

Determining Type I and Type II Errors when Applying Information Theoretic Change Detection Metrics for Data Association and Space Situational Awareness

Matthew P. Wilkins, Eamonn J. Moyer, and Islam I. Hussein
Applied Defense Solutions, Columbia, Maryland

Paul W. Schumacher, Jr.
Air Force Research Laboratory, Kihei, Hawaii

ABSTRACT

Correlating new detections back to a large catalog of resident space objects (RSOs) requires solving one of three types of data association problems: observation-to-track, track-to-track, or observation-to-observation. The authors previous work has explored the use of various information divergence metrics for solving these problems: Kullback-Leibler (KL) divergence, mutual information, and Bhattacharyya distance. In addition to approaching the data association problem strictly from the metric tracking aspect, we have explored fusing metric and photometric data using Bayesian probabilistic reasoning for RSO identification to aid in our ability to correlate data to specific RS Os. In this work, we will focus our attention on the KL Divergence, which is a measure of the information gained when new evidence causes the observer to revise their beliefs. We can apply the Principle of Minimum Discrimination Information such that new data produces as small an information gain as possible and this information change is bounded by ϵ . Choosing an appropriate value for ϵ for both convergence and change detection is a function of your risk tolerance. Small ϵ for change detection increases alarm rates while larger ϵ for convergence means that new evidence need not be identical in information content. We need to understand what this change detection metric implies for Type I α and Type II β errors when we are forced to make a decision on whether new evidence represents a true change in characterization of an object or is merely within the bounds of our measurement uncertainty. This is unclear for the case of fusing multiple kinds and qualities of characterization evidence that may exist in different metric spaces or are even semantic statements. To this end, we explore the use of Sequential Probability Ratio Testing where we suppose that we may need to collect additional evidence before accepting or rejecting the null hypothesis that a change has occurred. In this work, we will explore the effects of choosing ϵ as a function of α and β . Our intent is that this work will help bridge understanding between the well-trodden grounds of Type I and Type II errors and changes in information theoretic content.

1. INTRODUCTION

Building and maintaining a large catalog of resident space objects (RSOs) is computationally challenging. This is becoming even more challenging as new sensors come online allowing us to detect and track ever more minute space debris objects. Correlating new detections back to the existing catalog in a statistically rigorous manner requires solving one of three types of data association problems: the observation-to-track association (OTTA) problem, the track-to-track association (TTTA) problem, and the observation-to-observation association (OTOA) problem. The authors previous work [1–4] has explored the use of various information divergence metrics for solving these problems: the Kullback-Leibler divergence [5], mutual information [6], and the Bhattacharyya distance [7].

In addition to approaching the data association problem strictly from the metric tracking aspect, we have explored fusing metric and photometric data using Bayesian probabilistic reasoning for RSO identification to aid in our ability to correlate data to specific RSOs. [8, 9] Employing hierarchies as a world model for organizing and categorizing RSOs coupled with probabilistic programming allows one to update your prior probability for recognizing/identifying an RSO with a variety of evidence types (soft, hard, etc). Once you have generated your posterior distribution, we would like to notify a space operator of any “significant” change in the characterization hypotheses based upon the supplied evidence. The question becomes how to quantitatively decided if the change between the prior and posterior distributions is statistically meaningful. Should one wait for more data to confirm the change? If you decide to create a change detection alert, how do you assess whether your alert is within a desired false alarm percentile?

While we explored many different metrics to quantify the divergence between two distributions, in this case, we will focus our attention on the KL Divergence. The KL Divergence, also called the discrimination information, is a measure

of the difference between two probability distributions P and Q and is always positive semi-definite.

$$D_{KL}(P \parallel Q) = \sum_i P(i) \log \left[\frac{P(i)}{Q(i)} \right] \geq 0 \quad (1)$$

For the case when P and Q represent the prior and posterior beliefs regarding the identity of an RSO for characterization and/or data association purposes, respectively, the Kullback-Leibler divergence, denoted as $D_{KL}(P \parallel Q)$ in (1), is a measure of the information gained when new evidence causes the observer to revise their beliefs. In order to use the KL Divergence as a method of information theoretic change detection (or conversely convergence), we can apply the Principle of Minimum Discrimination Information (MDI). [10, 11] That is to say, given new facts, a new distribution Q should be chosen which is as hard to discriminate from the original distribution P as possible so that the new data produces as small an information gain as possible. We can now bound this information change by some small ϵ :

$$0 \leq D_{KL}(P \parallel Q) \leq \epsilon \quad (2)$$

Choosing an appropriate value for ϵ for both convergence and change detection is a function of your risk tolerance. Small ϵ for change detection increases alarm rates while larger ϵ for convergence means that new evidence need not be identical in information content. We need to understand what this change detection metric implies for Type I and Type II errors when we are forced to make a decision on whether new evidence represents a true change in characterization of an object or is merely within the bounds of our measurement uncertainty. This is unclear for the case of fusing multiple kinds and qualities of characterization evidence that may exist in different metric spaces or are even semantic statements.

2. SEQUENTIAL PROBABILITY RATIO TESTING

To this end, we explore the use of Wald's Sequential Probability Ratio Testing (SPRT) [12] where we suppose that we may need to collect additional evidence before accepting or rejecting the null hypothesis that a change has occurred. The idea behind sequential testing is that we collect sets of observations one at a time and then choose between the following options:

1. Accept the null hypothesis H_0 , and stop observation
2. Reject the null hypothesis H_0 , select H_1 and stop observation
3. Defer decision until we have collected another piece of information

The challenge is now to find out when to choose which of the above options. We would want to control the Type I (α) and Type II errors (β)

$$\alpha = P(\text{Deciding for } H_1 \text{ when } H_0 \text{ is true}) \quad (3)$$

and

$$\beta = P(\text{Deciding for } H_0 \text{ when } H_1 \text{ is true}) \quad (4)$$

Recall that the log-likelihood ratio test (LRT) [13] has a critical region of the following form:

$$\Lambda_n = \log \left[\frac{L(H_0; X_1, \dots, X_n)}{L(H_1; X_1, \dots, X_n)} \right] > K \quad (5)$$

In particular, Wald's SPRT has the following form:

1. If $\Lambda_n > b$, decide that H_1 is true and stop;
2. If $\Lambda_n < a$, decide that H_0 is true and stop;
3. If $a < \Lambda_n < b$, collect another observation to obtain $\Lambda_n + 1$.

where a and b ($a < 0 < b < \infty$) depend on the desired Type I and Type II errors, α and β . They may be chosen as follows:

$$a \approx \log\left(\frac{\beta}{1-\alpha}\right) \quad b \approx \log\left(\frac{1-\beta}{\alpha}\right) \quad (6)$$

3. SPRT AS APPLIED TO THE PRINCIPLE OF MINIMUM DISCRIMINATION INFORMATION

Unfortunately, the direct implementation of the SPRT to our application would result in long tests requiring large sample sizes that may not be achieved when tracking RSOs. In general, the amount of available observation data for characterization purposes will be variable and of unpredictable length. In some cases, we might have only one observation while, in other cases, we will have extremely dense characterization data.

Therefore, we implemented a truncated SPRT (TSPRT). [14, 15] Much like the above, a TSPRT chooses between the acceptance or rejection of the null hypothesis, or defers its decision until further observations are collected. Unlike the SPRT, if a decision is not made within a certain number of samples then at iteration \hat{K} the choice between hypotheses H_0 and H_1 depends on the value of the sum of observed realizations compared to a threshold $\hat{\tau}$. \hat{K} and $\hat{\tau}$ are related to the errors α and β through:

$$\hat{K} = [\Phi^{-1}(c_0\alpha) + \Phi^{-1}(c_1\beta)]^2 \left(\frac{\sigma}{\mu_1 - \mu_0}\right)^2 \quad (7)$$

$$\hat{\tau} = \sqrt{\hat{K}} [\mu_1\Phi^{-1}(c_0\alpha) + \mu_0\Phi^{-1}(c_1\beta)] \left(\frac{\sigma}{\mu_0 - \mu_1}\right) \quad (8)$$

where c_0 and c_1 are mixing values between zero and one that are chosen by the user. The values μ_0 , μ_1 , and σ are known functions of the hypotheses being tested. In the case of MDI for change detection, the hypotheses are as follows:

$$H_0 : D_{KL}(P \parallel Q) \text{ has mean } \theta_0 \text{ where } \theta_0 \leq \text{MDI} \quad (9)$$

$$H_1 : D_{KL}(P \parallel Q) \text{ has mean } \theta_1 \text{ where } \theta_1 > \text{MDI}. \quad (10)$$

Along with the sample variance, s^2 , μ_0 , μ_1 , and σ are determined through the means θ_0 and θ_1 by:

$$\mu_0 = -\frac{(\theta_1 - \theta_0)^2}{2s^2} \quad (11)$$

$$\mu_1 = \frac{(\theta_1 - \theta_0)^2}{2s^2} \quad (12)$$

$$\sigma^2 = \frac{(\theta_1 - \theta_0)^2}{s^2} \quad (13)$$

4. SIMULATION SETUP AND RESULTS

For our simulation setup, we generated a set of characterization hypotheses consisting of two possible categorizations $\mathbf{p}(t_i) = [p_1(t_i), p_2(t_i)]$, where we *a priori* are certain of the identity of the RSO as belonging to category 1. That is to say $\mathbf{p}(t_0) = [1, 0]$. We supplied soft evidence that the RSO was actually of category 2 which caused the probabilistic reasoning engine to update the Bayesian belief as shown in Figure 1. The hypotheses can be represented as a sequence of priors as follows:

$$\{\mathbf{p}(t_i) \mid \mathbf{p} \in \mathbf{p}_0, \mathbf{p}_1, \dots, \mathbf{p}_n\} \quad (14)$$

We can compute the KL Divergence from 1 and generate the set of divergences as shown in Figure 2:

$$\{D_{KL}(t_i) \mid D_{KL} \in D_{KL}(t_1), \dots, D_{KL}(t_n)\} \quad (15)$$

Applying the TSPRT above, we were able to determine the mean MDI for this particular set of priors and test for statistically significant change. Notice that in Figures *fig:case4_probVsTime* and *fig:case4_DivVsTime*, the KL Divergences surpass the 5%, $\beta = 10\%$ level. The user selectable mixing parameters c_1 and c_2 can be adjusted such that the MDI threshold is moved closer to the knee in the probability curve and is effectively a way to control the sensitivity of the TSPRT.

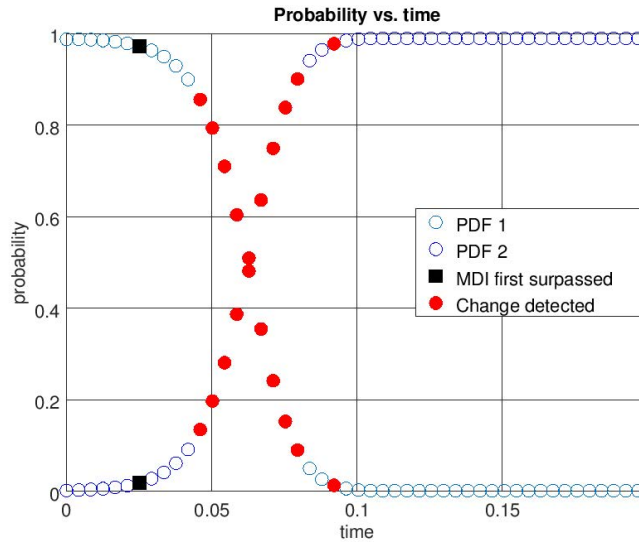


Fig. 1. Probabilities versus time showing the point at which the MDI was first surpassed and where statistically significant change was detected.

5. CONCLUSION

In this paper, we have developed a method to connect the KL Divergence and the principal of Minimum Discrimination Information to traditional false alarm and false negative rates. MDI is an extremely sensitive information change metric that allows us to detect even the most minute change in information content. However, the change detection threshold needs to be adjusted to correspond desired false alarm rate. By utilizing a truncated sequential probability ratio test, we can determine when posterior characterization distributions have changed in a statistically meaningful fashion.

REFERENCES

- [1] I. I. Hussein, C. W. T. Roscoe, M. P. Wilkins, and P. W. Schumacher, Jr., "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.
- [2] I. I. Hussein, C. W. T. Roscoe, M. P. Wilkins, and P. W. Schumacher, Jr., "On Mutual Information for Observation-to-Observation Association," *Space Flight Mechanics Meeting*, Williamsburg, VA, January 11–15 2015.
- [3] I. I. Hussein, C. W. T. Roscoe, M. P. Wilkins, and J. Paul W. Schumacher, "Track-to-Track Association Using Information Theoretic Criteria," *Proceedings of the AAS Astrodynamics Specialist Conference*, Vail, CO, August 10-14 2015. AAS 15-583.
- [4] I. I. Hussein, C. W. T. Roscoe, M. P. Wilkins, and J. Paul W. Schumacher, "Track-to-Track Association Using Bhattacharyya Divergence," *Proceedings of the Advanced Maui Optical and Space Surveillance Technologies Conference*, Wailea, HI, 2015.
- [5] S. Kullback and R. A. Leibler, "On information and sufficiency," *Annals of Mathematical Statistics*, Vol. 22, 1951, pp. 79–86.
- [6] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY: Wiley-Interscience, 1991.
- [7] A. Bhattacharyya, "On a measure of divergence between two statistical populations defined by their probability distributions," *Bulletin of the Calcutta Mathematical Society*, Vol. 35, 1943, pp. 99–109.
- [8] M. P. Wilkins, A. Pfeffer, B. Ruttenberg, P. W. Schumacher, and M. K. Jah, "Implications of Hierarchies for RSO Characterization and Identification," *Proceedings of the 24th International Symposium on Space Flight Dynamics*, Laurel, MD, May 2014.
- [9] B. Ruttenberg, M. P. Wilkins, and A. Pfeffer, "Reasoning on Resident Space Object Hierarchies Using Probabilistic Programming," *Proceedings of the 18th International Conference on Information Fusion*, Washington, D.C., 2015.
- [10] S. Kullback, *Information theory and statistics*. Wiley and Sons, 1959.

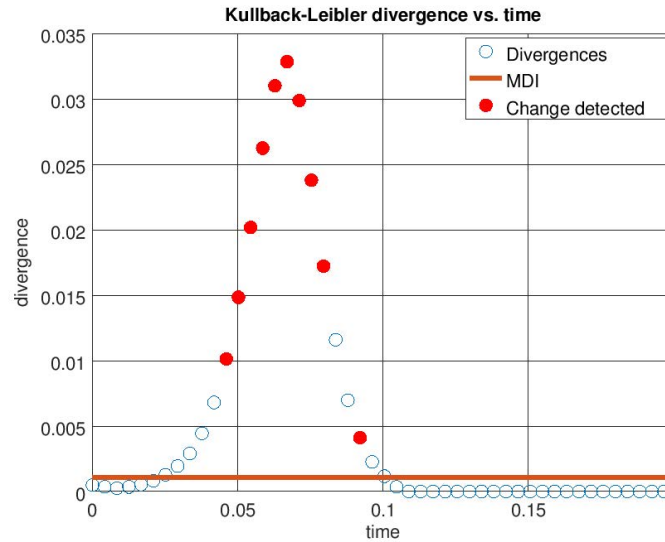


Fig. 2. Divergences versus time showing where the minimum discrimination information (MDI) threshold is and where statistically significant change was detected.

- [11] S. Kullback and M. A. Khairat, "A note on minimum information discrimination," *Annals of Mathematical Statistics*, Vol. 37, No. 1, 1966, pp. 279–280.
- [12] A. Wald, "Sequential Tests of Statistical Hypotheses," *Annals of Mathematical Statistics*, Vol. 16, No. 2, 1945, pp. 117–186.
- [13] J. Neyman and E. S. Pearson, "On the Problem of the Most Efficient Tests of Statistical Hypotheses," *Philosophical Transactions of the Royal Society A: Mathematical, Physical, and Engineering Sciences*, Vol. 231, 1933, pp. 289–337.
- [14] S. Tantaratana and H. Poor, "Asymptotic Efficiencies of Truncated Sequential Tests," *IEEE Transactions on Information Theory*, Vol. 28, No. 6, 1982, pp. 911–923.
- [15] S. D. Blostein and T. S. Huang, "Detecting Small, Moving Objects in Image Sequences Using Sequential Hypothesis Testing," *IEEE Transactions on Signal Processing*, Vol. 39, No. 7, 1991, pp. 1611–1629.