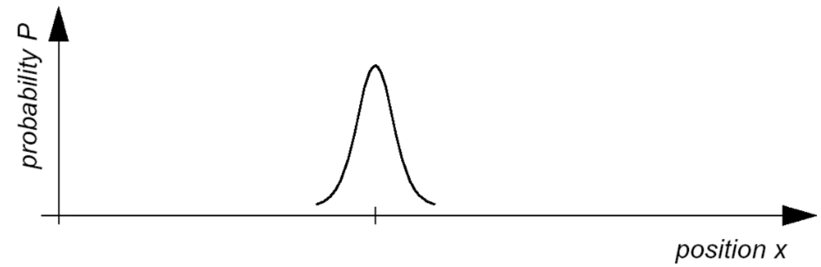
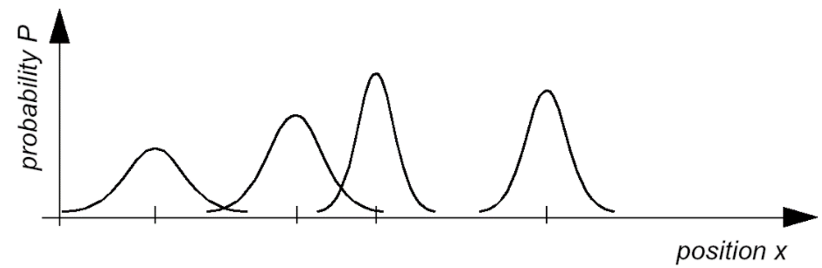
Probability Theory Review

Belief representation: how do we represent our belief (hypothesis) of where the robot is located?

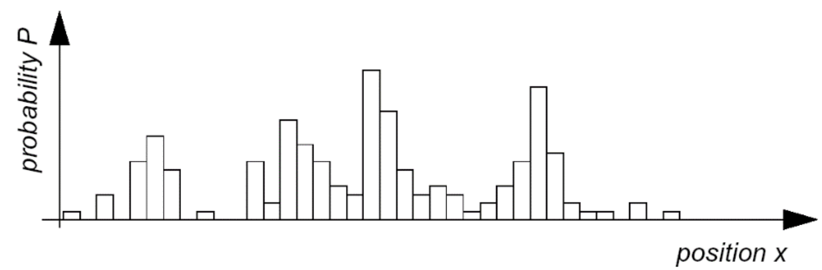
Continuous map with single hypothesis probability distribution



Continuous map with multiple hypotheses probability distribution



Discretized map with multiple hypotheses probability distribution



Discretized topological map with with multiple hypotheses probability distribution

Belief representation

- **Single-hypothesis belief:** The robot's belief about its position is expressed as a single point on a map
  - Advantage: no ambiguity, simplifies planning and decision making
  - Disadvantage: does not represent ambiguity/uncertainty

- **Multi-hypothesis belief**: allows the robot to track (possibly infinitely) many possible positions.

In both of the above, the beliefs are represented as *probabilities*

Discrete Random Variables

- X denotes a 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 probability that the random variable X takes on value x_i.
- $P(x_i)$ is called probability mass function.
- E.g.

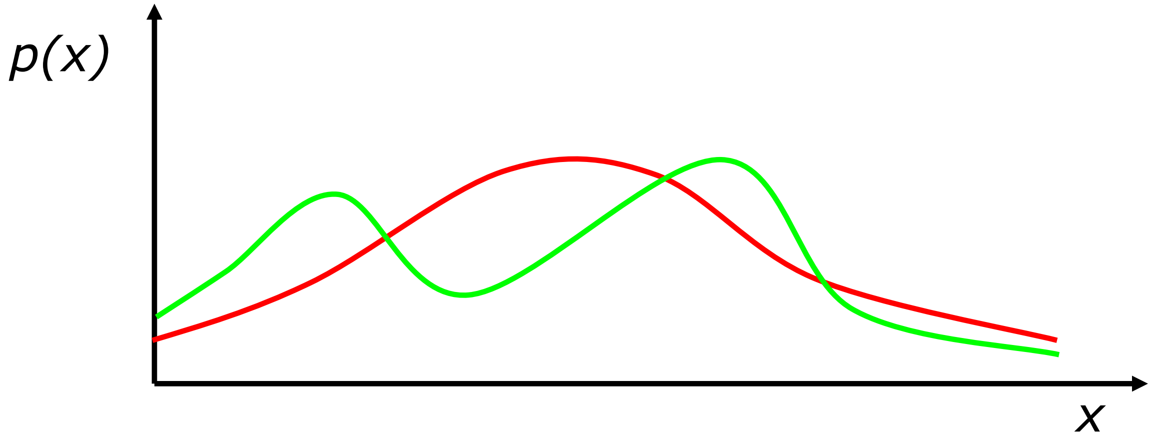
$$P(Raining) = 0.2$$

Continuous Random Variables

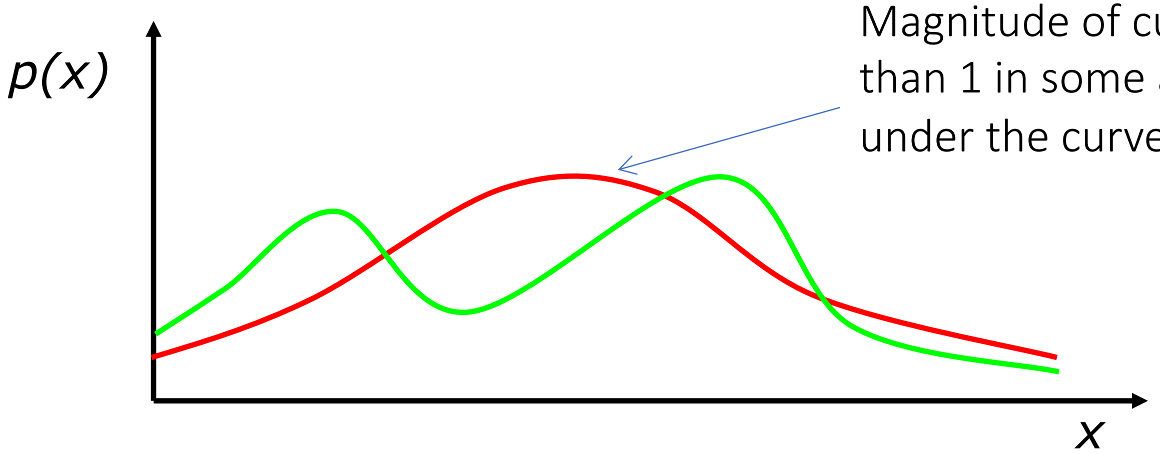
- X takes on values in the continuum.
- $p(X = x)$, or $p(x)$, is a probability density function.

$$P(x \in (a,b)) = \int_a^b p(x)dx$$

- E.g.



Probability Density Function



Magnitude of curve could be greater than 1 in some areas. The total area under the curve must add up to 1.

Since continuous probability functions are defined for an infinite number of points over a continuous interval, the probability at a single point is always 0.

Joint Probability

- *Notation*
  - $P(X = x \; and \; Y = y) \; = \; P(x, y)$

- If X and Y are independent then

$$P(x, y) \; = \; P(x) \, P(y)$$

Conditional Probability

- $P(x \mid y)$ is the probability of *x* given *y*

$$P(x \mid y) = \frac{P(x,y)}{P(y)}$$

$$\begin{aligned} P(x,y) &= P(x \mid y)\, P(y) \\ &= P(y \mid x)\, P(x) \end{aligned}$$

- If X and Y are independent then

$$P(x \mid y) = P(x)$$

An Example

Roll two dice, observe x_1 and x_2.

We know that there are 36 possible outcomes, all of which are equally likely (assuming the dice are *fair*).

It's easy to compute probabilities by simply counting outcomes:

- Probability $x_1 = 6$:

$$(6,1), (6,2), (6,3), (6,4), (6,5), (6,6) \rightarrow P = \frac{6}{36} = \frac{1}{6}$$

- Probability $x_1 = 6$ ***and*** x_2 is even:

$$(6,2), (6,4), (6,6) \rightarrow P = \frac{3}{36} = \frac{1}{12}$$

- Probability x_1 is even:

$$\begin{matrix} (2,1), (2,2), (2,3), (2,4), (2,5), (2,6) \\ (4,1), (4,2), (4,3), (4,4), (4,5), (4,6) \\ (6,1), (6,2), (6,3), (6,4), (6,5), (6,6) \end{matrix} \rightarrow P = \frac{18}{36} = \frac{1}{2}$$

Let's apply rules of conditional and joint probabilities:

Define events: A: x_1 is even; B: $\mathrm{x}_1 = 6$; C: $\mathrm{x}_2$is even; D: $\mathrm{x}_2 = 5$

From the previous page, we easily compute the following:

$$P(A) = \frac{1}{2}, \qquad P(B) = \frac{1}{6}, \qquad P(C) = \frac{1}{2}, \qquad P(D) = \frac{1}{6}.$$

Let's look at some combinations of events:

- $P(A, B) = \frac{1}{6} \neq P(A)P(B) = \frac{1}{6} \times \frac{1}{2} = \frac{1}{12} \rightarrow$ ***NOT independent***
- $P(A, C) = \frac{9}{36} = P(A)P(C) = \frac{1}{2} \times \frac{1}{2} \rightarrow$ ***independent***
- $P(B|A) = \frac{P(A,B)}{P(A)} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}$

This agrees with our intuition, since $\mathrm{x}_1 = 6$ in one third of the cases of x_1 being even:

(2,1), (2,2), (2,3), (2,4), (2,5), (2,6)
(4,1), (4,2), (4,3), (4,4), (4,5), (4,6)
(6,1), (6,2), (6,3), (6,4), (6,5), (6,6)

Law of Total Probability

Discrete case

$$\sum_x P(x) = 1$$

$$P(x) = \sum_y P(x, y)$$

$$P(x) = \sum_y P(x \mid y) P(y)$$

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 Theorem

We know that conjunction is commutative:

$$P(A,B) = P(B,A)$$

Using the definition of conditional probability:

$$P(B|A)P(A) = P(B,A) = P(A,B) = P(A|B)P(B)$$

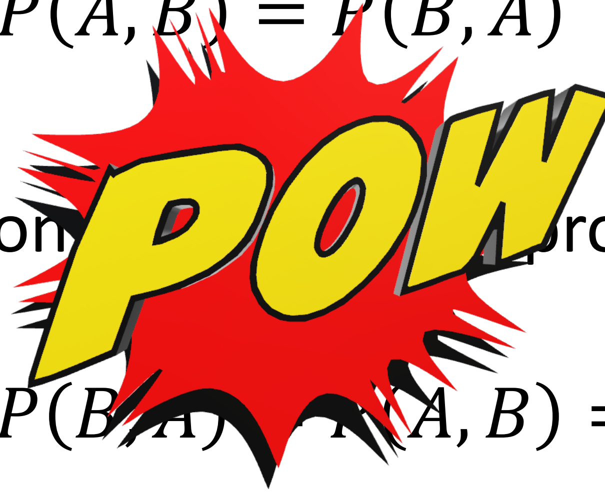
$$P(B|A)P(A) = P(A|B)P(B)$$

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

Bayes Theorem

We know that conjunction is commutative:

$$P(A,B) = P(B,A)$$



Using the definition … probability:

$$P(B|A)P(A) = P(B, \ldots) \ldots (A,B) = P \ldots$$

$$\ldots|A)P(A) = P(A|B)P(B)$$

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

Example

We roll one die, and an observer tells us things about the outcome.
We want to know if $X = 4$.

- Before we know anything, we believe $P(X = 4) = \frac{1}{6}$. PRIOR
- Now, suppose the observer tells us that X is even. EVIDENCE

$$P(X = 4 \mid X \text{ even}) = \frac{P(X{=}4, X \text{ even})}{P(X \text{ even})} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3} \quad \text{BAYES}$$

- We could also use Bayes to infer $P(X = \text{even} \mid X = 4)$:

$$P(X \text{ even} \mid X = 4) = \frac{P(X = 4, X \text{ even})}{P(X = 4)} = \frac{\frac{1}{6}}{\frac{1}{6}} = 1$$

Bayes Rule

$$P(x|y) = \frac{P(y \mid x)\ P(x)}{P(y)} = \frac{\text{likelihood} \cdot \text{prior}}{\text{evidence}}$$

x is robot pose and y is sensor data

$p(x)$ → *Prior* probability distribution

$p(x|y)$ → *Posterior* (conditional) probability distribution

$p(y|x)$ → *Likelihood*, model of the characteristics of the event

$p(y)$ → *Evidence*, does not depend on x

Bayes Rule

$$\Rightarrow \quad P(x|y) = \frac{P(y|x)\ P(x)}{P(y)} = \frac{\text{likelihood} \cdot \text{prior}}{\text{evidence}}$$

$$\Rightarrow \quad P(x|y) = \frac{P(y|x)\ P(x)}{\sum_x P(y|x)P(x)}$$

About likelihoods…

Why do we call the conditional probability $p(y|x)$ a *likelihood*, but we call $p(x|y)$ the *posterior*??

We define the likelihood $\mathcal{L}(x)$ to be a function of x, ***not*** a function of y :

$$\mathcal{L}(x) = p(y|x)$$

Note: $\mathcal{L}(x)$ is ***not*** a probability. In particular,

$$\sum_x \mathcal{L}(x) \neq 1$$

Normalization Coefficient

$$P(x|z) = \frac{P(z|x)P(x)}{P(z)}$$

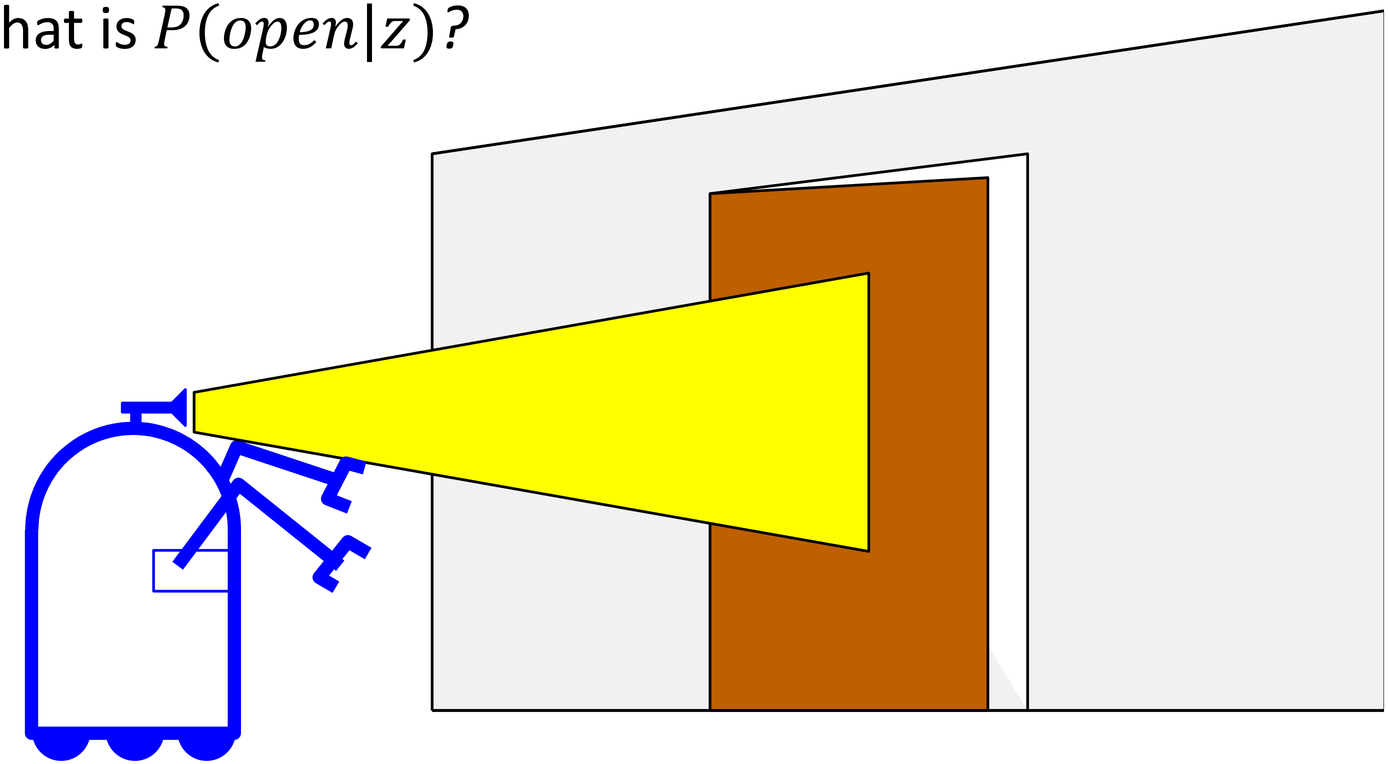
Note that the denominator is independent of x, and as a result will typically be the same for any value of x in the posterior $P(x|z)$.

Therefore, we typically represent the normalization term by the coefficient $\eta = [P(z)]^{-1}$ and Bayes equation is written as

$$P(x|z) = \eta P(z|x)P(x)$$

Simple Example of State Estimation

- Suppose a robot obtains measurement z *(e.g., distance sensor reports an obstacle 40cm in front of the robot)*
- 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.
- Bayes rule allows us to use causal knowledge:

Comes from sensor model.

$$P(open \mid z) = \frac{P(z \mid open)P(open)}{P(z)}$$

Use law of total probability: $P(z) = \sum_y P(z|y)P(y)$

Example

$$P(open \mid z) = \frac{P(z \mid open)P(open)}{P(z)}$$

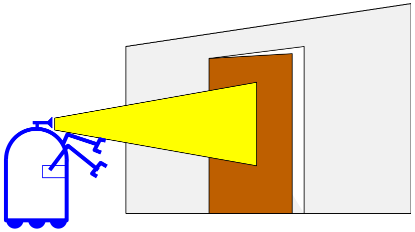
- $P(z|open) = 0.6$ $\qquad$ $P(z|\neg open) = 0.3$
- $P(open) = P(\neg open) = 0.5$

$$P(open \mid z) = \frac{P(z \mid open)P(open)}{P(z \mid open)\,p(open) + P(z \mid \neg open)\,p(\neg open)}$$

$$P(open \mid z) = \frac{0.6 \cdot 0.5}{0.6 \cdot 0.5 + 0.3 \cdot 0.5} = \frac{2}{3} = 0.67$$

z raises the probability that the door is open.

Lets try the measurement again…



$$P(x|z) = \eta\, P(z|x)P(x)$$

Given information:

$P(z_1|open) = 0.6$
$P(z_1|closed) = 0.3$
$P(open) = 0.5$
$P(closed) = 0.5$

$$P(open|z_1) = \eta\, P(z_1|open)P(open)$$

$$P(open|z_1) = \eta\, 0.6 * 0.5 = \eta\, 0.3$$

Unlike before, we don't yet have the answer because we still have the unknown term η that indicates that we need to normalize to get the true probability.

$$P(closed|z_1) = \eta\, P(z_1|closed)P(closed)$$

$$P(closed|z_1) = \eta\, 0.3 * 0.5 = \eta\, 0.15$$

$$\eta = (0.3 + 0.15)^{-1} = 2.22$$

$$\mathbf{P(open|z_1) = 0.67}$$

Combining Evidence

- Suppose our robot obtains another observation z_2. *e.g. we made a second sensor reading with the same sensor, and it reports an obstacle 35cm away*

- How can we integrate this new information?

- More generally, how can we estimate $P(x| z_1 \ldots z_n)$?

Generalizing the Condition with Bayes Theorem

The usual version of Bayes is conditioned on a single event:

$$P(x|z) = \frac{P(z|x)P(x)}{P(z)}$$

In fact, we can add any arbitrary context variables on the right side of the conditioning bar, so long as we apply them in ***every*** term.

$$P(x|z, Anything) = \frac{P(z|x, Anything)P(x|Anything)}{P(z|Anything)}$$

Multiple Measurements

$$P(x|z, Anything) = \frac{P(z|x, Anything)P(x|Anything)}{P(z|Anything)}$$

$$P(x|z_2, Anything) = \frac{P(z_2|x, Anything)P(x|Anything)}{P(z_2|Anything)}$$

$$P(x|z_2, z_1) = \frac{P(z_2|x, z_1)P(x|z_1)}{P(z_2|z_1)}$$

At time $t = 2$, everything earlier is merely context information.

Multiple Measurements (cont)

$$P(x|z_2, z_1) = \frac{P(z_2|x, z_1)P(x|z_1)}{P(z_2|z_1)}$$

At time $t = 2$

- $P(x|z_1)$ is the prior… what we believe about the state x, based on history of measurements before $t = 2$
- $P(x|z_2, z_1)$ is the posterior… what we believe about the state x, based on history of measurements, including $t = 2$
- $P(z_2|x, z_1)$ … If we really know the state x, then what we measured at time $t = 1$ won't affect what we expect to measure at time $t = 2$

Reference

- **Probabilistic Robotics** by Thrun, Burgard and Fox.. Chapter 2 (available on Piazza)