Contents

1 Continuous Probability Densities

Motivation

In real robots we often deal with continuous variables and continuous state spaces. Hence, we need to extend the notion of probability to continuous variables.

1.1 Continuous Probability Densities

In robotics we typically need to model a belief over continuous, multivariate random variables $x \in \mathbb{R}^n$. We do this using **probability density functions** (PDFs) $p(x)$ over the variables x, satisfying

$$\int p(x)dx = 1. \tag{1}$$

In terms of notation, for continuous variables we use lowercase letters for random variables, and uppercase letters to denote sets of them. We drop the notational conventions of making distinctions between random variables X and their realized values x, which helped us get used to thinking about probabilities, but will get in the way of clarity below.

A **unimodal** density has a single maximum, its **mode**. In general, however, a density can have multiple modes, in which case we speak of a **multimodal** density. The **mean** of a density is defined as

$$\mu = E_p[x] = \int xp(x)dx$$

irrespective of whether the density is unimodal or multimodal. Above, the notation $E_p[.]$ stands for "the expectation of . with respect to the density p".

1.2 Gaussian Densities

A Gaussian probability density on a scalar x given **mean** μ and **variance** σ^2 is given by

$$\mathcal{N}(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}. \tag{2}$$

This density is also known as the "bell curve". This density is very popular in robotics because it is so simple: it is the just the negative exponential of a quadratic around μ. In a Bayesian probability framework we interpret densities as knowledge, and the the **standard deviation** σ indicates the uncertainty we have about the quantity x.

The other reason of the Gaussian's popularity derives from the **central limit theorem**. This theorem says that the probability of the sum of a number of random variables, no matter what the density is on them, will tend to a Gaussian density. And, it does not have to be many random variables either: summing 4 random variables distributed randomly over an interval yields a cubic density, which already matches a Gaussian quite nicely.

A **multivariate Gaussian density** on $x \in \mathbb{R}^n$is obtained by extending the notion of a quadratic to multiple dimensions. We define the squared **Mahalanobis distance,**

$$\|x-\mu\|_{\Sigma}^{2} \triangleq (x-\mu)^{\top} \Sigma^{-1} (x-\mu) \tag{3}$$

where $\mu \in \mathbb{R}^n$ is the mean. The quantity Σ is an $n \times n$ **covariance matrix**, a symmetric matrix indicating uncertainty about the mean. The Mahalanobis distance is nothing but a weighted Euclidean distance, and is the multivariate equivalent of the scalar squared distance

$$\|x-\mu\|_{\sigma^2}^{2} \triangleq \left(\frac{x-\mu}{\sigma}\right)^{2}$$

but using the matrix inverse to weight this distance metric in an n-dimensional space. Armed with this, the equation for the multivariate Gaussian is

$$\mathcal{N}(x;\mu,\Sigma) = \frac{1}{\sqrt{|2\pi\Sigma|}} \exp\left\{-\frac{1}{2}\|x-\mu\|_{\Sigma}^{2}\right\}, \tag{4}$$

where the term $|2\pi\Sigma|$ in the normalization factor denotes the determinant of $2\pi\Sigma$.

Gaussian densities, whether scalar or multivariate, have some nice properties. A Gaussian density is unimodal and the mean μ is also its mode. In addition, any scalar marginal of a multivariate Gaussian is also a Gaussian. In fact, the probability density $P(y)$ of *any* linear combination $y = Hx$, with y an m-dimensional vector and H an $m \times n$ matrix, is also Gaussian with mean $H\mu$ and covariance $H\Sigma H^T$.

Exercise

1. Given a 2-dimensional density on (x, y), with mean $\mu = (\bar{x}, \bar{y})$ and covariance matrix

$$\Sigma = \begin{bmatrix} \sigma_x^2 & r \\ r & \sigma_y^2 \end{bmatrix}$$

   what is the variance of the marginals $p(x)$ and $p(y)$? Use the linearity property above.

2. Given the same 2D density, what is the mean and variance of the sum $z = x + y$?

3. Deeper thinking: what happens to the density if we push it through a nonlinear function, e.g., $z = \sqrt{x^2 + y^2}$, the Euclidean norm of the vector (x, y)? A qualitative answer is asked for.

1.3 Bayes Nets and Mixture Models

Continuous Bayes nets are exactly like discrete Bayes nets, except with continuous variables. Most of the concepts generalize effortlessly, and we will forego very formal definitions where we can.

An example crucial to the robotics domain is shown in Figure 1, which is the continuous equivalent of the dynamic Bayes net from Part I. In the figure, the continuous Markov chain backbone represents the evolution of the continuous state x_t over time, conditioned on the **controls** u_t. Notice we use new terminology here: we say *controls* rather than actions in this new, continuous world. For example, for the Duckiebot, the control u might be two-dimensional and represent the wheel speeds of the left and right wheels, respectively. Finally, the continuous measurements z_t at the bottom are conditioned on the states x_t.

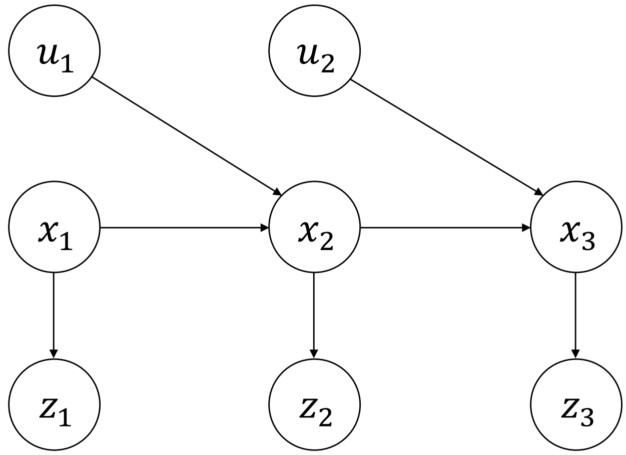

Figure 1: Continuous Bayes net modeling a robot with continuous states x_t, continuous measurements z_t, and continuous controls u_t.

We can mix and match discrete and continuous variables. A particularly simple Bayes net is a **Gaussian mixture model**. The joint density is

$$p(x, C) = p(x|C)P(C)$$

where $P(C)$ is a PMF on a discrete variable C that chooses between different Gaussians, , and $p(x|C)$ is the corresponding Gaussian mixture component. Even though each Gaussian density is unimodal, the marginal $p(x)$ is a multimodal density

$$p(x) = \sum_c p(x|C = c)p(C = c)$$

where the sum is over components. An example for a two-component mixture is shown in Figure 2.

Continuous probability densities present a representational challenge. We can no longer specify CPTs: either an equation needs to be available, as with the Gaussian density above, or an arbitrary density has to be somehow approximated. One such approximation is exactly using mixture densities: we can "mix" many simpler densities, like Gaussians, to approximate a more complicated density. This is known as **Parzen window density estimation**.

1.4 Continuous Measurement Models

In many cases it is both justified and convenient to model measurements as corrupted by zero-mean Gaussian noise. For example, a bearing measurement from a given pose $x \in SE(2)$ to a given 2D landmark l would be modeled as

$$z = h(x, l) + \eta, \tag{5}$$

where $h(.)$ is a **measurement prediction function**, and the noise η is drawn from a zero-mean Gaussian density with **measurement covariance** R. This yields the following conditional density $p(z|x, l)$ on the measurement z:

$$p(z|x, l) = \mathcal{N}(z; h(x, l), R) = \frac{1}{\sqrt{|2\pi R|}} \exp\left\{-\frac{1}{2}\left\|h(x, l) - z\right\|_R^2\right\}. \tag{6}$$

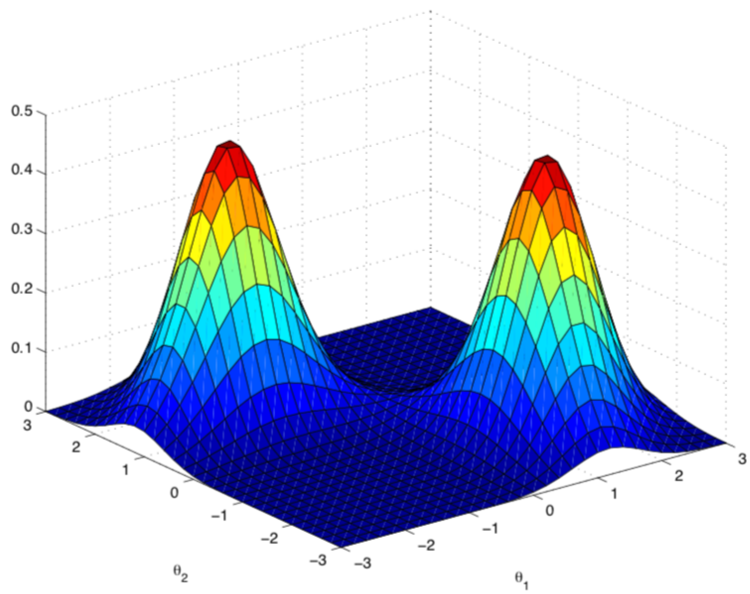

Figure 2: Gaussian mixture density with two components of about equal weight.

The measurement functions $h(.)$ are often nonlinear in practical robotics applications. Still, while they depend on the actual sensor used, they are typically not difficult to reason about or write down. The measurement function for a 2D bearing measurement is simply

$$h(x, l) = atan2(l_y - x_y, l_x - x_x) - x_\theta, \tag{7}$$

where $atan2$ is the well-known two-argument arctangent variant. Hence, the final **probabilistic measurement model** $p(z|x, l)$ is obtained as

$$p(z|x, l) = \frac{1}{\sqrt{|2\pi R|}} \exp\left\{-\frac{1}{2}\|atan2(l_y - x_y, l_x - x_x) - x_\theta - z\|_R^2\right\}. \tag{8}$$

Note that we will not *always* assume Gaussian measurement noise: to cope with the occasional data association mistake, we can use robust measurement densities, with heavier tails than a Gaussian density.

1.5 Continuous Motion Models

For a robot operating in the plane, probabilisitic **motion models** are densities of the form $p(x_{t+1}|x_t)$, specifying a **probabilistic motion model** which the robot is assumed to obey. This *could* be derived from odometry measurements, in which case we would proceed exactly as described above. Alternatively, such a motion model could arise from known control inputs u_t. In practice, we often use a conditional Gaussian assumption,

$$p(x_{t+1}|x_t, u_t) = \frac{1}{\sqrt{|2\pi Q|}} \exp\left\{-\frac{1}{2}\|g(x_t, u_t) - x_{t+1}\|_Q^2\right\}, \tag{9}$$

where $g(.)$ is a motion model, and Q a covariance matrix of the appropriate dimensionality, e.g., 3×3 in the case of robots operating in the plane.

Summary

We briefly summarize what we learned in this section:

1. Continuous probability densities generalize the notion of probability distributions to continuous random variables.
2. The scalar and multivariate Gaussian densities are useful and relatively simple.
3. Bayes nets generalize as well, and even allow for discrete and continuous variables in the same Bayes net, as in the case of mixture densities.
4. Continuous measurement models typically have a measurement prediction corrupted by noise, often modeled as Gaussian.
5. Continuous measurement models follow a similar pattern, but are conditioned on controls.