Standard Bayesian Approach to Quantized Measurements and Imprecise Likelihoods

Lawrence D. Stone
Metron Inc.
Reston, VA USA
Stone@metsci.com

***Abstract*— In this paper we show that the standard definition of likelihood function used in Bayesian inference simply and correctly handles imprecise likelihood functions and quantized measurements. Some recent papers have stated or implied that methods involving random sets, fuzzy membership functions, generalized likelihood functions, or Dempster-Shafer concepts are required to handle imprecise likelihood functions and quantized measurements. While it is true that one can use these methods, employing them adds unnecessary complication and possibly confusion to the solution of a simple problem. In the spirit of Occam's razor, we feel the simplest correct solution is the best.**

Keywords—Likelihood; Bayes; Random Sets; Fuzzy Logic; Dempster-Shafer; Imprecise; Quantized Measurements

I. Introduction

References [1] and [2] consider the problem of constructing a likelihood function for quantized measurements and propose that these types of measurements require a generalization of the standard notion of likelihood function that involves the use of random sets, concepts from fuzzy logic and Dempster-Shafer theory, as well as generalized or imprecise likelihood functions.

The purpose of this paper is to show that the standard concept of likelihood function as defined in [3] or [4] is sufficient to solve the general problem in an easy and straightforward manner. This is an important point because in the spirit of Occam's razor, the simplest correct solution to a problem is the best one. Simplicity allows readers to clearly understand the nature of the problem and its solution. It facilitates the use of the concept in applications and makes it easier to extend the concept to more challenging problems. It enables progress.

Of course, in performing research it is important to explore new and different methods and concepts. One of the virtues of the Dempster-Shafer (DS) model of uncertainty is that it expands the uncertainty calculus to situations in which it is not possible or reasonable to assign probabilities to the points of a finite state space S. Instead, one assigns probability (belief mass) to subsets. Assigning a belief mass to a subset means that the (probability) mass is allowed to "float" among the members of the subset. As an example, the notion of having no information about a parameter other than it is contained in S can be modeled by assigning (probability) mass 1 to the set S itself and 0 to all other subsets of S. A drawback to DS is that one must assign a (probability) belief mass to every subset of S. The size of the (probability) belief space is equal to $2^{|S|}$ where $|S|$ is the number of points in S. This becomes astronomically large as the number of points in S increases. For example, $2^{100} = 1.3 \times 10^{30}$. For even modest state spaces, the computational problems associated with DS become quite daunting. A special case of the DS model occurs when all the (probability) mass is assigned to singleton sets. This is called a Bayesian belief function [8] and is equivalent to the standard Bayesian probability function. When this special case holds, one is better off following the standard Bayesian calculus with its more reasonable computational load and larger set of tools (such as expectation) than are available in DS.

It is not a service to the research and applications communities to leave the impression that one of these more complex and sometimes difficult methods are required to solve a problem that can be handled by simpler, standard methods. What both communities desire is an understanding of when these alternate methods are required or provide an extra benefit.

This leads to the following open question. What situations involving quantized measurements or imprecise likelihood functions *require* the use of alternate, non-Bayesian models of uncertainty?

II. Bayesian Inference Formulation

Before beginning the discussion, we give the formulation of the basic Bayesian inference problem that is presented in [3] and is consistent with that in [4].

There is a unknown parameter Θ that we wish to estimate. There is a prior distribution p_0 on Θ such that

$$p_0(\theta) = \Pr\{\Theta = \theta\} \tag{1}$$

where Pr indicates either probability or probability density as appropriate. We obtain a measurements Z from a sensor. The measurement is viewed as a random variable whose distribution depends on θ. We define the likelihood function

$$l(z \mid \theta) = \Pr\{Z = z \mid \Theta = \theta\}. \quad (2)$$

If we receive a measurement $Z = z$, we compute the posterior distribution

$$p_1(\theta \mid Z = z) = \frac{l(z \mid \theta)\, p_0(\theta)}{\int l(z \mid \theta')\, p_0(\theta')\, d\theta'} \quad (3)$$

where integration is replaced by summation if the distribution on Θ is discrete.

III. Quantized Measurements

References [1] and [2] have illustrated the necessity of their approaches through the example of performing inference using quantized measurements. We use this same example to show that standard likelihood functions and the Bayesian inference process as given in (1) – (3) provide a straight-forward and correct way of incorporating quantized measurements into Bayesian inference. No generalization is required, and no extensions of the standard Bayesian probability concepts are needed.

We start with the digital voltmeter example given in [2]. Measurements are taken by a digital voltmeter that provides voltage readings to two decimal places. From the digital voltmeter measurement we wish to estimate the actual voltage Θ. Let p_0 be the prior on Θ. We consider three cases, measurements without noise, measurements with noise, and measurements where the quantization is unknown. We also consider a fourth case where the bins defining the quantization of the measurements are not intervals. This case is motivated by acoustic detections in deep ocean areas.

A. *Quantized Measurements without noise – known quantization*

If there is no noise added to the actual voltage, then any voltage in the interval (199.975, 1999.985] will produce a measurement of 199.98. The measurement space is a discrete set of points on the real line of the form $j \times 0.01$ where j is an integer such that $-\infty < j < \infty$. Any voltage in the set

$$S_j = (j \times 0.01 - 0.005, j \times 0.01 + 0.005] \quad (4)$$

will produce a measurement $Z = j \times 0.01$. From the definition of likelihood function in (2), we have

$$\begin{aligned} l(j \times 0.01 \mid \theta) &= \Pr\{Z = j \times 0.01 \mid \theta\} \\ &= \begin{cases} 1 \text{ if } \theta \in S_j \\ 0 \text{ otherwise.} \end{cases} \end{aligned} \quad (5)$$

For notational convenience, we shall use $Z = j$ for the measurement and $l(j \mid \theta)$ for the likelihood function in (5).

The posterior on the actual voltage Θ is computed by

$$\begin{aligned} p_1(\theta \mid Z = j) &= \frac{l(j \mid \theta)\, p_0(\theta)}{\int l(j \mid \theta')\, p_0(\theta')\, d\theta'} \\ &= \begin{cases} \dfrac{p_0(\theta)}{\int_{S_j} p_0(\theta')\, d\theta'} & \text{if } \theta \in S_j \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \quad (6)$$

B. *Quantized Measurements with noise – known quantization*

In this example we suppose the received voltage r at the digital voltmeter is the true voltage θ plus noise ε . Specifically, the true voltage is $\theta + \varepsilon$ where ε has the density function

$$f(y) = \Pr\{\varepsilon = y\} \text{ for } -\infty < \varepsilon < \infty .$$

The digital voltmeter produces measurements to two decimal places as above. In this case the likelihood function becomes

$$\begin{aligned} l(j \mid \theta) &= \Pr\{Z = j \mid \theta\} \\ &= \Pr\{j \times 0.01 - 0.005 < \theta + \varepsilon \le j \times 0.01 + 0.005\} \\ &= \Pr\{j \times 0.01 - \theta - 0.005 < \varepsilon \le j \times 0.01 - \theta + 0.005\} \\ &= \int_{j \times 0.01 - \theta - 0.005}^{j \times 0.01 - \theta + 0.005} f(y)\, dy. \end{aligned} \quad (7)$$

and the posterior distribution on Θ given the measurement $Z = j$ is

$$\begin{aligned} p_1(\theta \mid Z = j) &= \frac{p_0(\theta)\, l(j \mid \theta)}{\int p_0(\theta')\, l(j \mid \theta')\, d\theta'} \\ &= \frac{p_0(\theta) \int_{j \times 0.01 - \theta + 0.005}^{j \times 0.01 - \theta + 0.005} f(y)\, dy}{\int p_0(\theta') \left(\int_{j \times 0.01 - \theta' - 0.005}^{j \times 0.01 - \theta' + 0.005} f(y)\, dy \right) d\theta'} \end{aligned}$$

Example 1: As an example, let us consider the situation where

$$f(y) = \eta(y, 0, \sigma^2)$$

where $\eta(\cdot,0,\sigma^2)$ is the probability density function for the normal distribution with mean 0 and variance σ^2. The notation $\eta(\cdot,0,\sigma^2)$ is used to indicate the function of one variable obtained by fixing the values of the 2nd and 3rd variables at 0 and σ^2. We use a similar notation for a function of two variables when we wish to fix the value of one of the variables. Let

$$\Phi\left(z,\sigma^2\right)=\int_{-\infty}^{z}\eta\left(y,0,\sigma^2\right)dy \text{ for } -\infty<z<\infty. \quad (8)$$

Then the likelihood function $l(j\,|\,\theta)$ in (7) becomes

$$l\left(j\,|\,\theta\right)=\Phi\left(j\times 0.01-\theta+0.005,\sigma^2\right) -\Phi\left(j\times 0.01-\theta-0.005,\sigma^2\right). \quad (9)$$

Figure 1 shows plots of the likelihood function in (9) for $Z=10$ and $\sigma^2=0.0001, 0.0004, \text{and } 0.0016$.

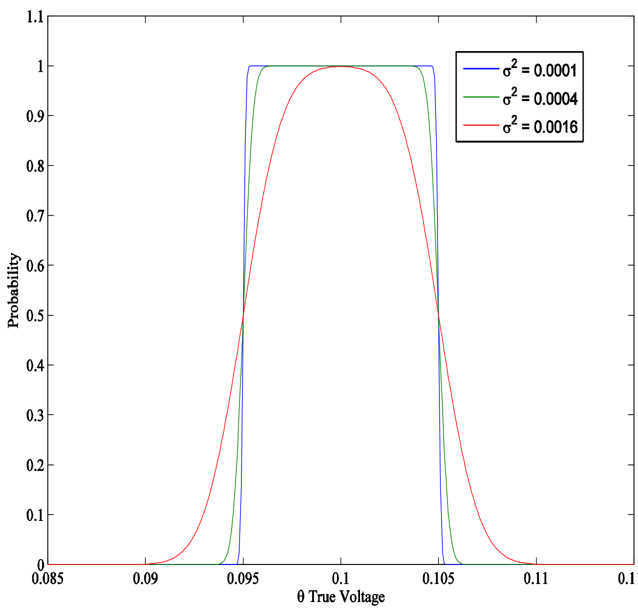

Fig. 1. Likelihood Functions for $j = 10$ (0.1 volt reading on voltmeter) when $\sigma^2=0.0001, 0.0004, \text{and } 0.0016$

The figure above looks very much like Figure 1 in [1]. In fact one can obtain that figure if he considers the case where the quantization has bin size 15 and the measurement $Z=j$ indicates the interval $(15j,15(j+1)]$. In this case the sets S_j in (4) become

$$S_j=(15j,15(j+1)],$$

and the likelihood function in (9) becomes

$$l\left(j\,|\,\theta\right)=\Phi\left(j\times(15+1)-\theta,\sigma^2\right)-\Phi\left(j\times 15-\theta,\sigma^2\right)$$

where θ plays the role of the variable z in [1].

C. Quantized Measurements when Quantization is Unknown

In the case where the quantization is unknown, we expand the state space on which we perform inference to simultaneously estimate the voltage and the quantization. To illustrate how this is done within the conventional Bayesian inference formalism, we consider the case where there is no noise added to the voltage and the quantized bins have a known and equal size. However, we do not know the anchor point for the bins.

Following Example 1, we take the bin size to be 0.01. However we don't know the anchor point of the bins. Specifically there is a unknown parameter Δ such that $-0.005<\Delta\le 0.005$ and

$$S_j\left(\Delta\right)=(j\times 0.01-0.005+\Delta, j\times 0.01+0.005+\Delta].$$

The inference problem is to estimate both θ and Δ.

In classic Bayesian fashion, we impose a prior distribution on $\left(\theta,\Delta\right)$ which represents our prior knowledge (or uncertainty) about $\left(\theta,\Delta\right)$. As an example we suppose that the priors on the two parameters are independent and the joint density on $\left(\theta,\Delta\right)$ is given by

$$g_0(\theta,\delta)=p_0(\theta)q_0(\delta) \quad \text{for } -\infty<\theta<\infty;\ -0.005<\delta\le 0.005.$$

The likelihood function for the observation $Z=j\times 0.01$ is

$$l\left(j\,|\left(\theta,\delta\right)\right)=\Pr\left\{Z=j\times 0.01\,|\left(\Theta,\Delta\right)=\left(\theta,\delta\right)\right\} =\begin{cases}1 \text{ if } \theta\in S_j\left(\delta\right)\\ 0 \text{ otherwise.}\end{cases}$$

Let g_1 be the posterior joint density on $\left(\theta,\Delta\right)$ given $Z=j$. Then

$$g_1\left((\theta,\delta)\,|\,Z=j\right) =\frac{l\left(j\,|\,(\theta,\delta)\right)p_0(\theta)q_0(\delta)}{\int_{-0.005}^{0.005}\int l\left(j\,|\,(\theta',\delta')\right)p_0(\theta')q_0(\delta')d\theta' d\delta'} =\begin{cases}\dfrac{p_0(\theta)q_0(\delta)}{\int_{-0.005}^{0.005}\int_{S_j(\delta')}p_0(\theta')q_0(\delta')d\theta' d\delta'} & \text{for } \theta\in S_j(\delta)\\ 0 \text{ otherwise.}\end{cases}$$

D. *Quantized Measurements with Noise – Bins that are not Intervals*

The bins, e.g., the sets S_j in (4), do not have to be intervals. They can be the unions of disjoint intervals or even more general sets. The approach given above will still work.

Example 2: Let us consider a case from underwater acoustics. When a passive acoustic sensor is located in a deep water area of the ocean, the sound propagation conditions often produce detection areas that are disjoint. For example, there may be good detection conditions from the sensor's location out to range 5 NM. This is typically called the direct path region. In addition there are often convergence zone regions at ranges of roughly 30 NM, 60 NM, and even farther out. A convergence zone is a region where the acoustic rays converge and produce low propagation loss and increased detection probability for the sensor. Suppose the convergence zones are 5 NM wide. It is often the case that the uncertainty the about source level (loudness of sound emitted) of a potential target, means that although one cannot calculate the detection probability as a function of range, one does know that if a target is detected it is one of these regions. In this case, a detection means that the target is in one of the above range intervals, i.e., its range is in the set

$$S = \{[0,5] \cup [27.5, 32.5] \cup [57.7, 62.5]\} \tag{10}$$

Generally one does not know the edges of the intervals in S exactly. Depending on the source level of the target and the level of ambient noise in the ocean, these areas can be a bit larger or smaller than the nominal numbers in (10). We will model this uncertainty with a likelihood function that is similar to the one given in Example 1 with the exception that there is only one bin $Z = 1$ corresponding to a detection. Specifically we let r denote the range of the target and ε be normally distributed with mean 0 and variance σ^2. Then the likelihood function l_d for a detection becomes

$$\begin{aligned} l_d(1 \mid r) &= \Pr\{Z = 1 \mid \text{ target at range } r\} \\ &= \begin{cases} \Pr\{r \le 5 + \varepsilon\} & \text{for } 0 \le r \le 20 \\ \Pr\{27.5 - \varepsilon \le r \le 32.5 + \varepsilon\} & \text{for } 20 < r \le 45 \\ \Pr\{57.5 - \varepsilon \le r \le 62.5 + \varepsilon\} & \text{for } 45 < r < \infty \end{cases} \end{aligned} \tag{11}$$

In terms of Φ defined in (8), the likelihood function in (11) becomes

$$l_d(1 \mid r) = \begin{cases} \min\{1 - \Phi(-5 - r, \sigma^2), 1 - \Phi(r - 5, \sigma^2)\} \\ \qquad \text{for } 0 \le r \le 20 \\ \min\{1 - \Phi(27.5 - r, \sigma^2), 1 - \Phi(r - 32.5, \sigma^2)\} \\ \qquad \text{for } 20 < r \le 45 \\ \min\{1 - \Phi(57.5 - r, \sigma^2), 1 - \Phi(r - 62.5, \sigma^2)\} \\ \qquad \text{for } 45 < r < \infty. \end{cases} \tag{12}$$

Figure 2 shows the likelihood function $l_d(1 \mid \cdot)$ when $\sigma^2 = 0.5$.

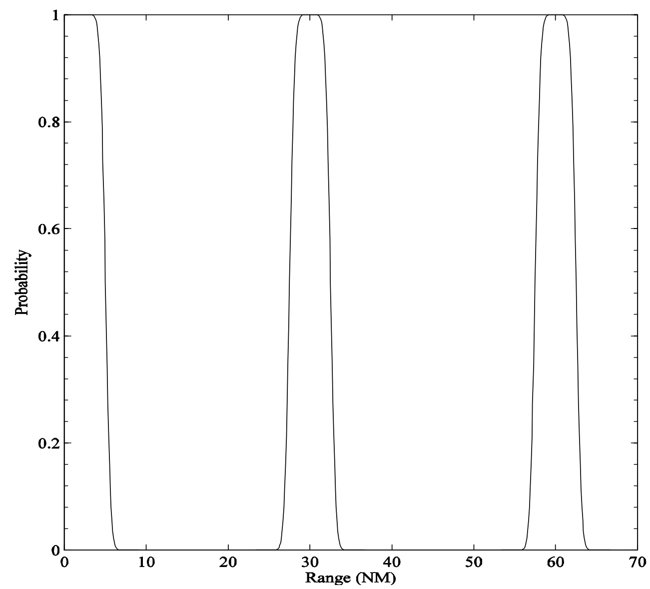

Fig. 2. Likelihood function $l_d(1 \mid \cdot)$ for accoustic detection in a convergence zone environment

IV. FUSION

In this section we show how one can fuse the information from quantized sensors.

Example 3. We consider two sensors of the type in Example 2. For simplicity we will consider a one dimensional state space. In most applications the position space would be two or three dimensional, but this one dimensional example will illustrate the methods involved. We take the target state space to be $X = [0 \text{ NM}, 70 \text{ NM}]$ and the prior to be uniform over X. Sensor 1 is located at $x = 0$ and sensor 2 is located at $x = 33$

The likelihood function $l_d^1(1 \mid \cdot)$ for a detection from sensor 1 is

$$l_d^1(1|x) = \begin{cases} \min\left\{1-\Phi(-5-x,\sigma^2), 1-\Phi(x-5,\sigma^2)\right\} \\ \qquad\qquad \text{for } 0 \le x \le 20 \\ \min\left\{1-\Phi(27.5-x,\sigma^2), 1-\Phi(x-32.5,\sigma^2)\right\} \\ \qquad\qquad \text{for } 20 < x \le 45 \\ \min\left\{1-\Phi(57.5-x,\sigma^2), 1-\Phi(x-62.5,\sigma^2)\right\} \\ \qquad\qquad \text{for } 45 < r < 70. \end{cases}$$

A plot of this likelihood function would look the same as Figure 2 with Range replaced by x on the horizontal axis.

The likelihood function for a detection from sensor 2 is

$$l_d^2(1|X) = \begin{cases} \min\left\{1-\Phi(28-r,\sigma^2), 1-\Phi(r-38,\sigma^2)\right\} \\ \qquad\qquad \text{for } 0 \le r \le 45 \\ \min\left\{1-\Phi(60.5-r,\sigma^2), 1-\Phi(r-65.5,\sigma^2)\right\} \\ \qquad\qquad \text{for } 45 < r \le 70. \end{cases}$$

Figure 3 shows the likelihood function $l_d^2(1|\cdot)$. Notice this likelihood is the same as the one in Figure 2 shifted to the right by 33 NM with the following differences. The second convergence zone is outside the state space $X = [0\text{ NM}, 70\text{ NM}]$ and the full direct path region is now in the state space and is centered around the sensor location at $x = 33$.

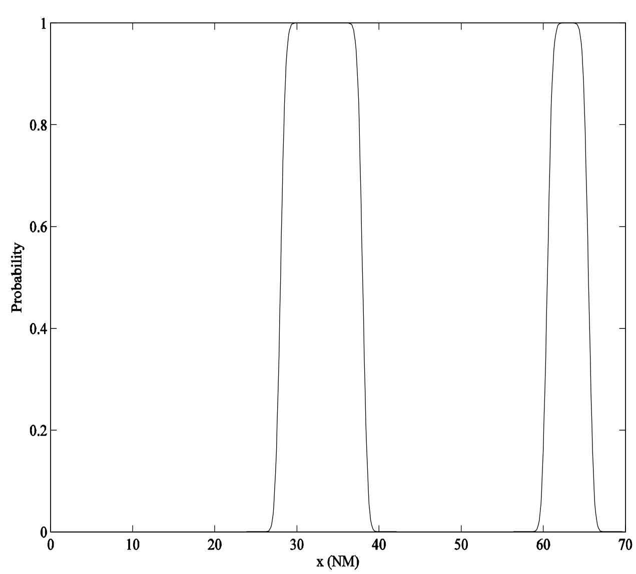

Fig. 3. Likelihood function $l_d^2(1|\cdot)$ for detection from sensor 2 located at $x = 33$

The joint likelihood function for detections from both sensor 1 and sensor 2 is the product of the two likelihood functions and is shown n Figure 4. Notice that direct path region near $x = 0$ has been zeroed out. The first convergence zone region centered a little to right of $x = 30$ is much narrower than the similar regions in Figures 2 and 3, and the second convergence zone region centered near $x = 62$ is narrower and has somewhat lower likelihood than in either Figure 2 or 3. The posterior on x is proportional to the likelihood function shown in Figure 4.

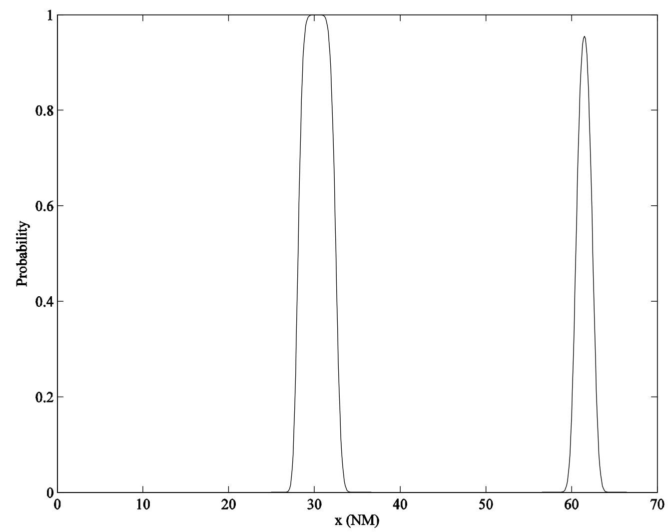

Fig. 4. Joint likelihood for a detection from sensors 1 and 2.

V. TRACKING MOVING TARGETS WITH QUANTIZED MEASUREMENTS

The target considered in section IV is stationary. One can also track moving targets using sensors with quantized measurements. Since the measurements do not satisfy the linear Gaussian assumptions required for a Kalman filter, it will usually be best to perform the tracking using a particle filter as described in [5] or chapter 3 of [6].

The general procedure is straight-forward. The particles are motion updated to the time of a measurement. The likelihood function for the quantized measurement is applied to the weight of each particle to produce the posterior distribution on target state. The particles are resampled as necessary and then motion updated to the time of the next measurement.

In particular, suppose that the target state distribution at time t is represented by the set of particles

$$\left\{(x_n(t), w_n(t)\right\} \text{ for } n = 1, \ldots, N$$

where x_n is the state of the nth particle and w_n is its probability. This set of particles represents a discrete probability approximation to the distribution on target state at the time t. Suppose we obtain a quantized measurement $Z = z$ such as the ones in Example 3. Let $l(z|x)$ be the likelihood function for this measurement. The posterior distribution on target state at time t is given by

$$\{(x_n(t), \tilde{w}_n(t)\} \text{ for } n = 1, \ldots, N \tag{13}$$

where

$$\tilde{w}_n(t) = \frac{l\big(z \mid x_n(t)\big) w_n(t)}{\sum_{n'=1}^{N} l\big(z \mid x_{n'}(t)\big) w_{n'}(t)}.$$

If the next measurement is received at time $t' > t$, the posterior particle filter representation in (13) can be motion-updated to the time t' to act as a proposal distribution for the incorporation of the measurement at time t'. Often the posterior in (13) is resampled before the motion update is performed. In addition other proposal distributions can used to improve the particle filter performance as discussed in [5].

One could employ a particle filter to track a submarine using the sensors described in Example 3. No extension beyond standard Bayesian likelihood functions and inference is required.

Reference [7] investigates the problem of tracking a target with quantized measurements. The authors assume a Gaussian motion model for the target and develop an approximate Minimum Mean Squared Error (MMSE) solution. They provide a numerical algorithm for obtaining this solution. This is very impressive work, and one must admire the authors for the cleverness of their solution. However, the solution is complex and does require many special assumptions. By contrast the particle filter approach is very general and simple. One is not constrained to Gaussian motion models or measurement errors. It is straight-forward to incorporate a wide variety of types of measurements. A possible drawback to using a particle filter is that it that it may be too computer intensive to be practical in some applications. However, the increasing capability of computers makes this less and less likely to be a problem.

VI. CONCLUSIONS

In the examples given above we have shown how to construct likelihood functions for quantized measurements using the standard Bayesian approach with standard likelihood functions. We have shown how to fuse information from one or multiple quantized measurements to compute a posterior distribution. We have also noted that quantized measurements can be applied to moving target problems using particle filters and likelihood functions for the quantized measurements in a straight-forward, Bayesian fashion.

The power of a likelihood function is that it converts measurements from (almost) any measurement space into a function on the target state space. This allows us to incorporate the information in these measurements into the posterior distribution on the target state space. The examples given above illustrate this process with quantized measurements, but the method is applicable to wide range of types of measurements and sensors. In particular, it is applicable to any measurement for which one can compute a likelihood function using the definition in (2). This is why likelihood functions are the common currency of information in Bayesian inference. The examples given above demonstrate this fact.

We have discussed above the virtues of using the simplest solution to a problem. In addition, there can be drawbacks to unnecessary complexity. For example, if we employ Dempster Shafer methods to handle quantized measurements, then we will be limited in applications to finite discrete state spaces since there has been no satisfactory extension of Dempster-Shafer theory to continuous state spaces. Even if the state space is finite, the computations involved with Dempster-Shafer methods grow exponentially with the size of the state space which limits its applicability to real problems.

REFERENCES

[1] B Ristic, "Bayesian estimation with imprecise likelihoods: Random set approach," IEEE Signal Processing Letters, vol. 18, pp. 394–398, July 2011.

[2] R Mahler, "General Bayes filtering of quantized measurements" pp. 346–352 in Proceedings of 14th International Conference on Information Fusion, Chicago, USA, July 5-8, 2011.

[3] L. Stone, A. Barlow, and T. Corwin, *Bayesian Multiple Target Tracking*, Boston, MA: Artech House, 1999

[4] J. Berger, Statistical Decsion Theory and Bayesian Analysis, 2nd ed, New York, NY: Springer-Verlag, 1985.

[5] B. Ristic, S. Arulampalam, and N Gordon, *Beyond the Kalman Filter,* Boston, MA. Artech House, 2004.

[6] L. Stone, et al, *Bayesian Multiple Target Tracking 2nd Edition,* Boston, MA. Artech House, to be published

[7] Z. Daun, V. Jilkov, and X, Lie, "State Estimation with Quantized Measurements: Approximate MMSE Approach" pp. 1067 – 1072 in Proceedings of 13th International Conference on Information Fusion, Edinburgh Scotland , July 26-29, 2010.

[8] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press 1976.