$D_{KL}(p_1||p_2) = \frac{1}{2}\left[\log\frac{|\Sigma_2|}{|\Sigma_1|} - d + \mathrm{tr}(\Sigma_2^{-1}\Sigma_1) + (\mu_2 - \mu_1)^T\Sigma_2^{-1}(\mu_2 - \mu_1)\right],$$
$$D_B(p_1||p_2) = \frac{1}{8}(\mu_1 - \mu_2)^T\Sigma_{avg}^{-1}(\mu_1 - \mu_2) + \frac{1}{2}\log\left(\frac{\det(\Sigma_{avg})}{\det(\Sigma_1)\det(\Sigma_2)}\right).$$

Figure 4 shows the accuracy of the Kullback-Leibler divergence estimate for a univariate Gaussian example. In Figure 4(a), we vary the value of k in the KNN estimate while holding $N = 20000$. While in Figure 4(b), we vary the value of N while holding $k = 50$. In general, a larger number of samples results in a more accurate estimation, however arbitrarliy increasing k while holding N fixed results in more error. Figure 5 demonstrates the accuracy of the Kullback-Leibler divergence for a six-dimensional Gaussian. In Figure 5(a) we hold $N = 50000$ and vary k while in Figure 5(b) we hold $k = 200$ and vary N. The error is much larger due to the increased size of the state space. In some sense, each sample from the distribution offers less information as there are an increased number of degrees of freedom in the state space. As a result, higher dimensional system will require a larger sample size for a given accuracy. As mentioned earlier, the value of k should be adjusted according to the sample size.[17]



(a) Sensitivity to kth nearest neighbor.

(b) Sensitivity to sample size N.

Figure 4. Univariate Kullback-Leibler Estimation.

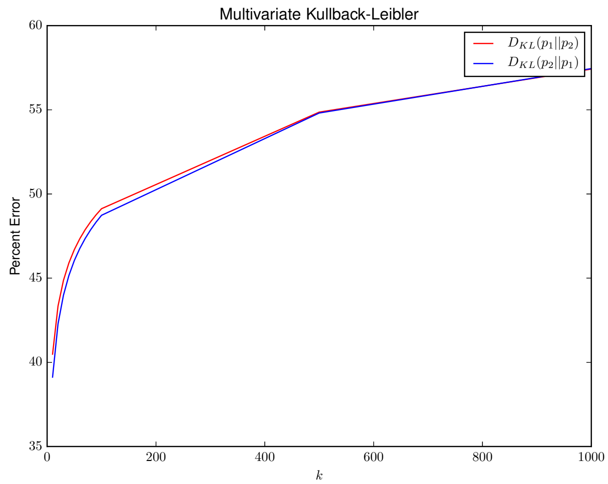
In Figure 6 and Figure 7, we perform a similar calculation and compute the Bhattacharyya divergence for the univariate and multivariate case. We determine the sensitivity of the computation of the Bhattacharyya divergence to variations in k and N by utilizing the same scenario as described previously for the Kullback-Leibler divergence. This analysis shows a similar behavior to the Kullback-Leibler results. The multivariate case is less accurate and furthermore a very large number of particles is required to achieve a similar level of accuracy as the univariate case.

As a final scenario, we determine the effect of dimensionality on the estimation of these information quantities. We hold $k = 50$ and $N = 20000$ and vary the dimension $d = \{1, 2, \ldots, 11, 12\}$. We define two Gaussian distributions by their mean, μ, and covariance, Σ, which are given by

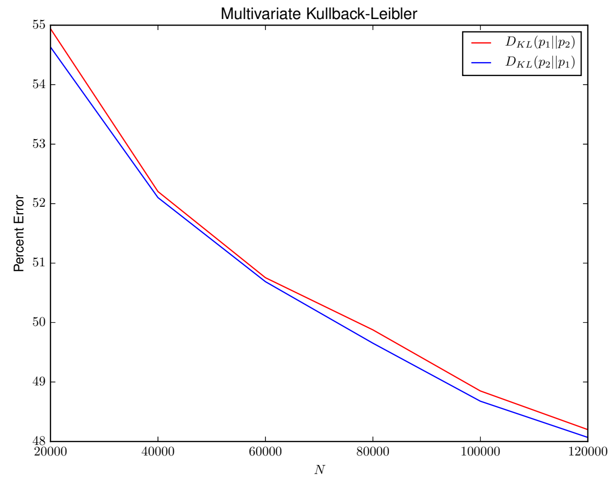
$$\mu_1 = \ ^{d},$$
$$\mu_1 = - \ ^{d},$$
$$\Sigma_1 = \Sigma_2 = \mathrm{diag}(\ ^{d}),$$

where ^{d} is defined as a d dimensional vector of ones and $\mathrm{diag}(\boldsymbol{x})$ generate a diagonal matrix with the vector $\boldsymbol{x}$ along the main diagonal.

In Figure 8(a) and Figure 8(b) we see that the divergence estimates become less accurate as the dimensionality grows. As a result, for higher dimensional systems, a corresponding increase in the number of

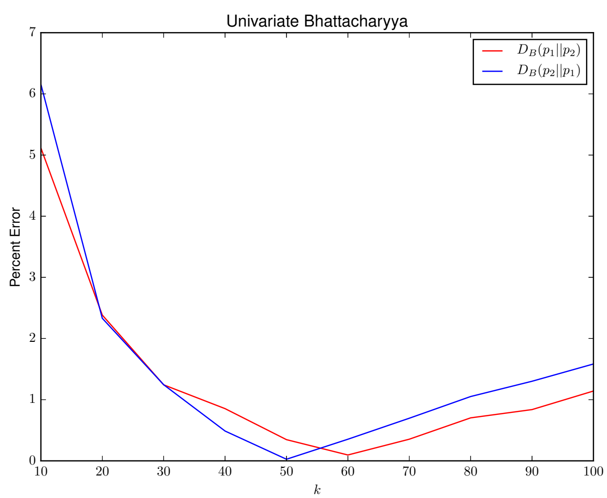


(a) Sensitivity to kth nearest neighbor.

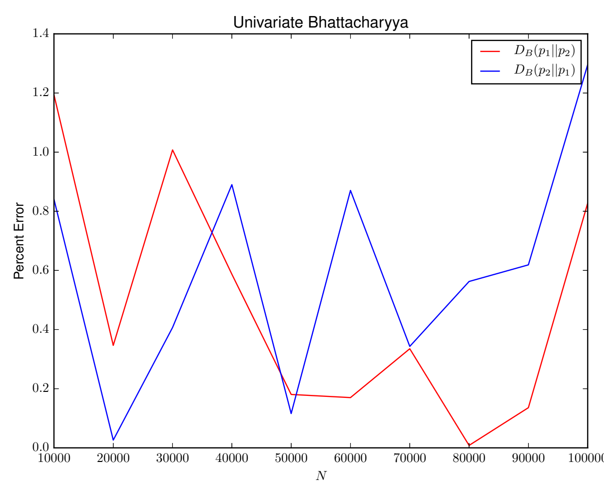


(b) Sensitivity to sample size N.

Figure 5. Multivariate Kullback-Leibler Estimation.

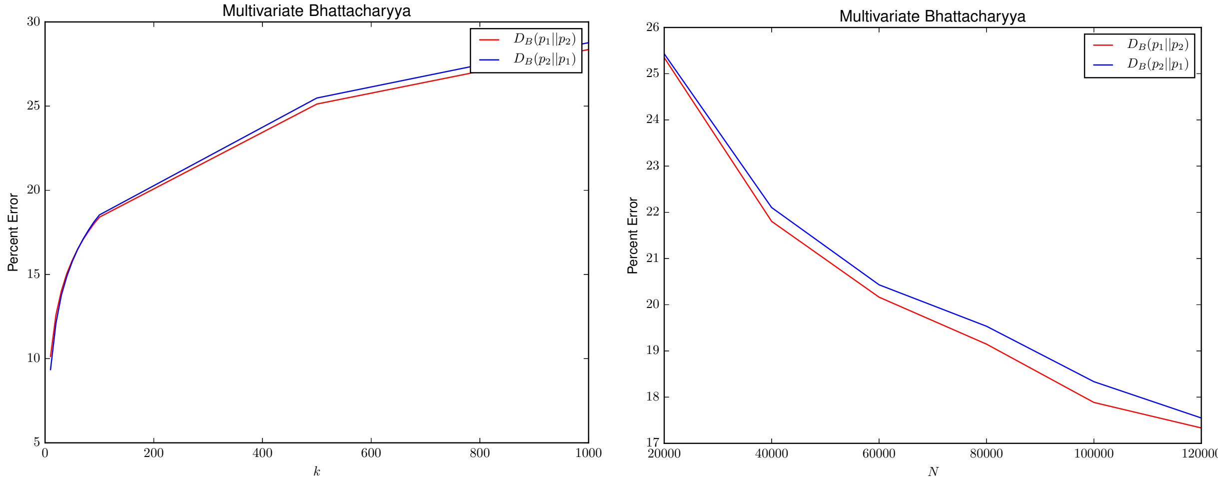


(a) Sensitivity to kth nearest neighbor.



(b) Sensitivity to sample size N.

Figure 6. Univariate Bhattacharyya Estimation.



(a) Sensitivity to kth nearest neighbor.

(b) Sensitivity to sample size N.

Figure 7. Multivariate Bhattacharyya Estimation.

particles is requred. Figure 8(c) shows that the entropy estimation is not sensitive to changes in dimensionality. However, there is interesting behavior between odd and even dimensioned systems, which is something we are currently investigating.

IV. Two SSA Applications

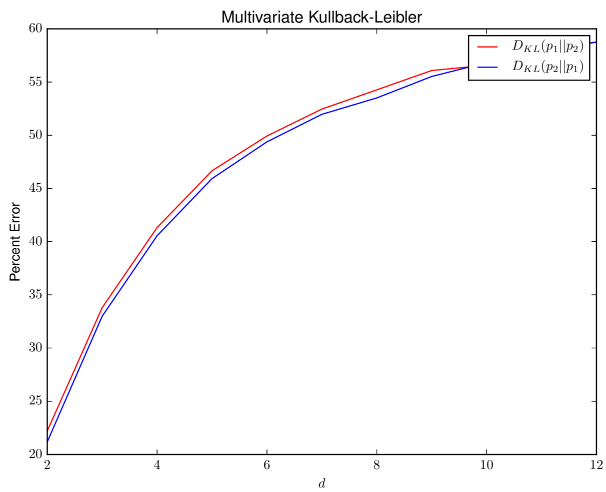
A. Conservation of Entropy for Hamiltonian Systems

In this section, we use the k-NN based entropy estimate to demonstrate that entropy is conserved for a Hamiltonian system when canonical coordinates are used to parametrize the system. While a more extensive analysis of the conservation of entropy will be published by the authors soon in an archival journal, here we present the basic result. Due to the conservation of volume in phase space of Hamiltonian systems, one can show using Louville Theorem that entropy for such systems is conserved. As a brief demonstration, we consider an object in a zero eccentricity, inclined (97.9 degrees) LEO orbit. The initial uncertainty in the orbit was chosen to be Gaussian such that the initial standard deviation in position was 10 m and 0.1 m/s for velocity (uniformly in all three Cartesian directions). Ten thousand particles were generated from this normal distribution. We emphasize that entropy is conserved only in canonical coordinates. Since position/velocity are not canonical, we transformed the particles' definition to Delaunay orbital elements, which are canonical. Using the k-NN algorithm, the particle cloud in Delaunay elements was used to approximate the underlying probability density function and the initial entropy was found to be 44.90855570500575. The uncertainty was propagated in Delaunay elements for 30 days. The resulting cloud was again used to approximate the underlying propagated density function. The final entropy was estimated to be 44.90800388940583, which is about $5.5 \times 10^{-2}\%$ error. Further investigation of the properties of entropy conservation need to be conducted, but this demonstrates (a) the use of the k-NN algorithm to estimate entropy and (b) that entropy is conserved for Hamiltonian systems.

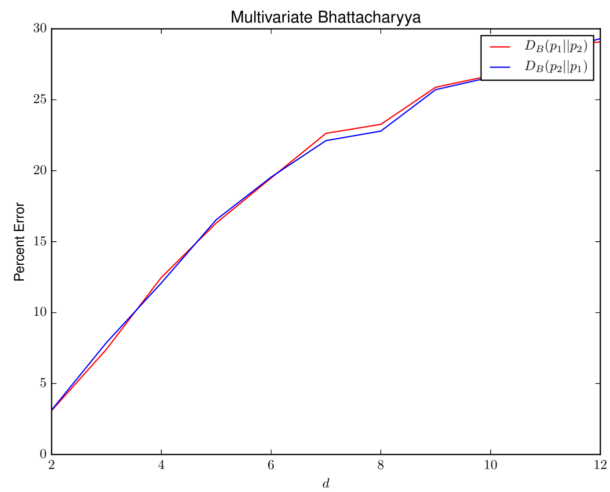
B. Using Bhattacharyya Divergence for UCT Correlation

As discussed in the introduction, using the UKF for long-time propagation of uncertainty looses its fidelity to the true probability distribution. Particle clouds, on the other hand, better represent the true probability distribution. In this section, we revisit the track-to-track association problem, in which we seek to correlate a collection of UCTs to a collection of propagated tracks. For more details on these problems, please see Ref. 2,3.

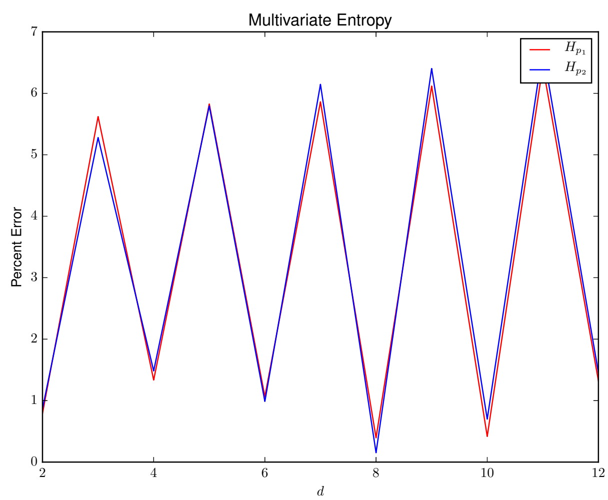
In this example, we consider two MEO objects in proximity of each other. The locations of the objects are normally randomly selected at some time t_* (chosen to be 30 minutes before midnight on January 1,



(a) Kullback-Leibler divergence estimation.



(b) Bhattacharyya divergence estimation.



(c) Entropy estimation.

Figure 8. Sensitivity of estimation to varying dimensions.

2004). Both objects have the same nominal mean position and velocity value but with various standard deviations. The nominal mean values are given in orbital elements in Table 1 for the orbital case considered, while the separation position from that mean value are varied. For position, it is varied from a value of 5 km to 200 km, while the velocity standard deviation is fixed at 0.5 m/s. Given the statistical nature of the initial locations of the RSOs, to measure performance we perform 1000 Monte Carlo runs per position standard deviation value. The average true positive correlation for both methods was recorded for each position standard deviation value and is used to compare the performance of the information divergence and CBTA solutions.

Table 1. Parameters of the True Orbit

Parameter	MEO Orbit Parameter Values
Semimajor Axis, km	26 600.0
Eccentricity	0.2
Inclination, deg	55.0
Argument of Perigee, deg	-120.0
Right Ascension of the Ascending Node, deg	207.0
True Anomaly, deg	20.0

Once the actual initial position and velocity for each object was selected at time t_*, the initial uncertainty used for the uncertainty propagation step was 35 km in position and 1.0 m/s in velocity (both being standard deviations uniformly assigned in all three Cartesian directions). One thousand particles were generated for each track and propagated to time t^*, where two UCTs are given. The time t^* was set to be 30 days after the initial propagation time t_*. At that point in time, an estimate of the Bhattacharyya divergence is computed using the k-NN algorithm, as discussed earlier in the paper. Only two-body dynamics are considered. In parallel to this, the true tracks were propagated forward in time for 30 days to time t^*. At which point we simulate an observation process that is equal to the true object state with an added zero mean Gaussian noise signal. The covariance for the observation process assumed a 10 km standard deviation in position (same for all three Cartesian directions) and 1 m/s for velocity (same for all three Cartesian directions), with no cross-correlation between position and velocity. This covariance matrix is denoted by $\mathbf{\Omega}_j$. Thus the mean of the UCTs is given by the actual measurement $\boldsymbol{\eta}_j$ and covariance is $\mathbf{\Omega}_j$.

The k-NN based estimate of the Bahttacharyya divergence solution was implemented in Monte Carlo simulations, as mentioned above. The result is shown in Figure 9. As can be seen in the figure, the Bhattacharyya divergence improves as the proximity between the initial two objects increases. As the separation distance between the two objects increases, we note that the true positive rate increases as well.

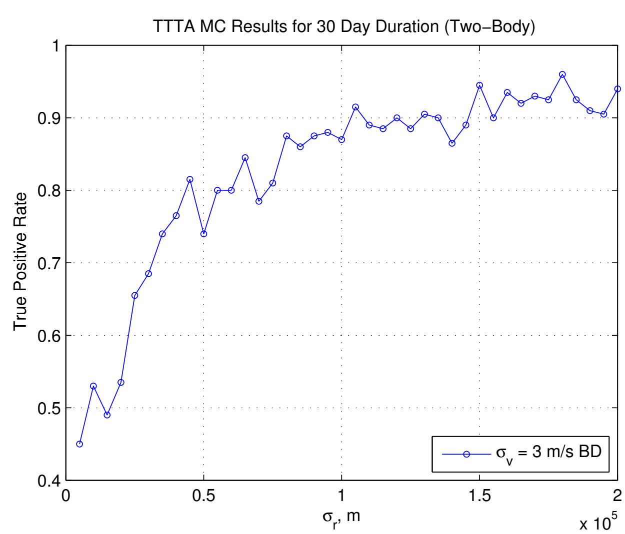


Figure 9. UCT Correlation results over a 30 day propagation period.

V. Conclusion

In this paper we reviewed some results from the information fusion literature on how to compute various types of information theoretic quantities from a particle cloud. Specifically, we focused on entropy and divergence estimation, which have several applications in SSA. We discussed some of the pitfalls of the main algorithm discussed in this paper, the k-NN algorithm, as well as some of the disadvantages of alternate techniques such as histogram methods. An alternative approach to the k-NN technique is to use an EM algorithm (or, alternatively, a k-means algorithm[18]) to approximate the PDF underlying the particle cloud. In future work, we will compare such GMM approximation approaches to the k-NN algorithm. Additionally, we plan on more fully implementing these approaches to existing SSA problems such as the various data association problems: track-to-track association (i.e., the UCT correlation problem), observation-to-observation association (i.e., the IOD problem), and the classical observation-to-track association problem.

References

[1] DeMars, K. J., Bishop, R., and Jah, M. K., "Entropy-Based Approach for Uncertainty Propagation of Nonlinear Dynamical Systems," *Journal of Guidance, Control, and Dynamics*, Vol. 36, No. 4, July 2013, pp. 1047–1057.

[2] Hussein, I. I., Roscoe, C. W. T., Wilkins, M. P., and Schumacher, Jr., P. W., "Track-to-Track Association Using Bhattacharyya Divergence," *Proceedings of the Advanced Maui Optical and Space Surveillance Technologies Conference*, Wailea, HI, September 15–18 2015.

[3] Hussein, I. I., Roscoe, C. W. T., Wilkins, M. P., and Schumacher, Jr., P. W., "UCT Correlation using the Bhattacharyya Divergence," *Space Flight Mechanics Meeting*, Napa, CA, February 14–18 2016.

[4] Huber, M. F., Bailey, T., Durrant-Whyte, H., and Hanebeck, U. D., "On entropy approximation for Gaussian mixture random vectors," *Multisensor Fusion and Integration for Intelligent Systems, 2008. MFI 2008. IEEE International Conference on*, Aug 2008, pp. 181–188.

[5] Hershey, J. R. and Olsen, P. A., "Approximating the Kullback Leibler Divergence Between Gaussian Mixture Models," *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*, Vol. 4, April 2007, pp. IV–317–IV–320.

[6] Scott, D. W., "On Optimal and Data-Based Histograms," *Biometrika*, Vol. 66, No. 3, 1979, pp. 605–610.

[7] Silverman, B. W., *Density Estimation for Statistics and Data Analysis*, Chapman and Hall/CRC, 1998.

[8] Hall, P. and Morton, S. C., "On the estimation of entropy," *Annals of the Institute of Statistical Mathematics*, Vol. 45, No. 1, 1993, pp. 69–88.

[9] Singh, H., Misra, N., Hnizdo, V., Fedorowicz, A., and Demchuk, E., "Nearest Neighbor Estimates of Entropy," *American Journal of Mathematical and Management Sciences*, Vol. 23, No. 3-4, 2003, pp. 301–321.

[10] Figueiredo, M. A. T. and Jain, A. K., "Unsupervised Learning of Finite Mixture Models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, 2002, pp. 381–396, 3.

[11] Cover, T. M. and Thomas, J. A., *Elements of Information Theory*, Wiley-Interscience, 1991.

[12] Hussein, I. I., Roscoe, C. W. T., Wilkins, M. P., and Schumacher, Jr., P. W., "Information Theoretic Criteria for Observation-to-Observation Association," *Proceedings of the Advanced Maui Optical and Space Surveillance Technologies Conference*, Wailea, HI, September 9–12 2014.

[13] Hussein, I. I., Roscoe, C. W. T., Wilkins, M. P., and Schumacher, Jr., P. W., "On Mutual Information for Observation-to-Observation Association," *Space Flight Mechanics Meeting*, Williamsburg, VA, January 11–15 2015.

[14] Rényi, A., "On measures of entropy and information," *Proceedings of the fourth Berkeley Symposium on Mathematics, Statistics and Probability 1960*, 1961, pp. 547–561.

[15] DeMars, K. J., Bishop, R. H., and Jah, M. K., "A Splitting Gaussian Mixture Method for the Propagation of Uncertainty in Orbital Mechanics," *Advances in the Astronautical Sciences*, Vol. 140, 2012, (Proceedings of the 21st AAS/AIAA Spaceflight Mechanics Meeting, New Orleans, LA, February 13–17 2011, Paper AAS 11-201).

[16] Andrieu, C., de Freitas, N., Doucet, A., and Jordan, M. I., "An Introduction to MCMC for Machine Learning," *Machine Learning*, Vol. 50, No. 1, 2003, pp. 5–43.

[17] Sricharan, K., Raich, R., and Hero, A., "Estimation of non-linear functionals of densities with confidence," *IEEE Trans on Information Theory*, Vol. 58, No. 7, 2012, pp. 4135–4159.

[18] Lloyd, S., "Least squares quantization in PCM," *IEEE Transactions on Information Theory*, Vol. 28, No. 2, Mar 1982, pp. 129–137.