Lecture 19, Nov 22, 2023

Bayesian Tracking

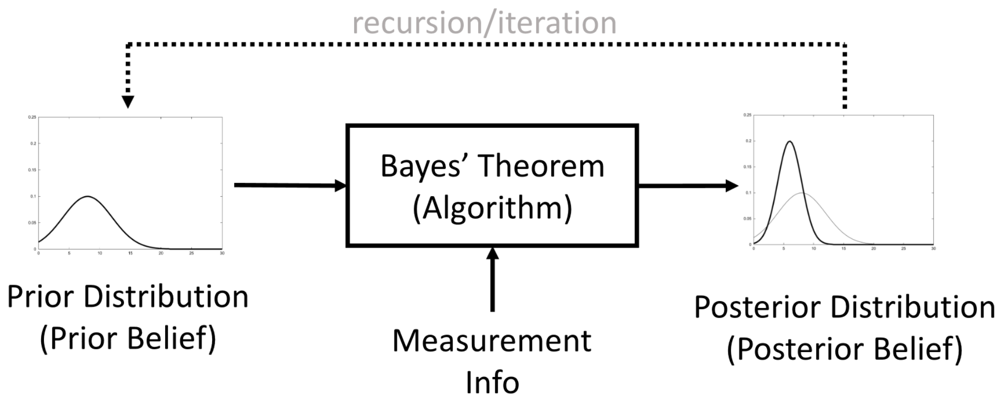

Figure 1: High-level overview of Bayesian localization.

- We wish to derive a recursive state estimation algorithm (i.e. iterating at each timestep) for a system with a finite state space, consisting of two main steps:
  1. The *prior update*, where the state estimate is predicted forward using the process model
  2. The *measurement update*, where the prior is combined with observation and measurements to correct it
- Let $\boldsymbol{x}_k \in \mathcal{X}$ be the vector-valued state at time k (assumed discrete, i.e. $\mathcal{X}$ is finite); let $\boldsymbol{y}_k$ be a vector-valued measurement that we can observe (continuous or discrete)
- We have a motion model $\boldsymbol{x}_k = \boldsymbol{f}_{k-1}(\boldsymbol{x}_{k-1}, \boldsymbol{v}_{k-1})$ and the observation model $\boldsymbol{y}_k = \boldsymbol{h}_k(\boldsymbol{x}_k, \boldsymbol{w}_k)$, where $\boldsymbol{v}_k, \boldsymbol{w}_k$ are independent noise terms with known PDFs; we also assume noise is independent of the initial condition $\boldsymbol{x}_0$
  - Note $\boldsymbol{u}_{k-1}$ is not explicitly included, but we can incorporate it by absorbing it into $\boldsymbol{f}_{k-1}$ and $\boldsymbol{h}_k$
- Let $\boldsymbol{y}_{1:k} = \{\, \boldsymbol{y}_1, \dots, \boldsymbol{y}_k \,\}$; we want to calculate $f(\boldsymbol{x}_k|\boldsymbol{y}_{1:k})$, i.e. the probability distribution of the state at time k, given all our measurements
- Assuming the *Markov property* (i.e. each state only depends on the prior state, and not the state history), we can formulate the problem as computing $f(\boldsymbol{x}_k|\boldsymbol{y}_{1:k})$ from $f(\boldsymbol{x}_{k-1}|\boldsymbol{y}_{1:k-1})$
- Prior update: compute $f(\boldsymbol{x}_k|\boldsymbol{y}_{1:k-1})$ in terms of $f(\boldsymbol{x}_{k-1}|\boldsymbol{y}_{1:k-1})$
  - By total probability, $f(\boldsymbol{x}_k|y_{1:k-1}) = \displaystyle\sum_{x_{k-1} \in \mathcal{X}} f(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, \boldsymbol{y}_{1:k-1}) f(\boldsymbol{x}_{k-1}|\boldsymbol{y}_{1:k-1})$
    * i.e. we introduce $\boldsymbol{x}_{k-1}$ and marginalize across it
  - $\boldsymbol{x}_k$ and $\boldsymbol{y}_{1:k-1}$ are conditionally independent given $\boldsymbol{x}_{k-1}$, because the distribution of $\boldsymbol{x}_{k-1}$ already incorporates the information from all previous measurements
    * $\boldsymbol{x}_k$ is a function of $\boldsymbol{v}_{k-1}$ only (because $\boldsymbol{x}_{k-1}$ is known)
    * $\boldsymbol{y}_{k-1}$ is a function of $\boldsymbol{w}_{k-1}$
    * $\boldsymbol{y}_{k-2}$ is a function of $\boldsymbol{x}_{k-2}$ and $\boldsymbol{w}_{k-2}$, but $\boldsymbol{x}_{k-2}$ is a function of $\boldsymbol{x}_{k-3}$ and $\boldsymbol{v}_{k-3}$, and so on
    * Therefore $\boldsymbol{y}_{1:k-1}$ is a function of $\boldsymbol{x}_{k-1}, \boldsymbol{v}_{1:k-3}, \boldsymbol{w}_{1:k-1}, \boldsymbol{x}_0$
    * $\boldsymbol{x}_k$, and $y_{1:k-1}$ depend only on random variables that are independent, so these two variables must be independent
  - Therefore the prior update is $f(\boldsymbol{x}_k|\boldsymbol{y}_{1:k-1}) = \displaystyle\sum_{x_{k-1} \in \mathcal{X}} f(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}) f(\boldsymbol{x}_{k-1}|\boldsymbol{y}_{1:k-1})$
    * The distribution $f(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})$ can be calculated exactly from our process model and noise distribution using change of variables
- Measurement update: compute $f(\boldsymbol{x}_k|y_{1:k-1})$, given $\boldsymbol{y}_k$ and $f(\boldsymbol{x}_k|\boldsymbol{y}_{1:k-1})$

- Using Bayes' rule, $f(\boldsymbol{x}_k|\boldsymbol{y}_{1:k}) = f(\boldsymbol{x}_k|\boldsymbol{y}_k, \boldsymbol{y}_{1:k-1})$
$$= \frac{f(\boldsymbol{y}_k|\boldsymbol{x}_k, \boldsymbol{y}_{1:k-1})f(\boldsymbol{x}_k|\boldsymbol{y}_{1:k-1})}{f(\boldsymbol{y}_k|\boldsymbol{y}_{1:k-1})}$$
- Once again, $\boldsymbol{y}_k$ and $\boldsymbol{y}_{1:k-1}$ are conditionally independent, given $\boldsymbol{x}_k$
  * $\boldsymbol{y}_k$ is a function of only $\boldsymbol{w}_k$, if given $\boldsymbol{x}_k$
  * Using a similar procedure we can show $\boldsymbol{y}_{1:k-1}$ is a function of $\boldsymbol{v}_{0:k-2}, \boldsymbol{w}_{1:k-1}, \boldsymbol{x}_0$, all of which are independent of $\boldsymbol{w}_k$
  * Therefore $\boldsymbol{y}_k, \boldsymbol{y}_{1:k-1}$ are conditionally independent on $\boldsymbol{x}_k$
  * $f(\boldsymbol{y}_k|x_k, \boldsymbol{y}_{1:k-1}) = f(\boldsymbol{y}_k|\boldsymbol{x}_k)$, and can be computed from our measurement model
- The term in the denominator is simply a normalization constant
  * We can compute it as $f(\boldsymbol{y}_k|\boldsymbol{y}_{1:k-1}) = \sum_{\boldsymbol{x}_k \in \mathcal{X}} f(\boldsymbol{y}_k|\boldsymbol{x}_k)f(\boldsymbol{x}_k|\boldsymbol{y}_{1:k-1})$ by total probability
- Therefore the measurement update is $f(\boldsymbol{x}_k|\boldsymbol{y}_{1:k}) = \dfrac{f(\boldsymbol{y}_k|\boldsymbol{x}_k)f(\boldsymbol{x}_k|\boldsymbol{y}_{1:k-1})}{\sum_{\boldsymbol{x}_k \in \mathcal{X}} f(\boldsymbol{y}_k|\boldsymbol{x}_k)f(\boldsymbol{x}_k|\boldsymbol{y}_{1:k-1})}$

Implementation

- Enumerate the state as $\mathcal{X} = \{1, 2, \ldots, N\}$
- Define $\boldsymbol{a}^i_{k|k} = \Pr(\boldsymbol{x}_k = i|\boldsymbol{y}_{1:k-1}), i = 1, \ldots, N$ as an array of N elements in which we store the posterior
  - Initialize $\boldsymbol{a}^i_{0|0} = \Pr(\boldsymbol{x}_0 = i)$
- Define $\boldsymbol{a}^i_{k|k-1} = \Pr(\boldsymbol{x}_k = i|\boldsymbol{x}_{1:k-1}), i = 1, \ldots, N$ to store the prior
- Recursive update:
  - $\boldsymbol{a}^i_{k|k-1} = \sum_{j=1}^{N} \Pr(\boldsymbol{x}_k = i|\boldsymbol{x}_{k-1} = j)\boldsymbol{a}^j_{k-1|k-1}$
    * $\Pr(\boldsymbol{x}_k = i|\boldsymbol{x}_{k-1} = j)$ can be calculated from $\boldsymbol{x}_k = \boldsymbol{f}_{k-1}(\boldsymbol{x}_{k-1}, \boldsymbol{v}_{k-1})$ and the distribution of $\boldsymbol{v}_k$
  - $\boldsymbol{a}^i_{k|k} = \dfrac{f(\boldsymbol{y}_k|\boldsymbol{x}_k = i)\boldsymbol{a}^i_{k|k-1}}{\sum_{j=1}^{N} f(\boldsymbol{y}_k|\boldsymbol{x}_k = j)\boldsymbol{a}^j_{k|k-1}}$
    * $f(\boldsymbol{y}_k|\boldsymbol{x}_k = i)$ can be calculated from $\boldsymbol{y}_k = \boldsymbol{h}_k(\boldsymbol{x}_k, \boldsymbol{w}_k)$ and the distribution of $\boldsymbol{w}_k$

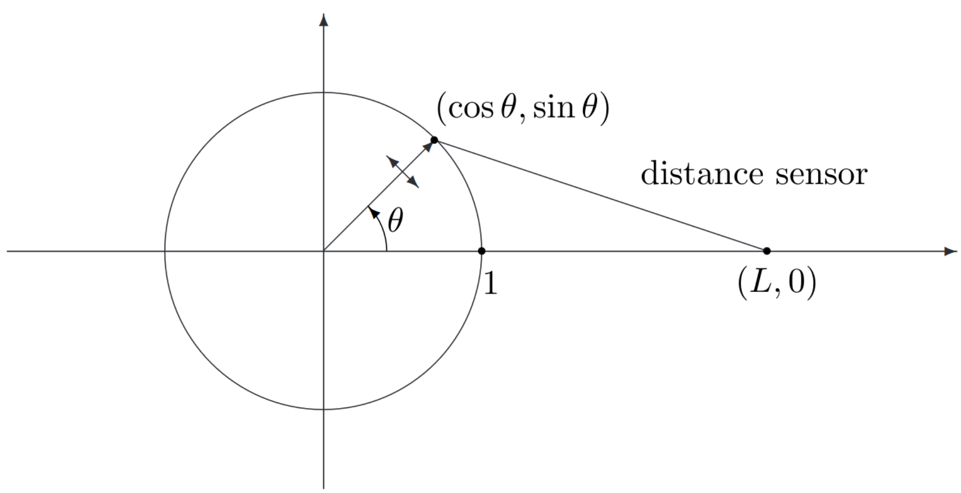

Figure 2: Setup for the example problem.

- Example: consider an object moving randomly on a circle, in discrete steps; our measurement is the distance to the object from a distance sensor located at $(L, 0)$
  - Let x_k be the object's location on the circle, then $\theta_k = \dfrac{2\pi x_k}{N}$
  - Set up the models:

    * The process model is $x_k = f(x_{k-1}, v_{k-1}) = (x_{k-1} + v_{k-1}) \bmod N$
    * The process noise is 1 with probability p, and -1 with probability $1 - p$
    * The measurement model is $y_k = h(x_k, w_k) = \sqrt{(L - \cos\theta_k)^2 + \sin^2\theta_k} + w_k$
    * The measurement noise is uniformly distributed over $[-e, e]$
- Using a change of variables we can now compute the PDFs of the process and sensor models
    * $f(x_k|x_{k-1}) = \begin{cases} p & x_k = (x_{k-1} + 1) \bmod N \\ 1 - p & x_k = (x_{k-1} - 1) \bmod N \\ 0 & \text{otherwise} \end{cases}$
    * $f(y_k|x_k) = \begin{cases} \frac{1}{2e} & \left| y_k - \sqrt{(L - \cos\theta_k)^2 + \sin\theta^2\theta_k} \right| \le e \\ 0 & \text{otherwise} \end{cases}$
- Initialize as $f(x_0) = \frac{1}{N} \forall x_0 \in \{ 0, 1, \ldots, N - 1 \}$ which assumes a state of maximum ignorance