

Lecture Notes: Non-Gaussian Distributions

Recall that in filtering problems, state variables are always represented by distributions rather than single numbers. As figure 1 shows, “you are here” is not a single location, it is a distribution of probable locations. When the distribution is Gaussian, as in figure 1, it can be represented using a function of e :

$$p(x_{t,t}) = \exp\left(\frac{-(x_{t,t} - \bar{x}_{t,t})^2}{\sigma_{x_{t,t}}^2}\right) \quad (1)$$

where $\bar{x}_{t,t}$ is the most probable location (center of the Gaussian), and $\sigma_{x_{t,t}}^2$ is the scale of the width of the Gaussian.

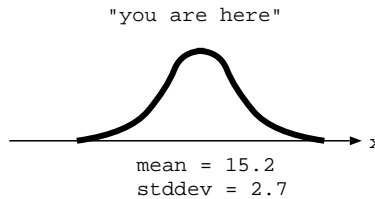


Figure 1: An example Gaussian distribution for a state variable.

However, there are many problems for which the distribution is not Gaussian. Figure 2 shows an example. For this shape, it is difficult to formalize the distribution using an equation.

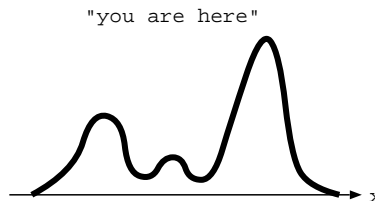
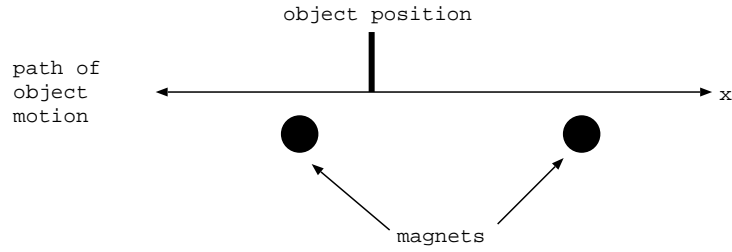
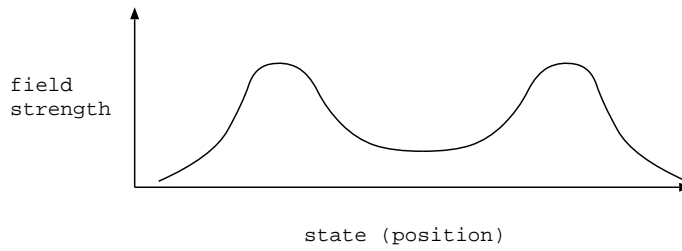


Figure 2: An example non-Gaussian distribution for a state variable.

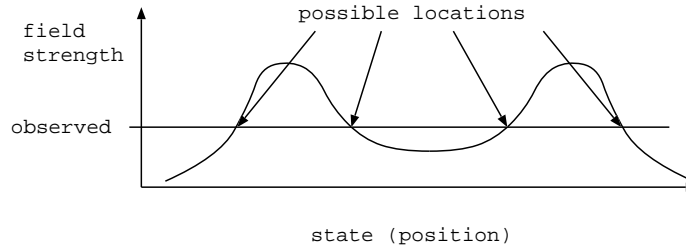
The following example illustrates a problem in which this can happen. Consider a 1D object moving in between two attractors, such as magnets. Figure 3(a) shows a digram. The object can sense an attractive force, but without direction. In other words, it senses a



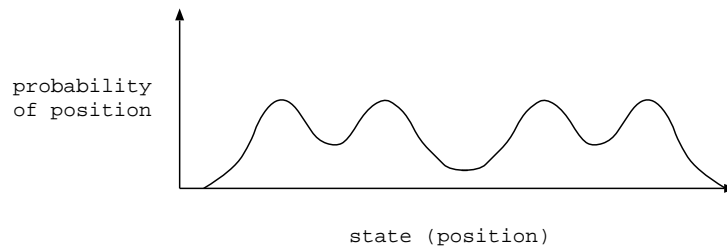
(a) Problem diagram. An object is moving back and forth along a string. Two magnets are placed at fixed positions near the string.



(b) The magnetic field strength from the sum of both magnets as a function of position on the string.



(c) The magnetic field strength sensed when at the object position shown in (a). The same reading would be obtained at four different positions on the string.



(d) The probability of position given the field strength shown in (c).

Figure 3: A problem demonstrating a non-Gaussian distribution.

magnetic field strength. Because the field strengths overlap between the two magnets, the total field strength at each position can be plotted as in figure 3(b). At the location shown in figure 3(a), the object senses a field strength that could arise from one of four possible positions, as shown in figure 3(c).

This poses the question: Given a measurement, what is the object’s probable location? In other words, what is the probability of position given a field strength? Figure 3(d) shows how this can be plotted for the example. Assume that the measurement is noisy. Therefore the field strength shown as an observation in figure 3(c) could have come from neighboring positions to the four possible locations, depending on the amount of noise in the observation. This is represented in figure 3(d) by the four hill-like shapes centered on the four locations identified in figure 3(c).

This example demonstrates how a non-Gaussian distribution can happen in a problem. Further, it demonstrates how the distribution may change over time. Looking back at figures 3(c)-(d), it is easy to see that if the object moves to another position, the whole shape of the distribution in figure 3(d) will change. Such a distribution is often called “intractable”. This means that the distribution is not easily modelled by an analytic function, or that even if it can be modelled, the function does not provide an easy mechanism for calculations.

Now recall the formula for recursive Bayesian estimation:

$$p(x_{0:t}|y_{0:t}) = \frac{p(x_t|x_{t-1})p(y_t|x_t)}{p(y_t|y_{0:t-1})}p(x_{0:t-1}|y_{0:t-1}) \quad (2)$$

Fortunately, it is applicable for any distributions, not just Gaussians. It is easy to see why Gaussians are preferred, because if each term in equation 2 is a Gaussian, then multiplying and dividing the terms is simply a calculation of adding and subtracting the exponents. However, the formula can be applied to any distribution.

In order to do so, we make use of several principles, including Monte Carlo approximation, importance sampling, and sequential importance sampling. In these notes we cover only the first of these topics. The next lecture notes covers the others.

Monte Carlo approximation is the practice of using a set of samples to approximate a distribution. Each sample is given a state and a weight. Figure 4 shows an example. In this picture, the state of each sample is represented pictorially by its position on the line. Its weight is represented by the height of the line. With enough samples, weighted and positioned appropriately, any distribution can be approximated.

Mathematically, the set of samples is denoted as

$$\chi = \{x^{(m)}, w^{(m)}\}_{m=1}^M \quad (3)$$

The notation (m) denotes an index, not an exponent. The state of each sample is represented in $x^{(m)}$ and is a vector, containing a value for each of the state variables. It may be thought of as a “guess” of the true state, and can be anywhere in the entire state space. The weight of each sample is $w^{(m)}$ and is a single number. It may be thought of as the “belief” in the guess.

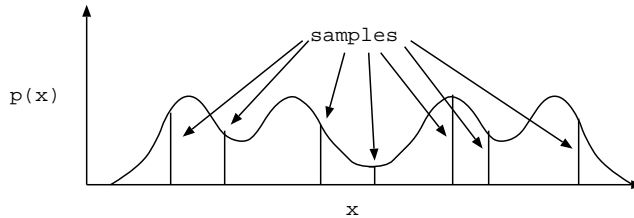


Figure 4: Monte Carlo approximation of a distribution.

If samples are selected (often called “drawn”) from the state space randomly, and with a sufficiently large M , χ can be used to approximate $p(x)$:

$$p(x) \approx \sum_{m=1}^M w^{(m)} \delta(x - x^{(m)}) \quad (4)$$

The Dirac delta function $\delta(\dots)$ is a mathematical construct used to indicate the value of a function at a given value. In this case, $\delta(x - x^{(m)})$ indicates the value of x at $x^{(m)}$.

Using the approximation, expectations can be calculated via summations instead of integrals. For example, analytically we would calculate:

$$E[x] = \int x \cdot p(x) dx \quad (5)$$

This can be approximated using χ as

$$E[x] \approx \sum_{m=1}^M x^{(m)} \cdot w^{(m)} \quad (6)$$

Note that the expected value is the mean. It makes sense to calculate this when working with Gaussian distributions, because it is the single most probable value. However, when working with non-Gaussian distributions, it is important to realize that the expected value may be a terrible “answer”. For example, in figure 3(d) the expected value is in the middle of the string, which is not a very likely location given the observation. In such a case it may make more sense to calculate local maxima or other answers. These can still be calculated from χ .

The particle filter does not work with an analytic model of the state distribution. Instead, it approximates it using a Monte Carlo method. This will be explained more next time.