Probabilistic Robotics

Slides from Autonomous Robots (Siegwart and Nourbaksh), Chapter 5
Probabilistic Robotics (S. Thurn et al.)
Today

- Overview of probability, Representing uncertainty
- Propagation of uncertainty, Bayes Rule
- Application to Localization and Mapping
Probabilistic Robotics

Key idea:
Explicit representation of uncertainty using the calculus of probability theory

- Perception = state estimation
- Action = utility optimization
Pr(A) denotes probability that proposition A is true

Axioms of Probability Theory

\[ 0 \leq \Pr(A) \leq 1 \]

\[ \Pr(True) = 1 \quad \Pr(False) = 0 \]

\[ \Pr(A \lor B) = \Pr(A) + \Pr(B) - \Pr(A \land B) \]
A Closer Look at Axiom 3

\[ \Pr(A \lor B) = \Pr(A) + \Pr(B) - \Pr(A \land B) \]
Syntax

- Basic element: random variable
- Similar to propositional logic: possible worlds defined by assignment of values to random variables.
- Boolean random variables
  - e.g., Cavity (do I have a cavity?)

- Discrete random variables
  - e.g., Weather is one of <sunny, rainy, cloudy, snow>

- Domain values must be exhaustive and mutually exclusive

- Elementary proposition constructed by assignment of a value to a random variable:
  - e.g., Weather = sunny, Cavity = false (abbreviated as ¬cavity)

- Complex propositions formed from elementary propositions and standard logical connectives e.g., Weather = sunny ∨ Cavity = false
Syntax

- **Atomic event:** A complete specification of the state of the world about which the agent is uncertain

  E.g., if the world consists of only two Boolean variables Cavity and Toothache, then there are 4 distinct atomic events:

  - Cavity = false ∧ Toothache = false
  - Cavity = false ∧ Toothache = true
  - Cavity = true ∧ Toothache = false
  - Cavity = true ∧ Toothache = true

- Atomic events are mutually exclusive and exhaustive
Prior probability

- Prior or unconditional probabilities of propositions
  \[ P(\text{Cavity} = \text{true}) = 0.1 \text{ and } P(\text{Weather} = \text{sunny}) = 0.72 \]
  correspond to belief prior to arrival of any (new) evidence.

- Probability distribution values for all possible assignments:
  \[ P(\text{Weather}) = <0.72,0.1,0.08,0.1> \text{ (normalized, i.e., sums to 1)} \]

- Joint probability distribution for a set of random variables gives the probability of every atomic event on those random variables:
  \[ P(\text{Weather,Cavity}) = \begin{bmatrix} 0.144 & 0.02 & 0.016 & 0.02 \\ 0.576 & 0.08 & 0.064 & 0.08 \end{bmatrix} \]
Joint Distribution

- Every question about the domain can be answered from joint probability distribution

- Two examples of joint distributions

\[
\begin{array}{c|cccc}
\text{Weather} = & \text{sunny} & \text{rainy} & \text{cloudy} & \text{snow} \\
\hline
\text{Cavity} = \text{true} & 0.144 & 0.02 & 0.016 & 0.02 \\
\text{Cavity} = \text{false} & 0.576 & 0.08 & 0.064 & 0.08 \\
\end{array}
\]
Conditional probability

- **Definition of conditional probability:**
  
  \[
P(a \mid b) = \frac{P(a \land b)}{P(b)}
  \]

- **Product rule** gives an alternative formulation:
  
  \[
P(a \land b) = P(a \mid b)P(b) = P(b \mid a)P(a)
  \]

- A general version holds for whole distributions, e.g.,
  \[
P(\text{Weather, Cavity}) = P(\text{Weather} \mid \text{Cavity}) \cdot P(\text{Cavity})
  \]

- View as a set of 4 × 2 equations, **not** matrix multiplication

- **Chain rule** is derived by successive product rule
  
  \[
P(X_1, \ldots, X_n) = P(X_1, \ldots, X_{n-1}) \cdot P(X_n \mid X_1, \ldots, X_{n-1})
  \]
  
  \[
  = P(X_1, \ldots, X_{n-2}) \cdot P(X_{n-1} \mid X_1, \ldots, X_{n-2}) \cdot P(X_n \mid X_1, \ldots, X_{n-1}) = \ldots
  \]
  
  \[
  = \prod_{i=1}^{n} P(X_i \mid X_1, \ldots, X_{i-1})
  \]
Inference by enumeration

- Start with the joint probability distribution:

<table>
<thead>
<tr>
<th></th>
<th>toothache</th>
<th>¬ toothache</th>
</tr>
</thead>
<tbody>
<tr>
<td>catch</td>
<td>0.108</td>
<td>0.012</td>
</tr>
<tr>
<td>¬ catch</td>
<td>0.072</td>
<td>0.008</td>
</tr>
<tr>
<td>cavity</td>
<td>0.016</td>
<td>0.064</td>
</tr>
<tr>
<td>¬ cavity</td>
<td>0.144</td>
<td>0.576</td>
</tr>
</tbody>
</table>

- Given a proposition, calculate what is its probability

- For any proposition \( \varphi \), sum the atomic events where it is true: \( P(\varphi) = \sum_{\omega:\omega \models \varphi} P(\omega) \)
Inference by enumeration

- Start with the joint probability distribution:

<table>
<thead>
<tr>
<th></th>
<th>toothache</th>
<th>¬ toothache</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>catch</td>
<td>¬ catch</td>
</tr>
<tr>
<td>cavity</td>
<td>.108</td>
<td>.012</td>
</tr>
<tr>
<td>¬ cavity</td>
<td>.016</td>
<td>.064</td>
</tr>
</tbody>
</table>

- For any proposition \( \varphi \), sum the atomic events where it is true: \( P(\varphi) = \sum_{\omega: \omega \models \varphi} P(\omega) \)

- \( P(\text{toothache}) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2 \)
Inference by enumeration

- Start with the joint probability distribution:

<table>
<thead>
<tr>
<th></th>
<th>toothache</th>
<th>¬ toothache</th>
</tr>
</thead>
<tbody>
<tr>
<td>catch</td>
<td>.108</td>
<td>.072</td>
</tr>
<tr>
<td>¬ catch</td>
<td>.012</td>
<td>.008</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>toothache</th>
<th>¬ toothache</th>
</tr>
</thead>
<tbody>
<tr>
<td>cavity</td>
<td>.016</td>
<td>.144</td>
</tr>
<tr>
<td>¬ cavity</td>
<td>.064</td>
<td>.576</td>
</tr>
</tbody>
</table>

- Can also compute conditional probabilities:

\[
P(¬cavity \mid toothache) = \frac{P(¬cavity \land toothache)}{P(toothache)} = \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4
\]
Discrete Random Variables

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

E.g. Recognize which room you (robot) is in

- This is just shorthand for $P(Room = \text{office}), P(Room = \text{kitchen}), P(Room = \text{bedroom}), P(Room = \text{corridor})$

$$P(Room) = \langle 0.7, 0.2, 0.08, 0.02 \rangle$$
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
  \[ P(x,y) = P(x) \cdot P(y) \]

- $P(x / y)$ is the probability of $x$ given $y$
  \[ P(x / y) = P(x,y) / P(y) \]
  \[ P(x,y) = P(x / y) \cdot P(y) \]

- If $X$ and $Y$ are independent then
  \[ P(x / y) = P(x) \]

(verify using definitions of conditional probability and independence)
Law of Total Probability, Marginals

Discrete case

Law of total probability

\[ \sum_x P(x) = 1 \]

Marginalization

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

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

Continuous case

\[ \int p(x) \, dx = 1 \]

\[ p(x) = \int p(x, y) \, dy \]

\[ p(x) = \int p(x \mid y) p(y) \, dy \]
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

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

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

Algorithm:

\[ \forall x : aux_{x \mid y} = P(y \mid x) \ P(x) \]

\[ \eta = \frac{1}{\sum_x aux_{x \mid y}} \]

\[ \forall x : P(x \mid y) = \eta \ aux_{x \mid y} \]
Bayes' Rule

- Product rule: \( P(a \land b) = P(a \mid b)P(b) = P(b \mid a)P(a) \)

\[ \Rightarrow \text{Bayes' rule:} \quad P(a \mid b) = \frac{P(b \mid a)P(a)}{P(b)} \]

- or in distribution form

\[ P(Y \mid X) = \frac{P(X \mid Y)P(Y)}{P(X)} \]

- Useful for assessing diagnostic probability from causal probability:

\[ P(\text{Cause} \mid \text{Effect}) = P(\text{Effect} \mid \text{Cause}) \cdot P(\text{Cause}) / P(\text{Effect}) \]

Note: posterior probability of meningitis still very small!
Bayes' Rule

- Baye’s rule

\[ P(Y \mid X) = \frac{P(X \mid Y)P(Y)}{P(X)} \]

- More general version conditionalized on some evidence

\[ P(Y \mid X, e) = \frac{P(X \mid Y, e)P(Y \mid e)}{P(X \mid e)} \]

- E.g., let \( M \) be meningitis, \( S \) be stiff neck:

\[ P(m \mid s) = \frac{P(s \mid m)P(m)}{P(s)} = \frac{0.8 \times 0.001}{0.1} = 0.008 \]

- Normalization same for \( m \) and \( \sim m \)

\[ P(Y \mid X) = \alpha P(X \mid Y)P(Y) \]
Bayes Rule with Background Knowledge

\[ P(x \mid y, z) = \frac{P(y \mid x, z) P(x \mid z)}{P(y \mid z)} \]
Conditional Independence

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

- Equivalent to
  \[ P(x \mid z) = P(x \mid z, y) \]
  and
  \[ P(y \mid z) = P(y \mid z, x) \]
- But this does not necessarily mean
  \[ P(x, y) = P(x)P(y) \]
  (independence/marginal independence)
Simple Example of State Estimation

- Suppose a robot obtains measurement $z$
- What is $P(\text{open}/z)$?
Causal vs. Diagnostic Reasoning

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

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

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

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

\[
P(\text{open} | z) = \frac{0.6 \cdot 0.5}{0.6 \cdot 0.5 + 0.3 \cdot 0.5} = \frac{0.3}{0.3 + 0.15} = 0.67
\]

- \( z \) 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 $P(x/ z_1...z_n)$?
Example: Second Measurement

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

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

\[
= \frac{\frac{1}{2} \cdot \frac{2}{3}}{\frac{1}{2} \cdot \frac{2}{3} + \frac{1}{5} \cdot \frac{3}{15}} = \frac{\frac{1}{3}}{\frac{1}{3}} + \frac{\frac{1}{5}}{\frac{3}{5}} = \frac{\frac{5}{8}}{\frac{8}{8}} = 0.625
\]

\( z_2 \) lowers the probability that the door is open
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 \(x\)

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

\[ P(\text{closed} \mid u) = \sum P(\text{closed} \mid u, x')P(x') \]
\[ = P(\text{closed} \mid u, \text{open})P(\text{open}) \]
\[ + P(\text{closed} \mid u, \text{closed})P(\text{closed}) \]
\[ = \frac{9}{10} \times \frac{5}{8} + \frac{1}{18} \times \frac{3}{8} = \frac{15}{16} \]

\[ P(\text{open} \mid u) = \sum P(\text{open} \mid u, x')P(x') \]
\[ = P(\text{open} \mid u, \text{open})P(\text{open}) \]
\[ + P(\text{open} \mid u, \text{closed})P(\text{closed}) \]
\[ = \frac{1}{10} \times \frac{5}{8} + \frac{0}{18} \times \frac{3}{8} = \frac{1}{16} \]
\[ = 1 - P(\text{closed} \mid u) \]
Bayes Filters: Framework

- **Given:**
  - Stream of observations $z$ and action data $u$: $d_t = \{u_1, z_1 \ldots, u_t, z_t\}$
  - Sensor model $P(z|x)$
  - Action model $P(x|u,x')$
  - 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**:

\[
Bel(x_t) = P(x_t \mid u_1, z_1 \ldots, u_t, z_t)
\]
Markov Assumption

Underlying Assumptions

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

Mathematical expressions:

\[ p(z_t \mid x_{0:t}, z_{1:t-1}, u_{1:t}) = p(z_t \mid x_t) \]

\[ p(x_t \mid x_{1:t-1}, z_{1:t}, u_{1:t}) = p(x_t \mid x_{t-1}, u_t) \]
Bayes Filters

\[ \text{Bel}(x_t) = P(x_t \mid u_1, z_1 \ldots, u_t, z_t) \]

Bayes

\[ = \eta \ P(z_t \mid x_t, u_1, z_1, \ldots, u_t) \ P(x_t \mid u_1, z_1, \ldots, u_t) \]

Markov

\[ = \eta \ P(z_t \mid x_t) \ P(x_t \mid u_1, z_1, \ldots, u_t) \]

Total prob.

\[ = \eta \ P(z_t \mid x_t) \int P(x_t \mid u_1, z_1, \ldots, u_t, x_{t-1}) \] 
\[ \ P(x_{t-1} \mid u_1, z_1, \ldots, u_t) \, dx_{t-1} \]

Markov

\[ = \eta \ P(z_t \mid x_t) \int P(x_t \mid u_t, x_{t-1}) \ P(x_{t-1} \mid u_1, z_1, \ldots, u_t) \, dx_{t-1} \]

Markov

\[ = \eta P(z_t \mid x_t) \int P(x_t \mid u_t, x_{t-1}) P(x_{t-1} \mid u_1, z_1, \ldots, z_{t-1}) \, dx_{t-1} \]

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

\[ z = \text{observation} \]
\[ u = \text{action} \]
\[ x = \text{state} \]
Algorithm **Bayes_filter** \((\text{Bel}(x), d)\):

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

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

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

- 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.
Belief Representation

- a) Continuous map with *single hypothesis*

- b) Continuous map with *multiple hypothesis*

- d) Discretized map with probability distribution

- d) Discretized topological map with probability distribution
Belief Representation: Characteristics

- **Continuous**
  - Precision bound by sensor data
  - Typically single hypothesis pose estimate
  - Lost when diverging (for single hypothesis)
  - Compact representation and typically reasonable in processing power.

- **Discrete**
  - Precision bound by resolution of discretisation
  - Typically multiple hypothesis pose estimate
  - Never lost (when diverges converges to another cell)
  - Important memory and processing power needed. (not the case for topological maps)
Bayesian Approach: A taxonomy of models

More general

Bayesian Programs

Bayesian Networks

DBNs

Markov Loc

MDPs

MCML

POMDPs

More specific

Bayesian Filters

semi-cont. HMMs

continuous HMMs

Kalman Filters

S\_t S\_t-1 A\_t

S\_t S\_t-1 O\_t A\_t

S\_t S\_t-1 A\_t

© R. Siegwart, I. Nourbakhsh

S: State
O: Observation
A: Action

Courtesy of Julien Diard

S\_t: State
O\_t: Observation
A\_t: Action

© R. Siegwart, I. Nourbakhsh