A Probability Primer for Robotics Students

Wolfram Burgard
University of Freiburg
Department of Computer Science
Germany
burgard@informatik.uni-freiburg.de
http://www.informatik.uni-freiburg.de/~burgard

Tutorial Goal
To familiarize you with probabilistic paradigm in robotics

- Axioms
- Basic techniques
- Bayes rule
- Recursive Bayesian updating
- Bayes filters
Robots are Inherently Uncertain

- Uncertainty arises from four major factors:
  - Environment stochastic, unpredictable
  - Robot stochastic
  - Sensor limited, noisy
  - Models inaccurate

Nature of Sensor Data

- Odometry Data
- Range Data
Advantages of Probabilistic Paradigm

- Can accommodate inaccurate models
- Can accommodate imperfect sensors
- Robust in real-world applications
- Best known approach to many hard robotics problems

Pitfalls

- Computationally demanding
- False assumptions
- Approximate
Axioms of Probability Theory

Pr(A) denotes probability that proposition A is true.

- 0 ≤ Pr(A) ≤ 1
- Pr(True) = 1  Pr(False) = 0
- Pr(A ∨ B) = Pr(A) + Pr(B) − Pr(A ∧ B)

A Closer Look at Axiom 3

Pr(A ∨ B) = Pr(A) + Pr(B) − Pr(A ∧ B)
Using the Axioms

\[
\begin{align*}
\Pr(A \lor \neg A) &= \Pr(A) + \Pr(\neg A) - \Pr(A \land \neg A) \\
\Pr(True) &= \Pr(A) + \Pr(\neg A) - \Pr(False) \\
1 &= \Pr(A) + \Pr(\neg A) - 0 \\
\Pr(\neg A) &= 1 - \Pr(A)
\end{align*}
\]

Discrete Random Variables

- \(X\) denotes a \textit{random variable}.
- \(X\) can take on a finite number of values in \(\{x_1, x_2, \ldots, x_n\}\).
- \(P(X=x_i)\), or \(P(x_i)\), is the \textit{probability} that the random variable \(X\) takes on value \(x_i\).
- \(P(\cdot)\) is called \textit{probability mass function}.

- E.g. \(P(Room) = \langle 0.7, 0.2, 0.08, 0.02 \rangle\)
Continuous Random Variables

- $X$ takes on values in the continuum.
- $p(X=x)$, or $p(x)$, is a probability density function.
  \[ \Pr(x \in [a,b]) = \int_a^b p(x) \, dx \]
- E.g.

Joint and Conditional Probability

- $P(X=x \text{ and } Y=y) = P(x,y)$
- If $X$ and $Y$ are independent then
  \[ P(x,y) = P(x) \, P(y) \]
- $P(x \mid y)$ is the probability of $x$ given $y$
  \[ P(x \mid y) = \frac{P(x,y)}{P(y)} \]
  \[ P(x,y) = P(x \mid y) \, P(y) \]
- If $X$ and $Y$ are independent then
  \[ P(x \mid y) = P(x) \]
### Law of Total Probability, Marginals

<table>
<thead>
<tr>
<th>Discrete case</th>
<th>Continuous case</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ \sum_x P(x) = 1 ]</td>
<td>[ \int p(x) , dx = 1 ]</td>
</tr>
<tr>
<td>[ P(x) = \sum_y P(x, y) ]</td>
<td>[ p(x) = \int p(x, y) , dy ]</td>
</tr>
<tr>
<td>[ P(x) = \sum_y P(x \mid y) P(y) ]</td>
<td>[ p(x) = \int p(x \mid y) p(y) , dy ]</td>
</tr>
</tbody>
</table>

### Bayes Formula

\[ P(x, y) = P(x \mid y) P(y) = P(y \mid x) P(x) \]

\[ \Rightarrow \]

\[ P(x \mid y) = \frac{P(y \mid x) \, P(x)}{P(y)} = \frac{\text{likelihood} \cdot \text{prior}}{\text{evidence}} \]
Normalization

\[ P(x \mid y) = \frac{P(y \mid x) \, P(x)}{P(y)} = \eta \, P(y \mid x) \, P(x) \]

\[ \eta = P(y)^{-1} = \frac{1}{\sum_x P(y \mid x) P(x)} \]

Algorithm:

\[ \forall x : \text{aux}_{x|y} = P(y \mid x) \, P(x) \]

\[ \eta = \frac{1}{\sum_x \text{aux}_{x|y}} \]

\[ \forall x : P(x \mid y) = \eta \, \text{aux}_{x|y} \]

Conditioning

- Total probability:

\[ P(x \mid y) = \int P(x \mid y, z) \, P(z \mid y) \, dz \]

- Bayes rule and background knowledge:

\[ P(x \mid y, z) = \frac{P(y \mid x, z) \, P(x \mid z)}{P(y \mid z)} \]
**Simple Example of State Estimation**

- Suppose a robot obtains measurement $z$.
- What is $P(open|z)$?

---

**Causal vs. Diagnostic Reasoning**

- $P(open|z)$ is **diagnostic**.
- $P(z|open)$ is **causal**.
- Often causal knowledge is easier to obtain.
- Count frequencies!
- Bayes rule allows us to use causal knowledge:

$$P(open|z) = \frac{P(z|open)P(open)}{P(z)}$$
Example

- \( P(z|\text{open}) = 0.6 \) \( P(z|\neg\text{open}) = 0.3 \)
- \( P(\text{open}) = P(\neg\text{open}) = 0.5 \)

\[
P(\text{open} | z) = \frac{P(z|\text{open})P(\text{open})}{P(z|\text{open})P(\text{open}) + P(z|\neg\text{open})P(\neg\text{open})}
\]

\[
P(\text{open} | z) = \frac{0.6 \cdot 0.5}{0.6 \cdot 0.5 + 0.3 \cdot 0.5} = \frac{2}{3} = 0.67
\]

- \( z \) raises the probability that the door is open.

Combining Evidence

- Suppose our robot obtains another observation \( z_2 \).

- How can we integrate this new information?

- More generally, how can we estimate \( P(x|z_1...z_n) \)?
Recursive Bayesian Updating

\[ P(x \mid z_1, \ldots, z_n) = \frac{P(z_n \mid x, z_1, \ldots, z_{n-1}) P(x \mid z_1, \ldots, z_{n-1})}{P(z_n \mid z_1, \ldots, z_{n-1})} \]

**Markov assumption**: \( z_n \) is independent of \( z_1, \ldots, z_{n-1} \) if we know \( x \).

\[ P(x \mid z_1, \ldots, z_n) = \frac{P(z_n \mid x) P(x \mid z_1, \ldots, z_{n-1})}{P(z_n \mid z_1, \ldots, z_{n-1})} = \eta \frac{P(z_n \mid x) P(x \mid z_1, \ldots, z_{n-1})}{\prod_{i=1}^{n} P(z_i \mid x) P(x)} \]

---

**Example: Second Measurement**

- \( P(z_2 \mid \text{open}) = 0.5 \quad P(z_2 \mid \neg\text{open}) = 0.6 \)
- \( P(\text{open} \mid z_1) = \frac{2}{3} \)

\[ P(\text{open} \mid z_2, z_1) = \frac{P(z_2 \mid \text{open}) P(\text{open} \mid z_1)}{P(z_2 \mid \text{open}) P(\text{open} \mid z_1) + P(z_2 \mid \neg\text{open}) P(\neg\text{open} \mid z_1)} \]

\[ = \frac{\frac{1}{2} \cdot \frac{2}{3}}{\frac{1}{2} \cdot \frac{2}{3} + \frac{3}{5} \cdot \frac{1}{3}} = \frac{5}{8} = 0.625 \]

- \( z_2 \) lowers the probability that the door is open.
**A Typical Pitfall**

- Two possible locations \(x_1\) and \(x_2\)
- \(P(x_1) = 0.99\)
- \(P(z|x_2) = 0.09\) \(P(z|x_1) = 0.07\)

![Graph showing probability distribution for two locations]

**Actions**

- Often the world is *dynamic* since
  - actions carried out by the robot,
  - actions carried out by other agents,
  - or just the time passing by
- How can we incorporate such actions?
**Typical Actions**

- The robot turns its wheels to move
- The robot uses its manipulator to grasp an object
- Plants grow over time...

- Actions are never carried out with absolute certainty.
- In contrast to measurements, actions generally increase the uncertainty.

**Modeling Actions**

- To incorporate the outcome of an action $u$ into the current “belief”, we use the conditional pdf

$$P(x|u,x')$$

- This term specifies the pdf that executing $u$ changes the state from $x'$ to $x$. 
Example: Closing the door

State Transitions

$P(x|u,x')$ for $u = \text{“close door”}$:

If the door is open, the action “close door” succeeds in 90% of all cases.
Integrating the Outcome of Actions

Continuous case:

\[ P(x \mid u) = \int P(x \mid u, x') P(x') dx' \]

Discrete case:

\[ P(x \mid u) = \sum P(x \mid u, x') P(x') \]

Example: The Resulting Belief

\[ P(\text{closed} \mid u) = \sum P(\text{closed} \mid u, x') P(x') \]
\[ = P(\text{closed} \mid u, \text{open}) P(\text{open}) \]
\[ + P(\text{closed} \mid u, \text{closed}) P(\text{closed}) \]
\[ = \frac{9}{10} * \frac{5}{8} + \frac{1}{8} * \frac{3}{16} = \frac{15}{16} \]

\[ P(\text{open} \mid u) = \sum P(\text{open} \mid u, x') P(x') \]
\[ = P(\text{open} \mid u, \text{open}) P(\text{open}) \]
\[ + P(\text{open} \mid u, \text{closed}) P(\text{closed}) \]
\[ = \frac{1}{10} * \frac{5}{8} + \frac{0}{1} * \frac{3}{8} = \frac{1}{16} \]
\[ = 1 - P(\text{closed} \mid u) \]
Bayes Filters: Framework

- **Given:**
  - Stream of observations $z$ and action data $u$:
    \[ d_i = \{u_1, z_2, \ldots, u_{t-1}, z_t\} \]
  - Sensor model $P(z|x)$.
  - Action model $P(x|u,u')$.
  - Prior probability of the system state $P(x)$.

- **Wanted:**
  - Estimate of the state $X$ of a dynamical system.
  - The posterior of the state is also called **Belief**:
    \[ Bel(x_t) = P(x_t | u_1, z_2, \ldots, u_{t-1}, z_t) \]

Markov Assumption

- Static world
- Independent noise
- Perfect model, no approximation errors
Bayes Filters

$$Bel(x_i) = P(x_i \mid u_1, z_2, \ldots, u_{i-1}, z_i)$$

Bayes

$$= \eta \ P(z_i \mid x_i, u_1, z_2, \ldots, u_{i-1}) \ P(x_i \mid u_1, z_2, \ldots, u_{i-1})$$

Markov

$$= \eta \ P(z_i \mid x_i) \ P(x_i \mid u_1, z_2, \ldots, u_{i-1})$$

Total prob.

$$= \eta \ P(z_i \mid x_i) \ \int P(x_i \mid u_1, z_2, \ldots, u_{i-1}, x_{i-1}) \ P(x_{i-1} \mid u_1, z_2, \ldots, u_{i-1}) \ dx_{i-1}$$

Markov

$$= \eta \ P(z_i \mid x_i) \ \int P(x_i \mid u_{i-1}, x_{i-1}) \ P(x_{i-1} \mid u_1, z_2, \ldots, u_{i-1}) \ dx_{i-1}$$

$$= \eta \ P(z_i \mid x_i) \ \int P(x_i \mid u_{i-1}, x_{i-1}) \ Bel(x_{i-1}) \ dx_{i-1}$$

Bayes Filter Algorithm

1. Algorithm **Bayes_filter** (Bel(x), d):
   2. $\eta$ = 0
   3. if d is a perceptual data item z then
      4. For all x do
         5. $Bel'(x) = P(z \mid x)Bel(x)$
         6. $\eta = \eta + Bel'(x)$
      7. For all x do
         8. $Bel'(x) = \eta^{-1}Bel'(x)$
   9. else if d is an action data item u then
      10. For all x do
           11. $Bel'(x) = \int P(x \mid u, x') \ Bel(x') \ dx'$
      12. return Bel'(x)
Bayes Filters are Familiar!

\[
Bel(x_t) = \eta \int P(z_t | x_t) \ P(x_t | u_{t-1}, x_{t-1}) \ Bel(x_{t-1}) \ dx_{t-1}
\]

- Kalman filters
- Particle filters
- Hidden Markov models
- Dynamic Bayes networks
- Partially Observable Markov Decision Processes (POMDPs)

Application

- Estimate the opening angle of a door
- and the state of other dynamic objects
- using a laser-range finder
- from a moving mobile robot and
- based on Bayes filters.

[Schulz et al., 01]
Lessons Learned

- Bayes rule allows us to compute probabilities that are hard to assess otherwise.
- Under the Markov assumption, recursive Bayesian updating can be used to efficiently combine evidence.
- Bayes filters are a probabilistic tool for estimating the state of dynamic systems.