Statistical Techniques in Robotics (16-831, F10) Lecture #02 (Thursday August 26)

# Introduction to Filtering

*Lecturer: Drew Bagnell* *Scribe: Hatem Alismail*

## 1 Probability

A Random Variable[1] is a function from the probability space to some other space, usually $\mathbb{R}^N$. Usually, random variables are denoted with a capital letter (e.g. $X$, $Y$, $A$ ...). For example, a random variable might be a coin flip, which takes one of two possible values: head or tail. For a fair coin: $P(X = head) = P(X = tail) = 0.5$.

In the discreet random variable case, such as a coin flip, the probability mass function (PMF) is a function that assigns a probability value for random variable taking a specific value. Values of the PMF must sum up to 1.

In the case of a continuous random variable, such as a sonar reading, the probability distribution function (PDF) is the function that assigns a probability value for a random variable being within an interval.

The PDF must always integrate to 1, $\int p(x)dx = 1$

**Note 1.** *For a continuous random variable $p(X = x) = 0$. That is, the probability the random variable taking the specific value $x$ is always zero.*

**Note 2.** *Only the integral $\int p(x)dx$, not $p(x)$, over all possible values integrates to 1.*

### 1.1 Probability Axioms

The axioms of probability are:

- Probabilities must be bounded between 0 and 1, $0 <= P(x) <= 1$
- Probability of truth is always 1, $P(\texttt{true}) = 1$
- Probability of false is always 0, $P(\texttt{false}) = 0$
- Probability of $X$ or $Y$ is the probability of $X$ plus the probability of $Y$ minus the probability of the intersection of $X$ and $Y$ (see Figure 1), $P(X \vee Y) = P(X) + P(Y) - P(X \wedge Y)$

From these axioms only we can *prove* several other statements, such as $P(\neg X) = 1 - P(X)$.

The Gaussian PDF is probably the most famous/useful PDF, which is parameterized by two values, the mean ($\mu$) and the variance ($\sigma^2$). It can be written as:

$$p(x) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right) \tag{1}$$

[1]It is not random and it is not a variable, it is easier to think of a random variable as an *unknown quantity.*

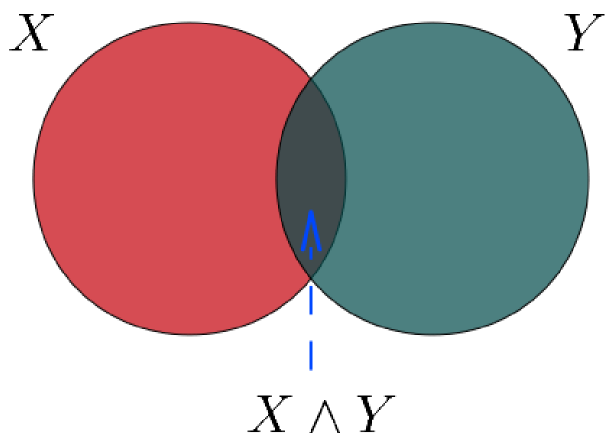


Figure 1: The probability of the event $X$ *or* $Y$ is the is the area of the red circle + the area of the blue circle - the intersection of the two circles so that we do not over count.

This definition can be generalized for $|x| > 1$, by replacing the mean with a mean vector and replacing the variance with a symmetric positive semidefinite matrix, called the covariance matrix $(\Sigma)$. This is called a multivariate Gaussian distribution can be written as:

$$p(x) = \det(2\pi\Sigma)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right) \tag{2}$$

#### 1.1.1 Bayes Rule

Relates $p(x|y)$ to its "inverse", $p(y|x)$. For $p(y) > 0$:

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)} = \frac{p(y|x)p(x)}{\sum_{x'} p(y|x')p(x')} \tag{3}$$

In many situations, it is convenient to avoid writing out the normalization factor explicitly. For this purpose, $\eta$, or $1/Z$ is used to denote the normalizer: $p(x|y) = \eta p(y|x)p(x)$.

## 2 State Estimation

State estimation is "the problem of estimating quantities from sensor data that are not directly observable, but can be inferred."

**Definition 3.** *A state is aspects of the environment and the robot that may impact the future.*

There are two types of states, a *static* state, such as walls, and a *dynamic* state, such as the robot's pose. The state at time $t$ will be denoted with $x_t$.

# 3 Filtering

In filtering, we want to obtain an estimate of $x_t$ over time. For example, in a localization problem $x_t$ is the pose of the robot. In mapping, the state is the map of the world, and in SLAM, the state is both, the robot's pose and the map of the environment.

Inputs to a general filtering problem are:

- Data: time indexed set of observations ($z_{1:t}$) and control actions ($u_{1:t}$). Generally, the robot observers the environment and performs a control action at every time step. Data can be denoted as: $d = \{z_{1:t}, u_{1:t}\}$.
- Probability distribution of the initial state $p(x_0)$.
- Motion/action model that relates the probability of the current state to the previous state and observations: $p(x_t|x_{t-1}, u_t)$
- Observation/measurement model, which is a model of how measurements are generated: $p(z_t|x_t)$.

See Figure 2 to see the evolution of control, states and observations.

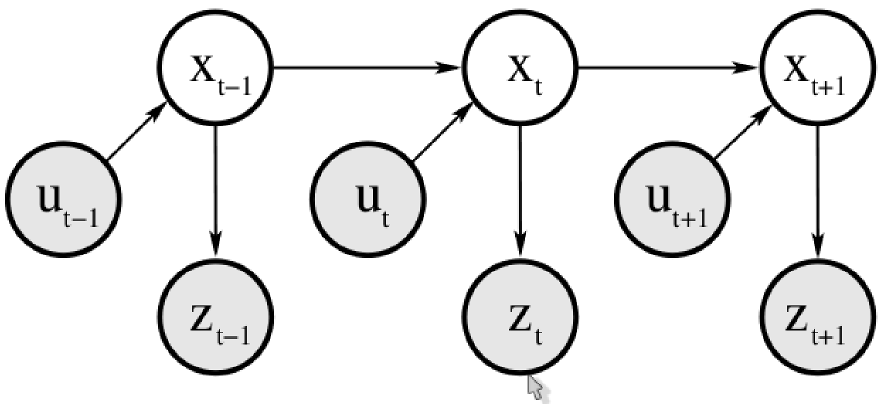


Figure 2: Evolution of controls, states, and observations

## 3.1 Belief distribution

A belief is a reflection of the robot's internal knowledge about the state and it is represented through conditional probability distributions.

$$bel(x_t) = p(x_t|z_{1:t}, u_{1:t}) \tag{4}$$

, which is the posterior probability over state variables conditioned on data (observations and actions).

### 3.2 The Bayes Filter

Bayes filter is a general algorithm to compute belief from observations and control data. A discrete Bayes filter algorithm is shown in Algorithm 1.

**Algorithm 1** Discrete_Bayes_Filter $(Bel(x), d)$

$\eta = 0$<br>
**if** $d$ is a *perceptual* data item $z$ **then**<br>
    **for all** $x$ **do**<br>
        $Bel'(x) = P(z|x)Bel(x)$<br>
        $\eta = \eta + Bel`(x)$<br>
    **end for**<br>
    **for all** $x$ **do**<br>
        $Bel'(x) = \eta^{-1}Bel'(x)$<br>
    **end for**<br>
**else if** $d$ is an *action* data item $u$ **then**<br>
    **for all** $x$ **do**<br>
        $Bel'(x) = \sum_{x'} P(x|u, x')Bel(x')$<br>
    **end for**<br>
**end if**<br>
**return** $Bel'(x)$

### 3.3 Derivation of the Bayes filter

Below is the mathematical derivation of the Bayes filter:

$$
\begin{aligned}
Bel(x_t) &= P(x_t|u_1, z_1, \ldots, u_t, z_t) \\
&= \eta P(z_t|x_t, u_1, z_1, \ldots, z_{t-1}, u_t)P(x_t|u_1, z_1, \ldots, z_{t-1}, u_t) && \text{Bayes rule} \\
&= \eta P(z_t|x_t)P(x_t|u_1, z_1, \ldots, z_{t-1}, u_t, x_{t-1}) && \text{Markov} \\
&= \eta P(z_t|x_t) \int P(x_t|, u_1, z_1, \ldots, z_{t-1}, u_t)P(x_{t-1}|u_1, z_1 \ldots, z_{t-1}, u_t)dx_{t-1} && \text{Total probability} \\
&= \eta P(z_t|x_t) \int P(x_{t-1}|u_t, x_{t-1})P(x_{t-1}|z_1, u_1, \ldots, z_{t-1}, u_{t-1})dx_{t-1} && \text{Markov (removed } u_t\text{)} \\
&= \eta P(z_t|x_t) \int P(x_t|u_t, x_{t-1})Bel(x_{t-1})dx_{t-1} && \text{Def. of } Bel(x_t)
\end{aligned}
$$

### 3.4 The Markov Assumption

The Markov assumption states that current state is conditionally independent from anything else given the previous state. While the Markovian world assumption may not be realistic in many situations[2], it simplifies the computations needed for state estimation.

[2]For example, correlations between adjacent laser beams. If there was a specular reflection at beam $a$, then it is likely that neighboring beams will experience similar behavior.

For an observation model, the Markov assumption can be expressed as:

$$P(z_t|x_0, x_1, u_1, z_1, \ldots, x_t, u_t, z_{t-1}) = P(z_t|x_t) \tag{5}$$

And for an action model[3]

$$P(x_{t+1}|x_0, x_1, u_1, z_1, \ldots, x_t, u_t, z_{t-1}) = P(x_{t+1}|x_t, u_t, z_{t-1}) \tag{6}$$

## 4 Example: Markov (Grid-based) Localization

Markov localization uses a histogram filter. The basic idea is to discretize the (continuous) state space into a number of bins. Each of these bins has a probability mass associated with it. If the number of bins is sufficient, the probability mass distribution over these bins is an accurate approximation of the underlying distribution in the state space. The integrals in the Bayes Filter are replaced with summations which are easier to compute and which need to be calculated over a finite number of cells.

For Markov localization, the state space is the pose of the robot. This space is quantized into a rectangular grid and each cell ($k$) in the grid has a probability mass ($p_{k,t}$) associated with it that equals the probability of the robot being somewhere within that cell in the pose space. Motion causes the pose distribution to shift and flatten out due to the increased uncertainty in position while measurements tend to create peaks in the distribution.[4]

Figure 3 shows an example of Markov Localization, a likelihood estimate is maintained for all the cells. Initially, (a) the robot does not know its position, hence the belief is uniform. (b) shows the robot's belief about its location after observing a door, the belief is larger in locations of doors in the map. However, a single measurement is not enough to resolve the 3 doors ambiguity; the robot can be in front of any of the doors. In (c) the robot moves, and the belief of its location shift in the direction of movement. In (d), the robot observes another door. Based on this observation, uncertainty is reduced, and robot is able to localize itself fairly accurately. Finally, (e) as the robot moves uncertainty increases.

It is worth mentioning that the number of cells grows exponentially in the number of dimensions, so for high dimensional states this method may not be practical.

**Algorithm 2** grid_based_localizer($\{p_{k,t-1}\}, u_t, z_t, m$)

1: **for all** $k$ **do**
2: $\quad p'_{k,t} = \sum_i p_{i,t-1}$**motion_model**$(x_k, u_t, x_i)$
3: $\quad p_{k,t} = \eta p'_{k,t}$**measurement_model**$(z_t, x_k, m)$
4: **end for**
5: **return** $\{p_{k,t}\}$

[3]In the slides, there was a missing $z_{t-1}$

[4]In general, actions increase uncertainty while measurements reduce uncertainty.

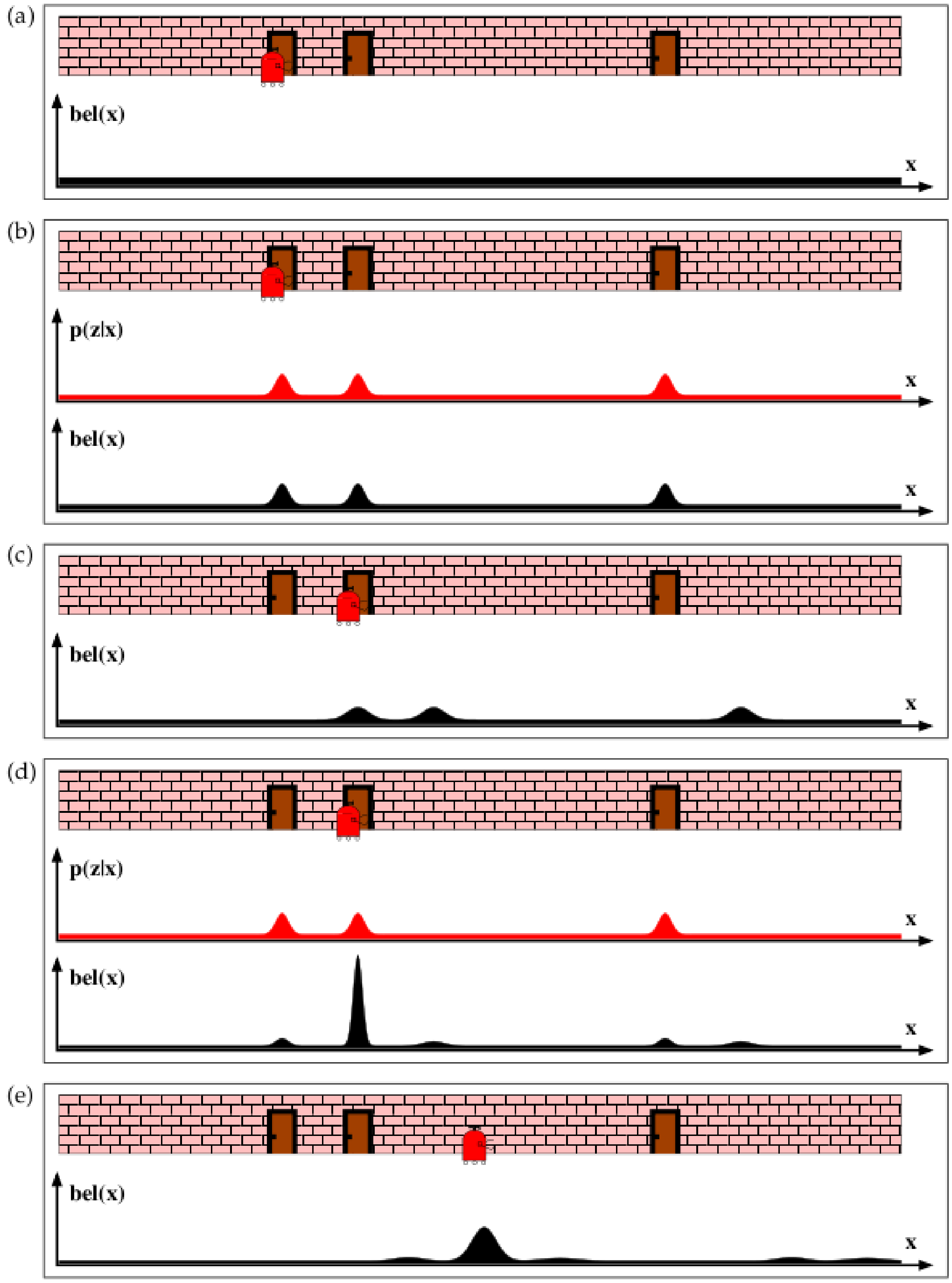


Figure 3: Basic idea of Markov localization: a robot during global localization

# 5 Forward Sensor Model

Bayesian Filtering techniques for robot state estimation require some probability density function that gives the likelihood of a measurement from a certain pose. This is $p(z_t|x_t, m)$ and is called the forward sensor model. A typical range finding sensor like a laser will return a number of measurements in a single scan $z_t = \{z_t^1, z_t^2, \ldots, z_t^K\}$. In reality measurements in a scan will depend on each other, but modeling this dependency is complicated so it will be assumed that scan readings are independent although this can lead to overconfident likelihoods.

$$p(z|x, m) = \prod_{k=1}^{K} p\left(z^k|x, m\right) \tag{7}$$

There are a number of sources of error in measurements (See Figure 4). These are

- Readings from objects not present in the map (people, furniture, etc)
- Noise around a 'true' measurement
- Maximum range reading (e.g. no return indicating $80m$ max range for SICK)
- 'Random' measurements

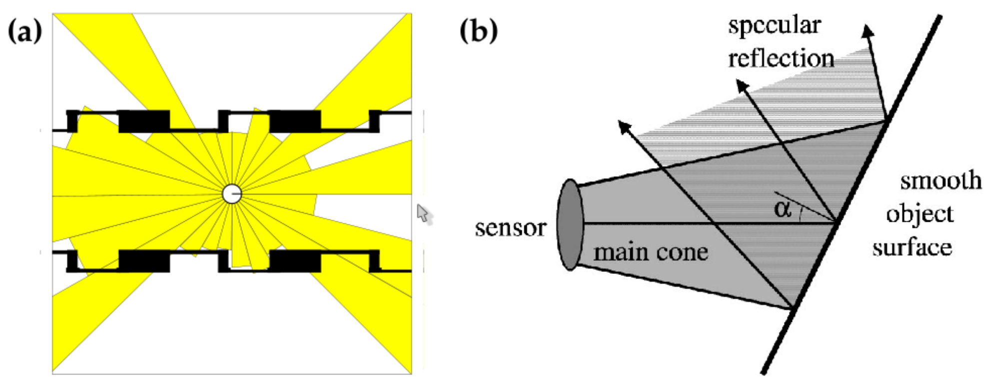


Figure 4: (a) typical ultrasound scan, (b) misreading in ultrasonic SENsing.(Figure 6.1 in the book)

## 5.1 Beam Models of Range Finders

Range finders measure range to nearby objects along a beam, such as laser ranger finders, or along a cone, such as ultrasonic sensors. The desired model $p(z_t|x,m)$ is a mixture of four densities that correspond to types of error (shown in Figure 5, which are:

1. **Local measurement noise.** In an ideal world, we'd except the range finder to give a perfect answer. However, due to some factors such as the finite resolution of the scanner, the atmosphere of the environment, etc. we'd expect to see some noise

2. **Unexpected objects.** Accounts for dynamic objects (such as people) that are absent from the robot's static map.

3. **Failures (max range).** In some cases of a reflection, or transparent objects, range measurements are reported as maximum range and fails to report a correct range.

4. **Random measurement.** When the sensor is *really* having a bad day...

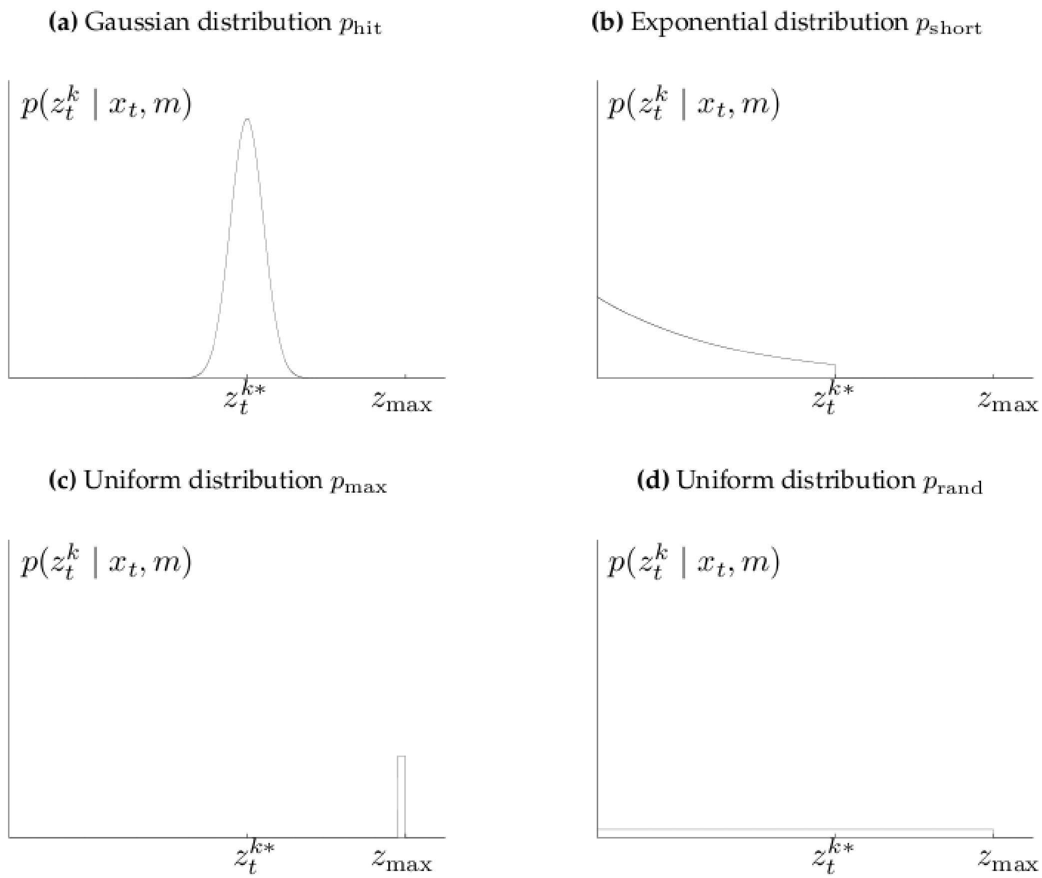


**Figure 6.3** Components of the range finder sensor model. In each diagram the horizontal axis corresponds to the measurement $z_t^k$, the vertical to the likelihood.

Figure 5: Components of error model

We can use a linear combination of the distributions to get a "Pseudo-density" distribution (Figure 6). The mixing coefficients should sum to one and can be determined experimentally by collecting ground truth range data and finding the coefficients that provide the best fit.

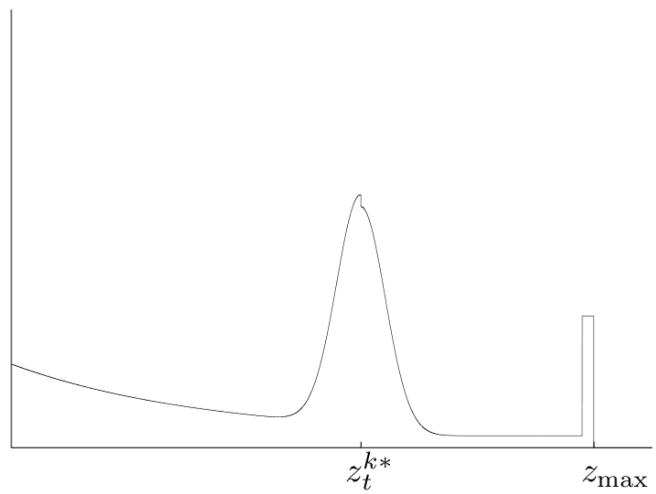


**Figure 6.4** "Pseudo-density" of a typical mixture distribution $p(z_t^k \mid x_t, m)$.

Figure 6: Combined error model

## Acknowledgments

- "Probabilistic Robotics" by S. Thurn, W. Burgard, and D. Fox.
  - Figures and equations
- Lecture notes from Fall 2008 by Supreeth Achar
  - Section 4 with some modifications.
  - Section 5 with some modifications.