Probability: Review

Pieter Abbeel
UC Berkeley EECS

Many slides adapted from Thrun, Burgard and Fox, Probabilistic Robotics
Why probability in robotics?

- Often state of robot and state of its environment are unknown and only noisy sensors available
  - Probability provides a framework to fuse sensory information
    → Result: probability distribution over possible states of robot and environment

- Dynamics is often stochastic, hence can’t optimize for a particular outcome, but only optimize to obtain a good distribution over outcomes
  - Probability provides a framework to reason in this setting
    → Ability to find good control policies for stochastic dynamics and environments
Example 1: Helicopter

- State: position, orientation, velocity, angular rate

- Sensors:
  - GPS: noisy estimate of position (sometimes also velocity)
  - Inertial sensing unit: noisy measurements from
    - 3-axis gyro [=angular rate sensor],
    - 3-axis accelerometer [=measures acceleration + gravity; e.g., measures (0,0,0) in free-fall],
    - 3-axis magnetometer

- Dynamics:
  - Noise from: wind, unmodeled dynamics in engine, servos, blades
Example 2: Mobile robot inside building

- **State:** position and heading

- **Sensors:**
  - Odometry (=sensing motion of actuators): e.g., wheel encoders
  - Laser range finder:
    - Measures time of flight of a laser beam between departure and return
    - Return is typically happening when hitting a surface that reflects the beam back to where it came from

- **Dynamics:**
  - Noise from: wheel slippage, unmodeled variation in floor
Axioms of Probability Theory

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

\[ \Pr(\Omega) = 1 \quad \Pr(\phi) = 0 \]

\[ \Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B) \]

\( \Pr(A) \) denotes probability that the outcome \( \omega \) is an element of the set of possible outcomes \( A \). \( A \) is often called an event. Same for \( B \). 
\( \Omega \) is the set of all possible outcomes. 
\( \phi \) is the empty set.
A Closer Look at Axiom 3

\[ \Pr(\Omega) = \Pr(A) + \Pr(B) - \Pr(A \cap B) \]
Using the Axioms

\[
\begin{aligned}
\Pr(A \cup (\Omega \setminus A)) &= \Pr(A) + \Pr(\Omega \setminus A) - \Pr(A \cap (\Omega \setminus A)) \\
\Pr(\Omega) &= \Pr(A) + \Pr(\Omega \setminus A) - \Pr(\phi) \\
1 &= \Pr(A) + \Pr(\Omega \setminus A) - 0 \\
\Pr(\Omega \setminus A) &= 1 - \Pr(A)
\end{aligned}
\]
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 $x_i$.
- $P(\cdot)$ is called probability mass function.
- *E.g., $X$ models the outcome of a coin flip, $x_1 = \text{head}, x_2 = \text{tail}, P(x_1) = 0.5, P(x_2) = 0.5$*
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
  \[
P(x \mid y) = P(x)
  \]

- Same for probability densities, just \( P \rightarrow p \)
# 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

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

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

Algorithm:

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

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

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

- Law of total probability:

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

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

\[ P(x \mid y) = \int P(x \mid y, z)P(z \mid y)dz \]
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 \[ P(y \mid z) = P(y \mid z, x) \]
Simple Example of State Estimation

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

- $P(\text{open} | z)$ is diagnostic.
- $P(z | \text{open})$ is causal. \(\text{count frequencies!}\)
- Often causal knowledge is easier to obtain.
- Bayes rule allows us to use causal knowledge:

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

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

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

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

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

- \( z \) 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 \( x \).

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

- $P(z_2|\text{open}) = 0.5$  \hspace{1cm}  $P(z_2|\neg\text{open}) = 0.6$

- $P(\text{open}|z_1) = \frac{2}{3}$

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

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

- $z_2$ lowers the probability that the door is open.
A Typical Pitfall

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