

Chapter 3

The Bayesian Approach

Bayesian inference methods [9] provide a well-studied toolkit for calculating a distribution of a quantity of interest given observed evidence (measurements). As such, they are well-suited for calculating a probability distribution of the final location of the aircraft given the data available from the Inmarsat satellite communication system. The resulting probability distribution is essential to prioritise search efforts. In this chapter, we provide a brief introduction to Bayesian methods. We assume a reasonable background in probability theory; the interested reader is referred to excellent resources such as [8, 9, 10, 18, 36, 39, 43] if further detail is desired.

The required probability density function (pdf) is the probability of the aircraft location given the available data. Bayes' rule defines a method to calculate this pdf using prior information, including knowledge of how aircraft move, and a model of how the measured data relate to the aircraft location and velocity. Mathematically, Bayes' rule is

$$p(\mathbf{x}|\mathbf{z}) = \frac{p(\mathbf{x}, \mathbf{z})}{p(\mathbf{z})} \tag{3.1}$$

$$= \frac{p(\mathbf{z}|\mathbf{x}) p(\mathbf{x})}{p(\mathbf{z})} \tag{3.2}$$

$$= \frac{p(\mathbf{z}|\mathbf{x}) p(\mathbf{x})}{\int p(\mathbf{z}|\mathbf{x}') p(\mathbf{x}') d\mathbf{x}'} \tag{3.3}$$

where the elements are:

1. \mathbf{x} is the random variable, or the *state*, which is the quantity of interest (e.g., the position of the aircraft);
2. \mathbf{z} is the measurement (e.g., the Inmarsat satellite communication data, which provides some form of positional data);
3. $p(\mathbf{x})$ is the prior pdf of the state (not incorporating the measurement, e.g., based on historical data);

4. $p(\mathbf{z}|\mathbf{x})$ is the pdf of the measurement conditioned on the state (e.g., this may be constructed by observing the distribution of measurements in cases where the state is known);
5. $p(\mathbf{x}|\mathbf{z})$ is the conditional pdf of interest (the posterior pdf), describing the distribution of state (e.g., aircraft location) taking into account the observed measurement.

The posterior probability density is based on the accumulated Inmarsat satellite communications data as well as all available contextual knowledge on the sensor characteristics, aircraft dynamic behaviour and environmental conditions and constraints. The method is based on the state space approach to time series modelling. Here, attention is focused on the state vector of a system. The state vector contains all relevant information required to describe the system under investigation at a given point in time. For example, in radar tracking problems this information would typically be related to the kinematic characteristics of the aircraft, such as position, altitude, speed, and heading. The measurement vector represents noisy observations that are related to the state vector. For example, the distance and bearing angle between the sensor and the object being measured. The state-space approach is convenient for handling multivariate data and nonlinear, non-Gaussian processes; it provides a significant advantage over traditional time series techniques for these problems; and has been extensively used in many diverse applications over the last 50 years [7]. An excellent summary of Bayesian techniques for state space models is given by [36].

In order to proceed, two models are required: first, the measurement model relates the noisy measurements to the state; and second, the system or dynamic model describes the evolution of the state with time. The measurement model used for BTO and BFO metadata is defined in a probabilistic form in Chap. 5. The dynamic model used to define the behaviour of the aircraft is defined in Chaps. 6 and 7.

If the measurement model and the system model are both linear and Gaussian, the optimal estimate can be calculated in closed form using the Kalman filter [25]. If either the system or measurement model is nonlinear or non-Gaussian, the posterior pdf will be non-Gaussian and standard analysis with a Kalman filter will be suboptimal. This results in the need for approximate computational strategies and the approach adopted in this study is introduced in this chapter. The application of the measurement and dynamics models to this approach is described in Chap. 8. The computational approach proceeds in essentially two stages: prediction and update. The prediction stage uses the aircraft dynamic model to step from the state pdf at one time to the pdf at the next time. The state is subject to unknown disturbances, modeled as random noise, and also unknown control inputs, such as turn commands, and so prediction generally translates, deforms, and broadens the state pdf. The update operation uses the latest measurement to modify (typically to tighten) the prediction pdf. This is achieved using Bayes theorem, (3.3), which is the mechanism for updating knowledge about the state in the light of extra information from new data.

3.1 The Problem and its Conceptual Solution

To define the problem of nonlinear filtering, let us introduce the state vector $\mathbf{x}(t) \in \mathbb{R}^n$, where n is the dimension of the state vector. Here t is continuous-valued time. The state evolution is best described using a continuous-time stochastic differential equation, sometimes specifically referred to as an Itô differential equation [23]. However, it is often more convenient to sample this at discrete time instants, in which case $\mathbf{x}_k \equiv \mathbf{x}(t_k)$ represents the state at the k th discrete sample time. The elapsed time between samples $\Delta_k = t_k - t_{k-1}$ is not necessarily constant. The state is assumed to evolve according to a continuous-time stochastic model:

$$d\mathbf{x}(t) = \mathbf{f}(\mathbf{x}(t), d\mathbf{v}(t), t, dt), \quad (3.4)$$

where $\mathbf{f}(\cdot)$ is a known, possibly nonlinear deterministic function of the state and $\mathbf{v}(t)$ is referred to as a process noise sequence, which caters for random disturbances in the aircraft motion.

A sensor collects measurements, which are a possibly nonlinear function of the state. Measurements occur at times t_k , for $k = \{1, 2, \dots, K\}$. The k th measurement is denoted $\mathbf{z}_k \in \mathbb{R}^m$ where m is the dimension of the measurement vector. The measurements are related to the state via the measurement equation:

$$\mathbf{z}_k = \mathbf{h}_k(\mathbf{x}_k, \mathbf{w}_k), \quad (3.5)$$

where $\mathbf{h}_k(\cdot)$ is a known, possibly nonlinear function and \mathbf{w}_k is a measurement noise sequence. The noise sequences $\mathbf{v}(t)$ and \mathbf{w}_k will be assumed to be white, with known probability density functions and mutually independent. The initial state is assumed to have a known pdf $p(\mathbf{x}_0)$ and also to be independent of noise sequences.

We seek estimates of \mathbf{x}_k based on the sequence of all available measurements up to time t_k , defining the measurement history $\mathbf{Z}_k \triangleq \{\mathbf{z}_1, \dots, \mathbf{z}_k\}$. From a Bayesian perspective, the problem is to recursively construct the posterior pdf $p(\mathbf{x}_k | \mathbf{Z}_k)$. In principle, the pdf $p(\mathbf{x}_k | \mathbf{Z}_k)$ may be obtained recursively in two stages: prediction and update. The prediction stage steps from the pdf of \mathbf{x} at time t_{k-1} , $p(\mathbf{x}_{k-1} | \mathbf{Z}_{k-1})$, to the pdf at the next time, $p(\mathbf{x}_k | \mathbf{Z}_{k-1})$, not incorporating any new measurements. The update stage takes the predicted pdf $p(\mathbf{x}_k | \mathbf{Z}_{k-1})$ and incorporates the new measurement \mathbf{z}_k occurring at time t_k to obtain the updated pdf $p(\mathbf{x}_k | \mathbf{Z}_k)$. If there is a requirement to evaluate the pdf at time t for which there is no measurement then this pdf is the predicted pdf and no update step needs to be performed.

3.1.1 Prediction

The prediction stage involves using the system model (3.4) to obtain the prediction density of the state at time step k via the Chapman–Kolmogorov equation:

$$\begin{aligned}
p(\mathbf{x}_k|\mathbf{Z}_{k-1}) &= \int p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{Z}_{k-1}) p(\mathbf{x}_{k-1}|\mathbf{Z}_{k-1}) d\mathbf{x}_{k-1}, \\
&= \int p(\mathbf{x}_k|\mathbf{x}_{k-1}) p(\mathbf{x}_{k-1}|\mathbf{Z}_{k-1}) d\mathbf{x}_{k-1}.
\end{aligned} \tag{3.6}$$

The first line of (3.6) is a statement of the law of total probability. The simplification $p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{Z}_{k-1}) = p(\mathbf{x}_k|\mathbf{x}_{k-1})$ used to progress from the first line of (3.6) to the second applies because (3.4) describes a Markov process of order one. The probabilistic model of the state evolution, $p(\mathbf{x}_k|\mathbf{x}_{k-1})$, is defined by the system equation (3.4) and the known statistics of $\mathbf{v}(t)$.

3.1.2 Update

At time t_k a measurement \mathbf{z}_k becomes available and the update stage is carried out. This involves an update of the prediction (or prior) pdf via Bayes' rule:

$$\begin{aligned}
p(\mathbf{x}_k|\mathbf{Z}_k) &= p(\mathbf{x}_k|\mathbf{z}_k, \mathbf{Z}_{k-1}) \\
&= \frac{p(\mathbf{z}_k|\mathbf{x}_k, \mathbf{Z}_{k-1}) p(\mathbf{x}_k|\mathbf{Z}_{k-1})}{p(\mathbf{z}_k|\mathbf{Z}_{k-1})} \\
&= \frac{p(\mathbf{z}_k|\mathbf{x}_k) p(\mathbf{x}_k|\mathbf{Z}_{k-1})}{p(\mathbf{z}_k|\mathbf{Z}_{k-1})},
\end{aligned} \tag{3.7}$$

where conditional independence has been used to write the likelihood function $p(\mathbf{z}_k|\mathbf{x}_k, \mathbf{Z}_{k-1}) = p(\mathbf{z}_k|\mathbf{x}_k)$, which is defined by the measurement model (3.5) and the known statistics of \mathbf{w}_k . The normalizing constant on the denominator can be expanded as

$$p(\mathbf{z}_k|\mathbf{Z}_{k-1}) = \int p(\mathbf{z}_k|\mathbf{x}_k) p(\mathbf{x}_k|\mathbf{Z}_{k-1}) d\mathbf{x}_k. \tag{3.8}$$

In the update stage (3.7), the measurement \mathbf{z}_k is used to modify the prior density to obtain the required posterior density of the current state.

Note that there is no requirement for all of the measurements to have the same statistical model or even contain the same type of information. For example, there could be multiple sensors operating on different modalities. For simplicity, we have not introduced explicit notation to change the measurement pdf for each k . For the accident flight three different types of measurement have been used. As discussed in Chap. 5, the satellite communications messages consist of R-channel and C-channel messages that have differing information content. Another quite different form of measurement is the areas of the ocean floor that have been searched without locating the aircraft and the debris that has been recovered. This measurement and its potential use to refine the ongoing search are discussed in Chap. 11.

The recurrence relations (3.6) and (3.7) form the basis for the optimal Bayesian solution. The recursive propagation of the posterior density, given by (3.6) and (3.7), is only a conceptual solution in the sense that in general it cannot be determined analytically. In most practical situations the analytic solution of (3.7) and (3.8) is intractable and numerical approximations have to be used. This has been a topic of significant research effort over the past 20 years [1, 20, 33]; a general overview of the method is presented next.

3.2 The Particle Filter

In the linear Gaussian case, the pdfs for $p(\mathbf{v}(t))$, $p(\mathbf{w}_k)$, $p(\mathbf{x}_0)$ are all Gaussian and the functions $\mathbf{f}(\cdot)$ and $\mathbf{h}(\cdot)$ are linear. It can then be easily shown that the posterior $p(\mathbf{x}_k|\mathbf{Z}_k)$ is also Gaussian and all of these pdfs can be summarised completely by their means and covariances. The Kalman filter is an algorithm that defines recursions for the mean and covariance of $p(\mathbf{x}_k|\mathbf{Z}_k)$ in terms of the means and covariances of the prior and noise processes. However, in general, the posterior does not take the same functional form as the prior and indeed it is not possible to even write a closed form expression for $p(\mathbf{x}_k|\mathbf{Z}_k)$. In this case an approximate solution is required. The solution used for the MH370 search definition is referred to as the particle filter and is a numerical approximation based on random sampling.

The fundamental concept in the particle filter is to approximate the pdf $p(\mathbf{x}_k|\mathbf{Z}_k)$ as a weighted combination of sample points

$$p(\mathbf{x}_k|\mathbf{Z}_k) \approx \sum_{p=1}^P w_k^p \delta(\mathbf{x}_k - \mathbf{x}_k^p), \quad (3.9)$$

where the w_k^p are referred to as weights and sum to unity, and the \mathbf{x}_k^p are referred to as particles. The convergence properties of this approximation in the limit as the number of particles P increases have been well studied, for example [14, 21]. Given this approximate pdf, it is simple to evaluate the expectation of any nonlinear function of the state, such as

$$\mathbb{E}[\mathbf{g}(\mathbf{x}_k)|\mathbf{Z}_k] \equiv \int \mathbf{g}(\mathbf{x}_k) p(\mathbf{x}_k|\mathbf{Z}_k) d\mathbf{x}_k \approx \sum_{p=1}^P w_k^p \mathbf{g}(\mathbf{x}_k^p). \quad (3.10)$$

The approximation of an integral using sample points as above is referred to as Monte Carlo integration and can be applied to both the Chapman–Kolmogorov prediction (3.6) and the Bayesian update (3.7).

The particle filter is an algorithm that provides a mechanism to recursively create a set of weighted particles approximating $p(\mathbf{x}_k|\mathbf{Z}_k)$ starting from a previous set of weighted particles approximating $p(\mathbf{x}_{k-1}|\mathbf{Z}_{k-1})$. It does this in two stages: first it

moves the particle sample points $\mathbf{x}_{k-1}^p \rightarrow \mathbf{x}_k^p$ to new locations using a pdf referred to as a *proposal* distribution, which is a tractable approximation of the pdf of interest. Second, it determines new particle weights to correct for the difference between the proposal and the true pdf. This process is known as importance sampling [1, 33].

The proposal distribution is a critical component of the particle filter. It is a function chosen by the designer subject to relatively loose constraints. Importantly, the proposal distribution must cover all of the state space where the true distribution is non-zero and its tails should be heavier than the tails of the true distribution. If the proposal is chosen poorly then many of the particles \mathbf{x}_k^p will be assigned very low weights and the filter efficiency will be low: a large number of particles will be required for satisfactory performance. A common version of the particle filter is the Sample-Importance-Resample (SIR) particle filter that uses the system dynamics as a proposal distribution. The SIR is popular because it is often relatively straightforward to sample from the dynamics and because the weight update equation is very simple when the dynamics is used as the proposal. The filter used in this book is a form of SIR particle filter.

For the SIR particle filter, for each particle \mathbf{x}_{k-1}^p a new \mathbf{x}_k^p is drawn from the transition density $p(\mathbf{x}_k|\mathbf{x}_{k-1}^p)$, and weights are updated by scaling the previous weights by the current measurement likelihood and re-normalising,

$$w_k^p = (W_k)^{-1} p(\mathbf{z}_k|\mathbf{x}_k^p) w_{k-1}^p, \quad (3.11)$$

where the normalising term is $W_k = \sum_{p=1}^P p(\mathbf{z}_k|\mathbf{x}_k^p) w_{k-1}^p$.

A key difficulty in particle filters is the issue of degeneracy, i.e., over time, many weights tend toward zero, and the corresponding particles are of little use. Resampling is used to combat this difficulty. The simplest approach is draw P new particles from the approximate distribution (3.9), such that particles with very large weights are likely to be replicated many times over, and those with very small weights are unlikely to be sampled. A variety of methods are possible, and can be found in [1, 33]. The sampling method used in this study is detailed in Chap. 8.

3.3 Rao–Blackwellised Particle Filter

One of the challenges in implementing a particle filter is that the number of particles required to make a good approximation to the desired posterior pdf can grow exponentially with the dimension of the state space. In some circumstances, it is possible to mitigate this by incorporating an analytic representation of the distribution of part of the state given a sample of the remainder of the state. For example, suppose that the measurement function can be decomposed into two parts

$$\mathbf{z}_k = \mathbf{h}_k^1(\mathbf{x}_k^1) + \mathbf{h}_k^2(\mathbf{x}_k^2) + \mathbf{w}_k, \quad (3.12)$$

where \mathbf{x}_k^1 and \mathbf{x}_k^2 are disjoint sub-vectors of the state \mathbf{x}_k . In this case we can write

$$p(\mathbf{x}_k^1, \mathbf{x}_k^2 | \mathbf{z}_t, \mathbf{Z}_{t-1}) = p(\mathbf{x}_k^1 | \mathbf{z}_t, \mathbf{Z}_{t-1}) p(\mathbf{x}_k^2 | \mathbf{x}_k^1, \mathbf{z}_t, \mathbf{Z}_{t-1}). \quad (3.13)$$

The two densities above can be estimated using different filters. When the function $\mathbf{h}_k^2(\mathbf{x}_k^2)$ is linear and the noise is Gaussian, the second density $p(\mathbf{x}_k^2 | \mathbf{x}_k^1, \mathbf{z}_t, \mathbf{Z}_{t-1})$ can be estimated using a Kalman filter, even if the first function $\mathbf{h}_k^1(\mathbf{x}_k^1)$ is nonlinear. The state vector that needs to be sampled is then \mathbf{x}_k^1 not $[\mathbf{x}_k^1, \mathbf{x}_k^2]$ and the sampling process can use fewer samples for a given degree of accuracy.

When a particle filter is used for the nonlinear part of the measurement problem, the conditioning of the second state density $p(\mathbf{x}_k^2 | \mathbf{x}_k^1, \mathbf{z}_t, \mathbf{Z}_{t-1})$ leads to a separate Kalman filter for each particle. Each Kalman filter uses the sampled value of the sub-state \mathbf{x}_k^1 as though it were the truth. This arrangement is referred to as a Rao–Blackwellised particle filter [15, 29, 38].

Open Access This chapter is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, duplication, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, a link is provided to the Creative Commons license and any changes made are indicated.

The images or other third party material in this chapter are included in the work's Creative Commons license, unless indicated otherwise in the credit line; if such material is not included in the work's Creative Commons license and the respective action is not permitted by statutory regulation, users will need to obtain permission from the license holder to duplicate, adapt or reproduce the material.