

Highest Density Gates for Target Tracking

F. Jay Breidt and Alicia L. Carriquiry ¹
Iowa State University

June 2, 1997

Abstract:

The problem of forming validation regions or gates for new sensor measurements obtained when tracking targets in clutter is considered. Target dynamics and measurement characteristics are modeled with possible non-Gaussianities or nonlinearities, so that some degree of approximation is usually required in the computation of the filtering densities for the target position and predictive densities for future measurements. Highest density gates are proposed as summaries of the predictive densities. These gates are constructed numerically, via simulation from the filtering density approximation. The algorithm results in gates that are “exact” (up to numerical accuracy) regardless of the approximation used for the filtering density. That is, given an approximation to the filtering density, the gating procedure accounts for all further nonlinearities and non-Gaussianities. Numerical examples show that when the predictive density is markedly non-Gaussian, highest density gates offer advantages over the more common rectangular and ellipsoidal gates.

Keywords: Clutter, nonlinear filtering, validation region.

¹Breidt and Carriquiry are Associate Professors, Department of Statistics, Iowa State University, Ames, IA 50011-1210. This research was partially funded by Research Grant #N00014-96-1-0279 from the Office of Naval Research, US Department of Defense. Carriquiry gratefully acknowledges the Institute of Statistics and Decision Sciences, Duke University, where she was a visitor in the spring of 1997.

Electronic mail: jbreidt@iastate.edu.

1 Introduction

Target tracking is an essential component of military surveillance as well as air traffic control systems. In a target tracking system, the observed data at scan n are sensor measurements reporting information from the coverage region of the sensors. The scans $n = 1, 2, \dots$ occur at discrete time instants. The observations at scan n , $\mathbf{z}_n \in \mathbb{R}^w$, are modeled as transformed, noisy versions of the true state of the target, $\mathbf{x}_n \in \mathbb{R}^v$, where typically $w < v$.

A fairly general formulation of the measurement characteristics and target dynamics is given by the generalized state space model

$$\begin{aligned} \mathbf{z}_n &= g_n(\mathbf{x}_n) + \mathbf{w}_n \\ \mathbf{x}_{n+1} &= h_n(\mathbf{x}_n, \mathbf{v}_n), \end{aligned} \tag{1}$$

where the additive noise $\{\mathbf{w}_n\}$ is an independent and identically distributed (iid) sequence of w -dimensional vectors with (possibly non-Gaussian) density f_w , $g_n : \mathbb{R}^v \rightarrow \mathbb{R}^w$ is a known function, $h_n : \mathbb{R}^v \times \mathbb{R}^v \rightarrow \mathbb{R}^v$ is a known function, and $\{\mathbf{v}_n\}$ is a sequence of v -dimensional iid random vectors with (possibly non-Gaussian) density f_v , independent of $\{\mathbf{w}_n\}$. We assume that the densities f_w and f_v are known. The assumption of additive measurement noise is not critical, but is notationally convenient in later sections. The discrete-time dynamics in the state equation are also not essential; often, it is more natural to model target dynamics via differential equations.

The functions $\{g_n\}$ and $\{h_n\}$ will in general be nonlinear. For example, when \mathbf{x}_n consists of variables describing the position and velocity of a target in Cartesian coordinates, a typical observation may consist of the range, bearing angle, and azimuth angle of the target.

Given model (1), the inferential problem is recursive state estimation; that is, the construction of a sequence of conditional probability density functions which describe the current knowledge about the state given the observations. Let $D_n = \{\mathbf{z}_1, \dots, \mathbf{z}_n\}$. Then, beginning with $f(\mathbf{x}_0 | D_0)$, we have the following well-known recursive relations:

- *State predictive density:*

$$f(\mathbf{x}_n | D_{n-1}) = \int f(\mathbf{x}_n | \mathbf{x}_{n-1}) f(\mathbf{x}_{n-1} | D_{n-1}) d\mathbf{x}_{n-1}. \tag{2}$$

- *Filtering density:*

$$f(\mathbf{x}_n | D_n) = \frac{f(\mathbf{z}_n | \mathbf{x}_n) f(\mathbf{x}_n | D_{n-1})}{f(\mathbf{z}_n | D_{n-1})}, \quad (3)$$

where the normalizing constant in the denominator is the *measurement predictive density*

$$f(\mathbf{z}_n | D_{n-1}) = \int f(\mathbf{z}_n | \mathbf{x}_n) f(\mathbf{x}_n | D_{n-1}) d\mathbf{x}_n. \quad (4)$$

If the functions g_n and h_n are linear and the densities f_v and f_w are Gaussian in model (1), then all of the predictive and filtering densities are Gaussian, and the above equations reduce to the Kalman filter, which recursively updates the required mean vectors and covariance matrices. In the nonlinear or non-Gaussian case, there is no general closed-form solution for the required densities, and some degree of approximation is usually required. For example, the extended Kalman filter handles nonlinearity through Taylor series linearization, and non-Gaussianity through matching of first and second-order moments, so that the required densities are approximated by Gaussian densities (e.g., Anderson and Moore, 1979). The literature on linear and nonlinear filtering methods is vast. For example, Alspach and Sorenson (1972), Kallianpur (1980), Di Masi and Runggaldier (1981), Kitagawa (1987), Meinhold and Singpurwalla (1989), Gordon, Salmond and Smith (1993), and Budhiraja and Mirković (1997) describe various filtering algorithms.

In practice, the approximate densities from a filtering algorithm are usually summarized with a point predictor, such as the mean, median or mode of the density. An additional useful summary is a $(1 - \alpha)100\%$ *prediction region*: a subset of the sample space with the property that

$$\Pr[\text{predictand} \in \text{region} | D_n] \geq 1 - \alpha.$$

Prediction regions are not unique; several different constructions will be described below.

A prediction region for a future measurement \mathbf{z}_{n+1} is sometimes called a *validation region* or *gate*. Gates are useful in practice for clutter reduction, because they divide the measurement space \mathbb{R}^w into those points likely to be associated with the target under track and those unlikely to be associated with the target. Unlikely measurement-to-track pairs are discarded in a procedure known as *gating* (e.g., Blackman, 1986).

In this paper we consider the construction of highest density prediction regions. For concreteness, we focus on the construction of highest density gates for a new measurement \mathbf{z}_{n+1} conditional on the available data D_n , though prediction regions for other predictands could be constructed in a similar fashion.

The method we propose requires that the measurement predictive density, state predictive density, or filtering density be available at each scan, even if in unnormalized form. Further, we assume that simulation from the available density is feasible. The method is general in that it can be applied to the nonlinear, non-Gaussian tracking problem, regardless of the approximation used for the filtering density. Increasing the dimension of the measurement vector \mathbf{z}_{n+1} (for example, when tracking a formation of targets) causes no difficulty in principle, since construction of the highest density gate is resolved by calculation of scalar-valued variables.

The paper is organized as follows. In Section 2 we review some standard methods of gate construction. The gating method we propose is described in Section 3. Some illustrative examples are presented in Section 4 and discussion is in Section 5.

2 Background

A gate is a partition of \mathbb{R}^w such that a measurement within the gate is associated to the track being followed with a certain probability, called the coverage of the gate. Two common gates are *rectangular* and *ellipsoidal* gates, which we briefly review.

For a new measurement \mathbf{z}_{n+1} we want $\Pr\{\mathbf{z}_{n+1} \in \text{gate} \mid D_n\} \geq 1 - \alpha$. Let $\boldsymbol{\epsilon}_{n+1} = \mathbf{z}_{n+1} - g_{n+1}(\widehat{\mathbf{x}}_{n+1}) = (\epsilon_{1,n+1}, \dots, \epsilon_{w,n+1})' \in \mathbb{R}^w$ be the innovation vector at the $(n+1)$ st scan, where $g_{n+1}(\widehat{\mathbf{x}}_{n+1})$ is the predicted value of the transformed state resulting from the filtering algorithm. Let $\boldsymbol{\Psi}_{n+1} = [\psi_{jk,n+1}]$ denote the covariance matrix of $\boldsymbol{\epsilon}_{n+1}$.

Let $\varphi(\mathbf{z}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denote the w -dimensional Gaussian density with $w \times 1$ mean vector $\boldsymbol{\mu}$ and $w \times w$ positive definite covariance matrix $\boldsymbol{\Sigma}$; that is,

$$\varphi(\mathbf{z}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-w/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{z} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{z} - \boldsymbol{\mu})\right\}.$$

Rectangular and ellipsoidal gates assume that the distribution of the innovation vector is Gaussian, with density $\varphi(\boldsymbol{\epsilon}; \mathbf{0}, \boldsymbol{\Psi}_{n+1})$.

2.1 Rectangular gates

A rectangular gate (RG) for a w -dimensional measurement vector is the hyper-rectangle given by

$$\text{RG}_{n+1} = \left\{ \mathbf{z} \in \mathbb{R}^w : |\epsilon_j| \leq \xi_{\alpha/w} \psi_{jj,n+1}^{1/2}, (j = 1, \dots, w) \right\}, \quad (5)$$

where $\boldsymbol{\epsilon} = \mathbf{z} - g_{n+1}(\widehat{\mathbf{x}}_{n+1})$, $\Phi(\xi_{\alpha/w}) = 1 - \alpha/w$, and $\Phi(\cdot)$ is the cumulative distribution function of a standard Gaussian. Assuming that the innovations are Gaussian, the overall coverage probability is

$$\begin{aligned} \Pr(\mathbf{z}_{n+1} \in \text{RG}_{n+1} | D_n) &= \Pr\left(\bigcap_{j=1}^w \{|\epsilon_{j,n+1}| \leq \xi_{\alpha/w} \psi_{jj,n+1}^{1/2}\} | D_n\right) \\ &\geq 1 - \sum_{j=1}^w \Pr(|\epsilon_{j,n+1}| > \xi_{\alpha/w} \psi_{jj,n+1}^{1/2} | D_n) \\ &= 1 - w\alpha/w = 1 - \alpha, \end{aligned} \quad (6)$$

by Bonferroni's inequality. For example, for $\xi_{\alpha/w} = 2.24$ and $w = 2$, the rectangular gate has coverage probability greater than or equal to 0.95.

2.2 Ellipsoidal gates

An ellipsoidal gate (EG) is given by

$$\text{EG}_{n+1} = \left\{ \mathbf{z} \in \mathbb{R}^w : \boldsymbol{\epsilon}' \boldsymbol{\Psi}_{n+1}^{-1} \boldsymbol{\epsilon} \leq \kappa_\alpha \right\}, \quad (7)$$

where κ_α is the upper tail α -quantile of a χ^2 distribution with w degrees of freedom. Under the assumption of Gaussian innovations, EG_{n+1} has coverage probability equal to $1 - \alpha$, as we now show.

Let $\boldsymbol{\Lambda}_{n+1}$ denote the $w \times w$ diagonal matrix whose elements are $\lambda_{j,n+1}$, the eigenvalues of the positive definite matrix $\boldsymbol{\Psi}_{n+1}^{-1}$, and let \mathbf{E}_{n+1} be the corresponding matrix of eigenvectors such that $\mathbf{E}'_{n+1} \mathbf{E}_{n+1} = \mathbf{E}_{n+1} \mathbf{E}'_{n+1} = \mathbf{I}$. Then

$$d_{n+1}^2 = \boldsymbol{\epsilon}'_{n+1} \boldsymbol{\Psi}_{n+1}^{-1} \boldsymbol{\epsilon}_{n+1} = \boldsymbol{\epsilon}'_{n+1} \mathbf{E}'_{n+1} \boldsymbol{\Lambda}_{n+1}^{-1} \mathbf{E}_{n+1} \boldsymbol{\epsilon}_{n+1} = \boldsymbol{\epsilon}_{n+1}^{*'} \boldsymbol{\Lambda}_{n+1}^{-1} \boldsymbol{\epsilon}_{n+1}^*,$$

where $\boldsymbol{\epsilon}_{n+1}^* = \mathbf{E}_{n+1} \boldsymbol{\epsilon}_{n+1}$. It follows for Gaussian innovations that

$$d_{n+1}^2 = \sum_{j=1}^w \frac{\epsilon_{j,n+1}^{*2}}{\lambda_{j,n+1}} \sim \chi_w^2,$$

so that the values for κ_α such that $\Pr \{d_{n+1}^2 \leq \kappa_\alpha\} = 1 - \alpha$ can be obtained from the χ_w^2 tail probabilities. For example, for $\alpha = 0.05$ and $w = 2$, $\kappa_\alpha = 5.99$.

3 Highest Density Gates

It is clear that a rectangular gate is a fairly crude approximation, since it uses a Gaussian assumption and ignores any dependence structure in the innovations vector. Ellipsoidal gates take into account the dependence structure, but may be poor if the Gaussian approximation is poor. Heavy tails, skewness, or multimodalities would all lead to poor approximations, and each of these features may be present even in very simple nonlinear or non-Gaussian systems, as in the following example due to Kitagawa (1987). The state dynamics consist of a random walk with a heavy-tailed innovation,

$$x_{n+1} = x_n + v_n,$$

and the measurements consist of the random walk plus Gaussian noise,

$$z_n = x_n + w_n.$$

From Figure 4 of Kitagawa (1987), it is clear that skewness and bimodality may occur even in this simple model, as probability density is shifted from one likely state location to another over time. In the more complicated context of target tracking, clutter, maneuvers, and multiple targets may all lead to distinctly non-Gaussian features.

These difficulties with standard gating procedures suggest that other approaches should be considered. An ideal gate would have minimal volume in the measurement space for a given coverage probability, $1 - \alpha$. In fact, a prediction region has minimal volume for a given coverage if and only if, for $\mathbf{z} \in$ region and $\mathbf{z}^* \notin$ region,

$$f(\mathbf{z} | D_n) \geq f(\mathbf{z}^* | D_n)$$

(Box and Tiao, 1973); that is, every sensor observation inside the region has probability density at least as large as every sensor observation outside the region, which is one way to quantify the “likelihood” of association with the target.

The *highest density gate* (HDG) then consists of all $\mathbf{z} \in \mathbb{R}^w$ with high density:

$$\text{HDG}_{n+1}(q_\alpha) = \{\mathbf{z} \in \mathbb{R}^w : f(\mathbf{z} | D_n) \geq q_\alpha\}, \quad (8)$$

where q_α is the largest constant such that $\Pr\{f(\mathbf{z} | D_n) \geq q_\alpha | D_n\} = 1 - \alpha$. Construction of the HDG in any dimension thus depends only on the value q_α , which is the α th quantile of the *scalar* random variable $f(\mathbf{z}_{n+1} | D_n)$.

Hyndman (1996) suggests using an empirical quantile from a large number N of simulated observations as an estimator of q_α . In our context, if it is feasible to simulate from the measurement predictive density, we have the following algorithm:

1. Draw $\mathbf{z}^{(1)}, \dots, \mathbf{z}^{(N)}$ iid from $f(\mathbf{z}_{n+1} | D_n)$.

2. Form the scalars

$$Q_{n+1}(\mathbf{z}^{(i)}) = f(\mathbf{z}_{n+1} | D_n) |_{\mathbf{z}_{n+1}=\mathbf{z}^{(i)}}$$

for $i = 1, \dots, N$.

3. Sort the $Q_{n+1}(\mathbf{z}^{(i)})$'s to obtain the order statistics

$$Q_{1:N} \leq \dots \leq Q_{N:N}.$$

4. Estimate q_α by $\hat{q}_\alpha = Q_{[N\alpha]:N}$ (where $[x]$ is the largest integer not exceeding x), so that

$$\text{HDG}_{n+1}(\hat{q}_\alpha) = \{\mathbf{z} \in \mathbb{R}^w : f(\mathbf{z} | D_n) \geq \hat{q}_\alpha\}$$

is an approximate $(1 - \alpha)100\%$ highest density gate for \mathbf{z}_{n+1} .

This algorithm relies on simulation to avoid the difficulties of numerical integration in \mathbb{R}^w (Hyndman, 1996). All statements of independence are made conditionally on D_n . Note that the measurement predictive density need not be normalized in the above algorithm.

Let $F_{n+1}(q) = \Pr\{Q_{n+1}(\mathbf{z}^{(i)}) \leq q | D_n\}$ and let $f_{n+1}(q) = dF_{n+1}(q)/dq$. Then, using well-known asymptotic theory (e.g., Mood, Graybill and Boes, 1974), \hat{q}_α is approximately Gaussian with mean q_α and variance

$$\frac{\alpha(1 - \alpha)}{N f_{n+1}^2(q_\alpha)}.$$

This result may be useful for choosing a suitably large value of N .

If $f(\mathbf{z}_{n+1} | D_n)$ is not readily available for simulation, but the state predictive density $f(\mathbf{x}_{n+1} | D_n)$ is available, then replace steps 1 and 2 of the above algorithm by the following:

1. (a) Draw $\mathbf{x}^{(i0)}$ iid from $f(\mathbf{x}_{n+1} | D_n)$ for $i = 1, \dots, N$.
 (b) Draw $\mathbf{w}^{(i)}$ iid from f_w for $i = 1, \dots, N$.
 (c) Form $\mathbf{z}^{(i)} = g_{n+1}(\mathbf{x}^{(i0)}) + \mathbf{w}^{(i)}$ iid from $f(\mathbf{z}_{n+1} | D_n)$ for $i = 1, \dots, N$.
2. (a) Draw $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(M)}$ iid from $f(\mathbf{x}_{n+1} | D_n)$.
 (b) For $i = 1, \dots, N$, form the scalars

$$\begin{aligned}
 Q_{n+1}(\mathbf{z}^{(i)}) &= M^{-1} \sum_{j=1}^M f_w \left\{ \mathbf{z}^{(i)} - g_{n+1}(\mathbf{x}^{(j)}) \right\} \\
 &\xrightarrow{\text{a.s.}} \int f_w \left\{ \mathbf{z}^{(i)} - g_{n+1}(\mathbf{x}_{n+1}) \right\} f(\mathbf{x}_{n+1} | D_n) d\mathbf{x}_{n+1} \\
 &= f(\mathbf{z}_{n+1} | D_n) \Big|_{\mathbf{z}_{n+1}=\mathbf{z}^{(i)}},
 \end{aligned}$$

where the almost sure (a.s.) convergence is obtained as $M \rightarrow \infty$, and the last equality follows from equation (4).

Finally, if only the filtering density $f(\mathbf{x}_n | D_n)$ is available for simulation, replace steps 1 and 2 by the following:

1. (a) Draw $\mathbf{u}^{(i0)}$ iid from $f(\mathbf{x}_n | D_n)$ for $i = 1, \dots, N$.
 (b) Draw $\mathbf{v}^{(i0)}$ iid from f_v for $i = 1, \dots, N$.
 (c) Form $\mathbf{x}^{(i0)} = h_n(\mathbf{u}^{(i0)}, \mathbf{v}^{(i0)})$ iid from $f(\mathbf{x}_{n+1} | D_n)$ for $i = 1, \dots, N$.
 (d) Draw $\mathbf{w}^{(i)}$ iid from f_w for $i = 1, \dots, N$.
 (e) Form $\mathbf{z}^{(i)} = g_{n+1}(\mathbf{x}^{(i0)}) + \mathbf{w}^{(i)}$ iid from $f(\mathbf{z}_{n+1} | D_n)$ for $i = 1, \dots, N$.
2. (a) Draw $\mathbf{u}^{(1)}, \dots, \mathbf{u}^{(M)}$ iid from $f(\mathbf{x}_n | D_n)$.
 (b) Draw $\mathbf{v}^{(1)}, \dots, \mathbf{v}^{(M)}$ iid from f_v .
 (c) Form $\mathbf{x}^{(j)} = h_n(\mathbf{u}^{(j)}, \mathbf{v}^{(j)})$ iid from $f(\mathbf{x}_{n+1} | D_n)$ for $j = 1, \dots, M$.
 (d) For $i = 1, \dots, N$, form the scalars

$$\begin{aligned}
 Q_{n+1}(\mathbf{z}^{(i)}) &= M^{-1} \sum_{j=1}^M f_w \left\{ \mathbf{z}^{(i)} - g_{n+1}(\mathbf{x}^{(j)}) \right\} \\
 &\xrightarrow{\text{a.s.}} \int f_w \left\{ \mathbf{z}^{(i)} - g_{n+1}(\mathbf{x}_{n+1}) \right\} f(\mathbf{x}_{n+1} | D_n) d\mathbf{x}_{n+1} \\
 &= f(\mathbf{z}_{n+1} | D_n) \Big|_{\mathbf{z}_{n+1}=\mathbf{z}^{(i)}},
 \end{aligned}$$

where the almost sure (a.s.) convergence is obtained as $M \rightarrow \infty$, and the last equality follows from equation (4).

Of course, the numerical details of these algorithms must be worked out on a case-by-case basis. For example, the last two algorithms rely on Monte Carlo integration, but in some cases the measurement predictive density will be available analytically and in other cases the dimension of the integrals may be reduced by analytic computations.

To determine if a new sensor observation \mathbf{z}_{n+1}^* is in $\text{HDG}_{n+1}(\hat{q}_\alpha)$, simply evaluate $Q^* = f(\mathbf{z}_{n+1} | D_n) |_{\mathbf{z}_{n+1}=\mathbf{z}_{n+1}^*}$ and compare to \hat{q}_α ; if $Q^* \geq \hat{q}_\alpha$, then \mathbf{z}_{n+1}^* is in the gate. If $f(\mathbf{z}_{n+1} | D_n)$ is not available analytically, evaluate it via Monte Carlo integration as in step 2 of each of the last two algorithms; i.e., if

$$M^{-1} \sum_{j=1}^M f_w \left\{ \mathbf{z}_{n+1}^* - g_{n+1}(\mathbf{x}^{(j)}) \right\} \geq \hat{q}_\alpha,$$

then \mathbf{z}_{n+1}^* is in the gate.

4 Examples

We illustrate the performance of the HDG algorithm with some simulated numerical examples. We expect the highest density gates to outperform both the rectangular and ellipsoidal gates whenever the approximations to the filtering and predictive densities are highly non-Gaussian. The first two examples illustrate the differences between the standard gates and the highest density gates for some hypothetical predictive distributions: a skewed bivariate distribution and a bimodal bivariate distribution. The third and fourth examples involve tracking models: a classic tracking model with linear, Gaussian state dynamics and nonlinear measurements (range and bearing angle), and a multi-model with target dynamics switching between a maneuvering state and a non-maneuvering state according to the transitions of an unobserved Markov chain. This last example also includes clutter. The extended Kalman filter is used to approximate the sequence of filtering densities in the classic tracking model, and the bootstrap filter is used to approximate the sequence of filtering densities in the multi-model.

Example 1: Skewed predictive density

In this example, we use a correlated bivariate log-Gaussian as a hypothetical predictive density; specifically,

$$f(\mathbf{z}_{n+1} | D_n) = \varphi \left(\begin{bmatrix} \log z_{n+1,1} \\ \log z_{n+1,2} \end{bmatrix}; \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0.25 & 0.25 \\ 0.25 & 1.00 \end{bmatrix} \right) \frac{I_{\{z_{n+1,1} > 0\}} I_{\{z_{n+1,2} > 0\}}}{z_{n+1,1} z_{n+1,2}}.$$

This density has all of its mass on the first quadrant, and could be thought of as an appropriate predictive distribution for a target maneuvering near a physical boundary, like a submarine along a coastline. The density is highly skewed.

We generated $N = 10,000$ draws \mathbf{z}_{n+1} from this predictive distribution and computed the value of \hat{q}_α as described in Section 3, with $\alpha = 0.05$.

Both the rectangular and ellipsoidal gates shown in Figure 1 fail to recognize the skewness of the distribution and consequently cover large parts of the measurement space in which there is no probability density. By contrast, the highest density gate in Figure 1 yields a sensible prediction region with the proper support.

Example 2: Bimodal predictive density

In this example, we use a bimodal mixture of two bivariate Gaussian distributions as a hypothetical predictive distribution; specifically,

$$f(\mathbf{z}_{n+1} | D_n) = p\varphi(\mathbf{z}_{n+1}; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) + (1 - p)\varphi(\mathbf{z}_{n+1}; \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2),$$

where $p = 0.3$, $\boldsymbol{\mu}_1 = (5.0, 10.0)'$, $\boldsymbol{\mu}_2 = (14.0, 20.0)'$,

$$\boldsymbol{\Sigma}_1 = \begin{bmatrix} 1.0 & 1.0 \\ 1.0 & 4.0 \end{bmatrix}, \text{ and } \boldsymbol{\Sigma}_2 = \begin{bmatrix} 4.0 & 4.2 \\ 4.2 & 9.0 \end{bmatrix}.$$

We generated $N = 10,000$ draws \mathbf{z}_{n+1} from this mixture predictive distribution, and computed the value of \hat{q}_α as described in Section 3, with $\alpha = 0.05$. In Figure 2 we show the 95% rectangular, ellipsoidal, and highest density gates, together with the contour plots of the mixture predictive density $f(\mathbf{z}_{n+1} | D_n)$. The rectangular gate is very large, and includes portions of \mathbb{R}^2 with very low likelihood under the predictive distribution. As expected, the rectangular gate does not recognize the bimodality of the predictive density. The ellipsoidal

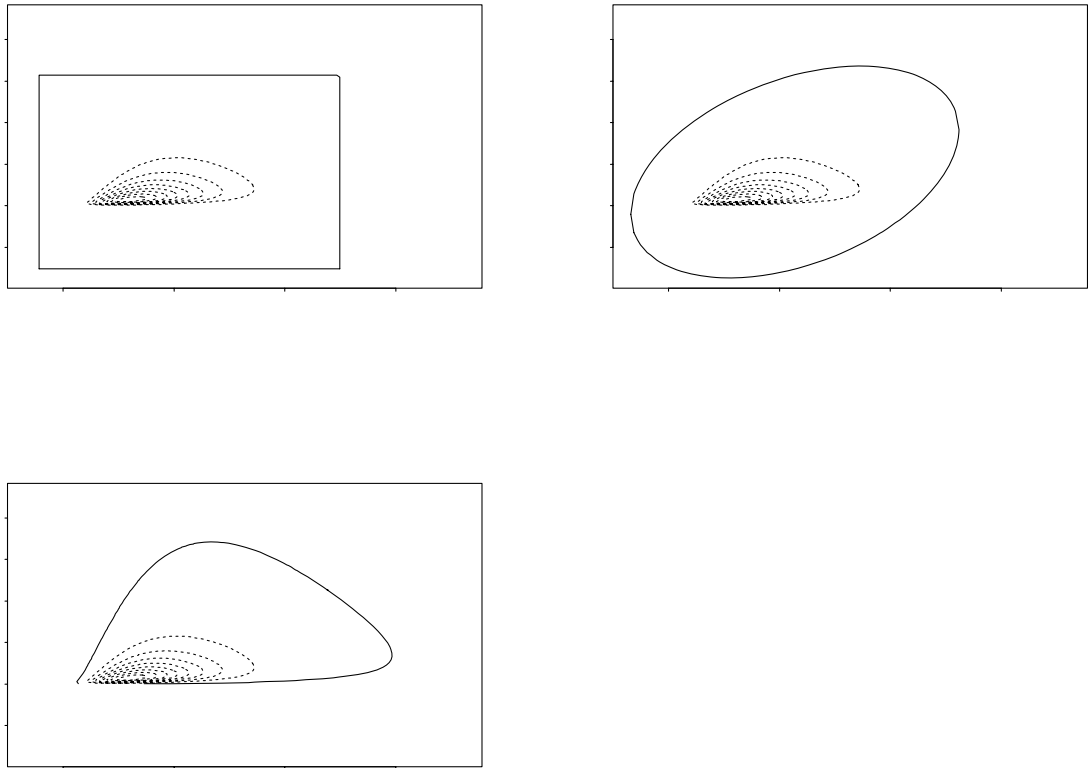


Figure 1: *Contours of a skewed predictive density (dotted lines) and 95% rectangular, ellipsoidal, and highest density gates (solid lines).*

gate recognizes the correlation structure in the innovations vector and so has smaller volume than the rectangular gate, but still does not reflect the fact that the predictive distribution is bimodal. Figure 2 also shows the highest density gate. Under the model, the HDG has coverage probability of 95% (up to numerical accuracy) and minimal volume. Indeed, the HDG is composed of two separate regions, corresponding to the two areas of highest probability given by the two components of the mixture predictive density.

Example 3: Extended Kalman filter for simple target tracking

Consider model (1) for the case in which the state vector $\mathbf{x}_n = (x_n, \dot{x}_n, y_n, \dot{y}_n)'$ has four components: position (x_n, y_n) and velocities (\dot{x}_n, \dot{y}_n) of a target at time n . The target dynamics evolve linearly from scan to scan, where scans are separated by a small time increment, Δ . The measurement vector \mathbf{z}_n has components of measured range and measured bearing angle, so that model (1) becomes

$$\begin{aligned} \mathbf{z}_n &= g(\mathbf{x}_n) + \mathbf{w}_n = \begin{bmatrix} (x_n^2 + y_n^2)^{1/2} \\ \tan^{-1}(y_n/x_n) \end{bmatrix} + \mathbf{w}_n \\ \mathbf{x}_{n+1} &= \mathbf{H}\mathbf{x}_n + \mathbf{v}_n = \begin{bmatrix} 1 & \Delta & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \Delta \\ 0 & 0 & 0 & 1 \end{bmatrix} \mathbf{x}_n + \mathbf{v}_n, \end{aligned}$$

where $\{\mathbf{w}_n\}$ is an iid sequence of Gaussian random vectors with density $f_w(\mathbf{w}) = \varphi(\mathbf{w}; \mathbf{0}, \mathbf{\Sigma}_w)$ and $\{\mathbf{v}_n\}$ is an iid sequence of Gaussian random vectors with density $f_v(\mathbf{v}) = \varphi(\mathbf{v}; \mathbf{0}, \mathbf{\Sigma}_v)$, independent of $\{\mathbf{w}_n\}$. Both the measurement error covariance matrix $\mathbf{\Sigma}_w$ and the state covariance matrix $\mathbf{\Sigma}_v$ are diagonal, with elements $(0.0001, 0.0001)$, and $(0.01, 0.001, 0.01, 0.001)$, respectively.

For this nonlinear model, the approximate filtering density that results from application of the EKF recursions is $f(\mathbf{x}_n | D_n) = \varphi(\mathbf{x}_n; \mathbf{x}_{n|n}, \mathbf{\Omega}_{n|n})$, where $\mathbf{x}_{n|n}$ is the filtered state vector at scan n and $\mathbf{\Omega}_{n|n}$ is its covariance matrix, obtained by cycling over the following steps:

1. Compute the innovations vector $\boldsymbol{\epsilon}_n = \mathbf{z}_n - g(\hat{\mathbf{x}}_n)$ and its approximate covariance matrix $\mathbf{\Psi}_n = \hat{\mathbf{G}}_n \mathbf{\Omega}_n \hat{\mathbf{G}}_n' + \mathbf{\Sigma}_w$, where

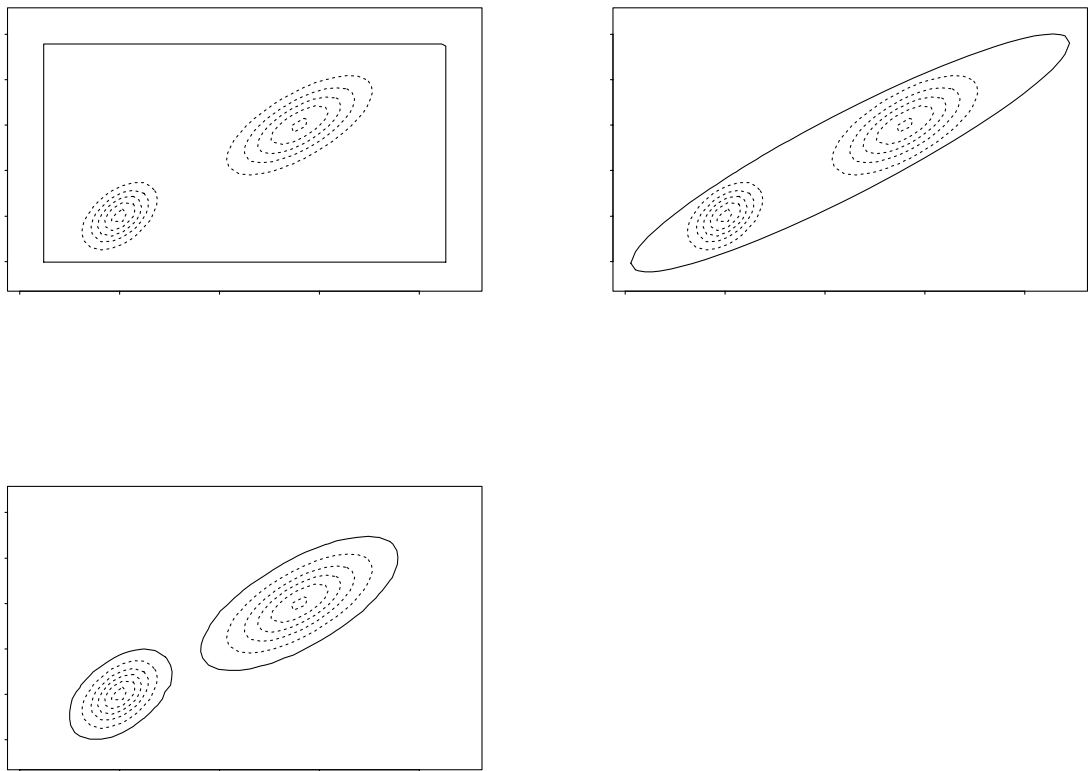


Figure 2: *Contours of a bimodal predictive distribution (dotted lines) and 95% rectangular, ellipsoidal, and highest density gates (solid lines).*

$$\hat{\mathbf{G}}_n = \frac{\partial g(\mathbf{x}_n)}{\partial \mathbf{x}'_n} \Big|_{\mathbf{x}_n = \hat{\mathbf{x}}_n}.$$

2. Obtain the filtered state vector and its covariance matrix as

$$\begin{aligned} \mathbf{x}_{n|n} &= \hat{\mathbf{x}}_n + \mathbf{\Omega}_n \hat{\mathbf{G}}'_n \mathbf{\Psi}_n^{-1} \boldsymbol{\epsilon}_n \\ \mathbf{\Omega}_{n|n} &= \mathbf{\Omega}_n - \mathbf{\Omega}_n \hat{\mathbf{G}}'_n \mathbf{\Psi}_n^{-1} \hat{\mathbf{G}}_n \mathbf{\Omega}_n. \end{aligned}$$

3. One step ahead prediction and the prediction covariance matrix are computed from the filtered values as

$$\begin{aligned} \hat{\mathbf{x}}_{n+1} &= \mathbf{H} \mathbf{x}_{n|n} \\ \mathbf{\Omega}_{n+1} &= \mathbf{H} \mathbf{\Omega}_{n|n} \mathbf{H}' + \mathbf{\Sigma}_v. \end{aligned}$$

For $\Delta = 0.01$ time units, and $\mathbf{x}_0 = \mathbf{0}$, we simulated a target trajectory over 50 time periods and applied the EKF to the simulated data to obtain filtered and predicted state values and an approximation to the filtering density at each time period. The goal was to compute a 95% gate for the 51st measurement using the available data $D_{50} = \{\mathbf{z}_1, \dots, \mathbf{z}_{50}\}$. The rectangular, ellipsoidal, and highest density gates are shown in Figure 3, together with the simulated, measured, and filtered trajectory of the target. Note that the rectangular gate constructed in measurement space is a polar rectangle in state space. It is somewhat larger than the ellipsoidal and highest density gates, which are quite similar, as would be expected when the noise is Gaussian, given that the nonlinearity in the model is moderate. Indeed, for a linear, Gaussian system, the ellipsoidal gate is the highest density gate.

Example 4: Bootstrap filter for multi-model with clutter

In this simplified example, the measurement space and state space are both $(0, 1)$. As shown by the dotted line in Figure 4, the target remains in one spot for five scans then, over the course of ten scans, maneuvers to a new spot where it remains for five scans. State dynamics are modeled as

$$x_{n+1} = x_n + v_n \sigma_v^{(s_n)}, \quad \{v_n\} \text{ iid } \varphi(v; 0, 1),$$

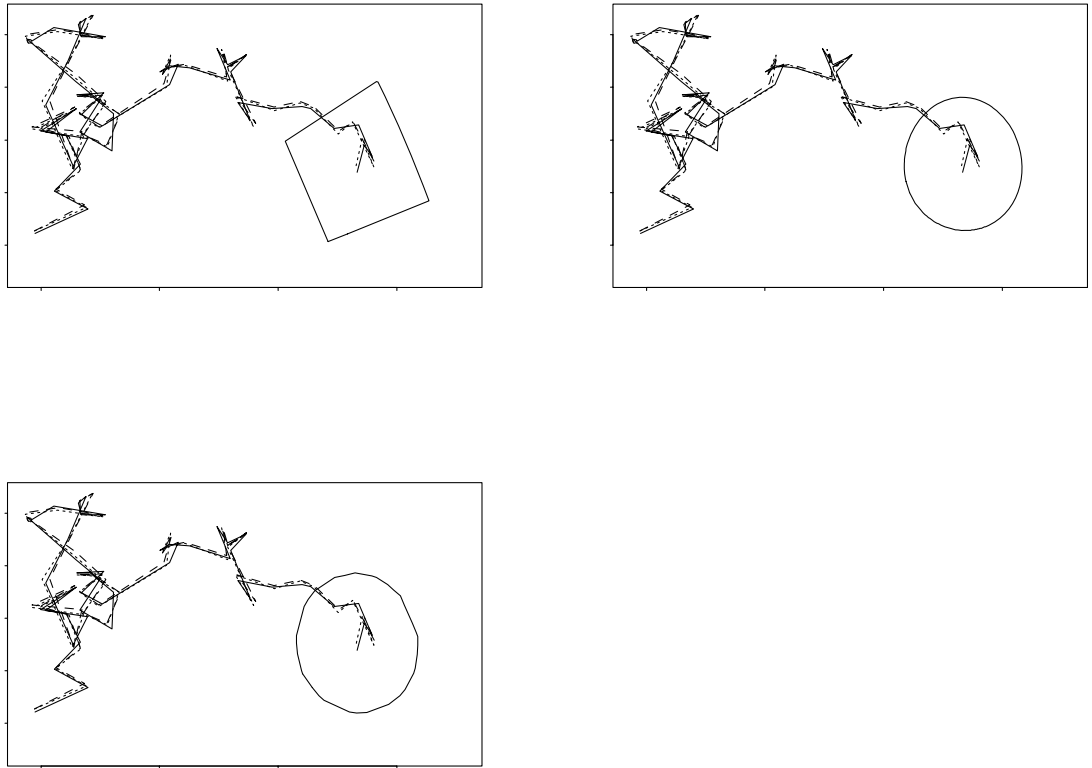


Figure 3: Classic tracking model and 95% rectangular, ellipsoidal, and highest density gates from the extended Kalman filter.

where $\sigma_v^{(0)} = 0.001$, $\sigma_v^{(1)} = 0.15$, and $s_n \in \{0, 1\}$ evolves according to a two-state Markov chain with transition probabilities

$$\begin{bmatrix} 0.8 & 0.2 \\ 0.9 & 0.1 \end{bmatrix}.$$

State 1 is a high-variability state corresponding to a maneuvering target, and state 0 corresponds to a non-maneuvering target. Maneuvers are expected to be relatively rare and short-lived.

Measurements of the target location, plotted as solid dots in Figure 4, have additive Gaussian measurement error with variance $\sigma_w^2 = (0.004)^2$. Clutter, plotted as open circles in Figure 4, is also observed. At each scan, the number of such spurious observations is a Poisson random variable with mean three. Clutter is uniformly distributed on the measurement space. No external information is available to classify observations into clutter and true measurements.

Some modifications to the filtering algorithm in (2) and (3) are required to reflect the more complex data structure of this model (Avitzour, 1995; Gordon, 1997). Let m_n be the total number of sensor observations at the n th scan, so that $\mathbf{z}_n = (z_{1n}, \dots, z_{m_n n})'$. Define \mathcal{H}_{kn} to be the event that sensor observation k comes from the target and all other observations are spurious ($k = 1, \dots, m_n$). Then for the state vector $\mathbf{x}_n = (x_n, s_n)'$, (3) becomes

$$\begin{aligned} f(\mathbf{x}_n | D_n) &\propto f(\mathbf{z}_n | \mathbf{x}_n) f(\mathbf{x}_n | D_{n-1}) \\ &= f(z_{1n}, \dots, z_{m_n n} | m_n, \mathbf{x}_n) f(m_n | \mathbf{x}_n) f(\mathbf{x}_n | D_{n-1}) \\ &\propto \sum_{k=1}^{m_n} f(z_{kn} | m_n, \mathbf{x}_n, \mathcal{H}_{kn}) \Pr\{\mathcal{H}_{kn} | m_n, \mathbf{x}_n\} f(\mathbf{x}_n | D_{n-1}) \\ &= \sum_{k=1}^{m_n} \varphi(z_{kn}; x_n, \sigma_w^2) m_n^{-1} f(\mathbf{x}_n | D_{n-1}). \end{aligned}$$

No modification is required for the computation of the state predictive density. The gate is computed using the predictive density of a valid measurement; i.e., $f(z_{kn} | \mathcal{H}_{kn}, D_{n-1})$.

A bootstrap filter (Gordon, Salmond and Smith, 1993) with prediction resample size of 1000 and update resample size of 10000 is used to approximate the sequence of filtering densities, taking into account the nonlinear evolution of the state and the presence of clutter. Highest density gates for a one-scan-ahead measurement, plotted as solid vertical lines in Figure 4, are computed at each scan using $N = 1000$ simulated observations and $M = 1000$. Note

that no actual gating is performed; all observations, even spurious ones, are used in updating the filtering densities. (Such a situation might arise if gates are used not for screening but for allocating of sensor resources.)

Means for the sequence of state predictive densities are plotted as a dashed line in Figure 4. Rectangular gates are centered about the dashed line and are scaled using the empirical variance of the measurement predictive density. The ellipsoidal gates coincide with the rectangular gates in this one-dimensional example.

Unlike the rectangular gates, the highest density gates are not necessarily symmetric about the mean prediction. In addition, the HDGs become discontinuous at a number of scans, as plausible track patterns arise in the clutter by chance. The rectangular gates respond by becoming huge, with endpoints sometimes falling outside the permissible state space. On the other hand, when the predictive density is unimodal, the rectangular gates are typically too small, failing to capture the non-Gaussian tail behavior. Overall, the highest density gates effectively draw attention away from the clutter and focus it on the part of the measurement space most likely to contain a valid observation.

5 Discussion

We have presented a method to construct highest density gates that is applicable in the nonlinear, non-Gaussian target tracking problem, regardless of the method used to approximate the filtering density. Our method makes no assumption about the form of the approximation; we only assume that simulation from some relevant probability density function in the recursive state estimation problem is feasible. Construction of the gate is carried out numerically, and is resolved by calculation of a scalar variable, \hat{q}_α , an estimate of the α th quantile of the scalar random variable $f(\mathbf{z}_{n+1}|D_n)$. Once \hat{q}_α is available, then it is a simple matter to decide whether a new measurement is inside or outside the gate.

Most filtering procedures involve some approximation to the filtering density, such as a Gaussian or Gaussian mixture approximation (e.g., Alspach and Sorenson, 1972), a numerical approximation (e.g., Budhiraja and Mirković, 1997), a Monte Carlo approximation (e.g., Gordon, Salmond and Smith, 1993), etc. The gating method we propose produces “exact” gates under any approximation to the filtering density. That is, *starting* from the approximate filtering

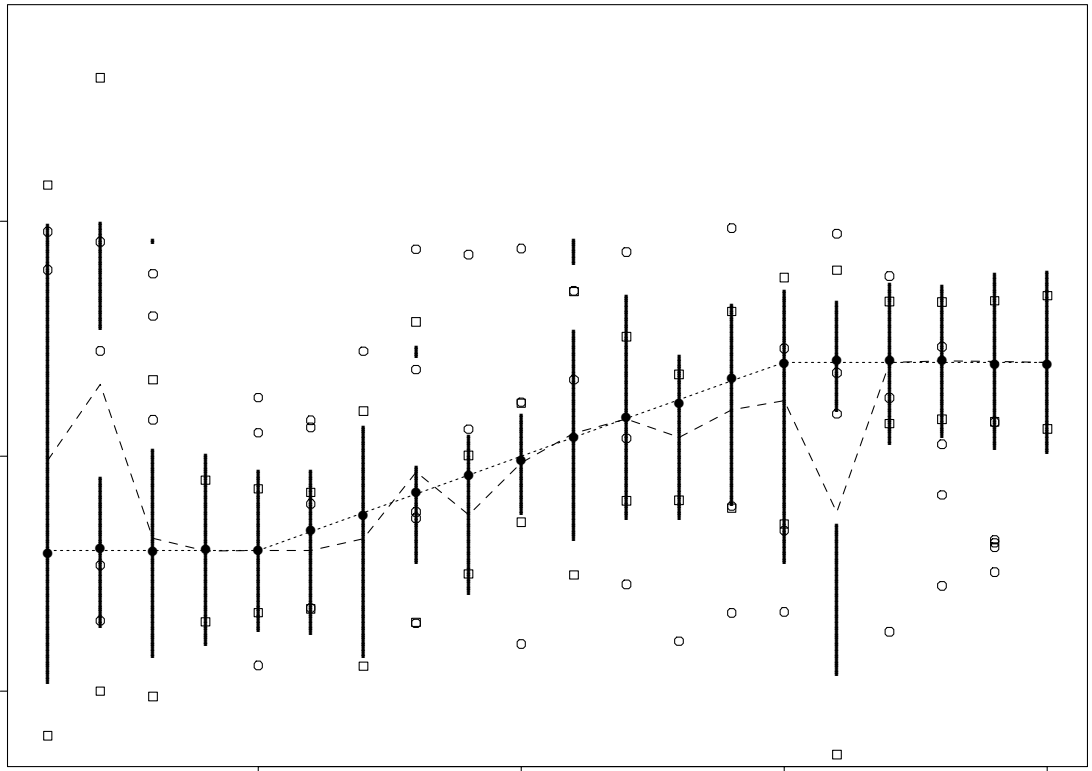


Figure 4: Multi-model and 95% gates from a bootstrap filtering algorithm. The dotted line is the true state trajectory, the solid dots are true measurements, the open circles are clutter, and the dashed line is the mean of the state predictive density at each scan. The open squares show the endpoints of the rectangular/ellipsoidal gates, and the vertical bars show the highest density gates.

density, the HDG accounts for all further nonlinearities and non-Gaussianities, and is exact up to numerical accuracy. By contrast, ellipsoidal gates are exact only under Gaussian, linear state-space models.

The performance of the gating algorithm was illustrated in Section 4 with four examples. The first two examples showed the potential advantages of highest density gates for predictive distributions with distinctly non-Gaussian features, which often arise in realistic tracking problems. The third example showed the near-correspondence of the highest density gate with the ellipsoidal gate in the application of the extended Kalman filter to the classic target tracking model (which is moderately nonlinear) with Gaussian noise. In this case, there is no apparent advantage of the highest density gate over the ellipsoidal gate, but then the model may be overly simplistic in a real target tracking situation.

The final example shows the highest density gates constructed from a bootstrap filter algorithm applied to a hidden Markov maneuver model in clutter. In this example, bimodal and skewed predictive distributions arise due to target maneuvers and clutter, and consequently the highest density gates are quite different from standard ellipsoidal and rectangular gates.

In this paper we have addressed the simplest target tracking problem, with one sensor measuring one target in clutter. We are currently working on multi-target problems. Our objective is to derive improved recursive procedures for constructing highest density gates in these more complex scenarios.

Appendix: Recursive Construction of Highest Density Gates

If simulation from $f(\mathbf{z}_{n+1} | D_n)$ is expensive, it may be desirable to re-use draws $\mathbf{z}^{(i)}$ from scan to scan. Suppose that at the n th scan, we have draws $Q_{n+1}(\mathbf{z}^{(i)})$ iid F_{n+1} . The α th quantile of this distribution determines the highest density gate for \mathbf{z}_{n+1} . For $h \geq 1$,

$$\begin{aligned} F_{n+1+h}(q) &= \Pr \{Q_{n+1+h} \leq q | D_{n+h}\} \\ &= \mathbb{E} \left[I_{\{Q_{n+1+h} \leq q\}} | D_{n+h} \right] \\ &= \frac{\int I_{\{Q_{n+1+h}(\mathbf{z}) \leq q\}} \frac{f(\mathbf{z} | D_{n+h})}{f(\mathbf{z} | D_n)} f(\mathbf{z} | D_n) d\mathbf{z}}{\int \frac{f(\mathbf{z} | D_{n+h})}{f(\mathbf{z} | D_n)} f(\mathbf{z} | D_n) d\mathbf{z}} \end{aligned}$$

$$\begin{aligned}
& \simeq \frac{\sum_{j=1}^N I_{\{Q_{n+1+h}(\mathbf{z}^{(j)}) \leq q\}} \frac{f(\mathbf{z}^{(j)} | D_{n+h})}{f(\mathbf{z}^{(j)} | D_n)}}{\sum_{j=1}^N \frac{f(\mathbf{z}^{(j)} | D_{n+h})}{f(\mathbf{z}^{(j)} | D_n)}} \\
& = \sum_{j=1}^N I_{\{Q_{n+1+h}(\mathbf{z}^{(j)}) \leq q\}} \pi_j,
\end{aligned}$$

where

$$\pi_j = \frac{\frac{f(\mathbf{z}^{(j)} | D_{n+h})}{f(\mathbf{z}^{(j)} | D_n)}}{\sum_{j=1}^N \frac{f(\mathbf{z}^{(j)} | D_{n+h})}{f(\mathbf{z}^{(j)} | D_n)}}.$$

The approximation above is an example of importance sampling (e.g., Gelman et al., 1995). The α th quantile of the resulting discrete cumulative distribution function can be computed directly, and is an approximation to the quantile needed for construction of the highest density gate for \mathbf{z}_{n+1+h} . The quality of the approximation depends on the relationship of $f(\mathbf{z}^{(j)} | D_n)$ to $f(\mathbf{z}^{(j)} | D_{n+h})$; if some measurements $\mathbf{z}^{(j)}$ are likely under the latter density, but unlikely under the former, then the approximation may be poor. This could occur, for example, if the target is maneuvering heavily, or h is large.

References

- Alspach, D.L. and Sorenson, H.W. (1972). Nonlinear Bayesian estimation using Gaussian sum approximations. *IEEE Transactions on Automatic Control* **AC-17**, 439–448.
- Anderson, B.D.O. and Moore, J.B. (1979). *Optimal Filtering*. Prentice-Hall, Englewood Cliffs, New Jersey.
- Avitzour, A. (1995). A stochastic simulation Bayesian approach to multitarget tracking. *IEE Proceedings on Radar, Sonar and Navigation* **142**, 41–44.
- Blackman, S. S. (1986). *Multiple-Target Tracking with Radar Applications*. Artech House, Massachusetts.

Box, G.E.P. and Tiao, G.C. (1973). *Bayesian Inference in Statistical Analysis*. Wiley, New York.

Budhiraja, A. and Mirković, D. (1997). A nonlinear filtering algorithm for multimodels. Unpublished manuscript, Department of Applied Mathematics, Brown University.

Di Masi, G. B., and Runggaldier, W. J. (1981). Continuous time approximations for the nonlinear filtering problem. *Applied Mathematics and Optimization* **7**, 233-245.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995). *Bayesian Data Analysis*. Chapman & Hall. New York.

Gordon, N. (1997). A hybrid bootstrap filter for target tracking in clutter. *IEEE Transactions on Aerospace and Electronic Systems* **33**, 353–358.

Gordon, N.J., Salmond, D.J., and Smith, A.F.M. (1993). A novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings on Radar and Signal Processing* **140**, 107–113.

Hyndman, R.J. (1996). Computing and graphing highest density regions. *American Statistician* **50**, 120–126.

Kallianpur, G. (1980). *Stochastic Filtering Theory*. Springer-Verlag, New York.

Kitagawa, G. (1987). Non-Gaussian state-space modeling of nonstationary time series. (with discussion.) *J. American Statistical Association* **82**, 1032–1063.

Meinhold, R.J. and Singpurwalla, N.D. (1989). Robustification of Kalman filter models. *J. American Statistical Association* **84**, 479–486.

Mood, A.M., Graybill, F.A. and Boes, D.C. (1974). *Introduction to the Theory of Statistics*, 3rd edition. McGraw-Hill, New York.