Probabilistic Robotics

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

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

Probabilistic Robotics

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

- *Perception* = *state estimation*
- *Action* = *utility optimization*

Uncertainty

Let action A_t = leave for airport $_t$ minutes before flight
Will A_t get me there on time?
Problems:

1. *partial observability (road state, other drivers' plans, etc.)*
2. *noisy sensors (traffic reports)*
3. *uncertainty in action outcomes (flat tire, etc.)*
4. *immense complexity of modeling and predicting traffic*

Hence a purely logical approach either

1. *risks falsehood: "A_{25} will get me there on time", or*
2. *leads to conclusions that are too weak for decision making:*

"A_{25} will get me there on time if there's no accident on the bridge and it doesn't rain and my tires remain intact etc etc."

(A_{1440} might reasonably be said to get me there on time but I'd have to stay overnight in the airport …)

Methods for handling uncertainty

- Rules with fudge factors:
  - A_{25} $|\rightarrow_{0.3}$ *get there on time*
  - *Sprinkler* $|\rightarrow_{0.99}$ *WetGrass*
  - *WetGrass* $|\rightarrow_{0.7}$ *Rain*

- Issues: Problems with combination, e.g., *Sprinkler* causes *Rain*??

- Probability
  - *Model agent's degree of belief*
  - *Given the available evidence,*
  - A_{25} *will get me there on time with probability 0.04*

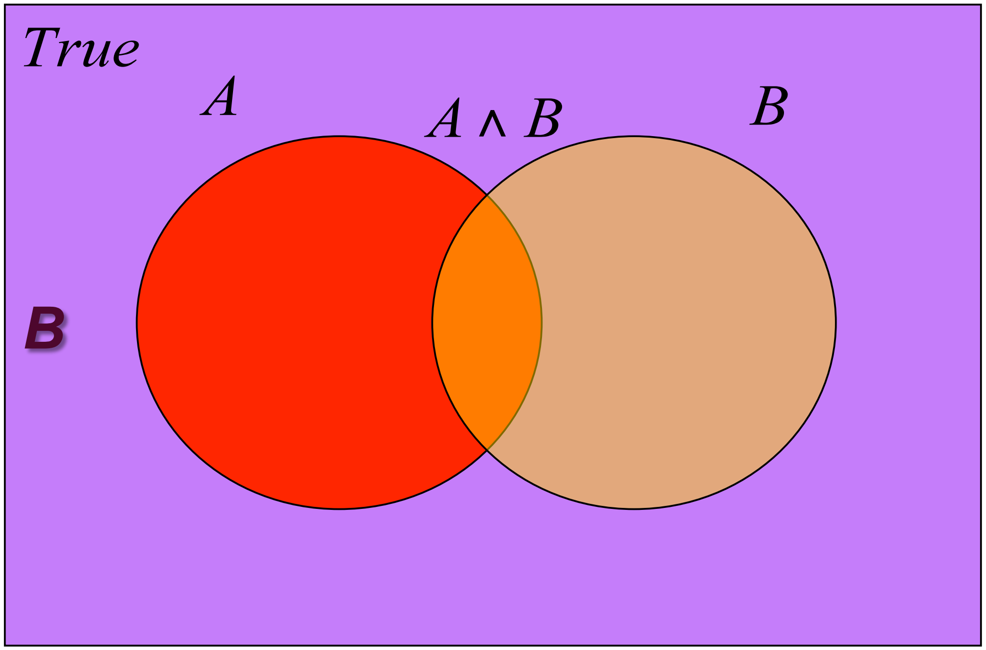
Axioms of Probability Theory

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

- $$0 \le \Pr(A) \le 1$$
- $$\Pr(True) = 1 \qquad \Pr(False) = 0$$
- $$\Pr(A \vee B) = \Pr(A) + \Pr(B) - \Pr(A \wedge B)$$

A Closer Look at Axiom 3

$$\Pr(A \vee B) = \Pr(A) + \Pr(B) - \Pr(A \wedge B)$$



Using the Axioms

$$
\begin{aligned}
\Pr(A \vee \neg A) &= \Pr(A) + \Pr(\neg A) - \Pr(A \wedge \neg A) \\
\Pr(True) &= \Pr(A) + \Pr(\neg A) - \Pr(False) \\
1 &= \Pr(A) + \Pr(\neg A) - 0 \\
\Pr(\neg A) &= 1 - \Pr(A)
\end{aligned}
$$

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:

  $Cavity = false \wedge Toothache = false$
  $Cavity = false \wedge Toothache = true$
  $Cavity = true \wedge Toothache = false$
  $Cavity = true \wedge Toothache = true$

- Atomic events are mutually exclusive and exhaustive

Prior probability

- Prior or unconditional probabilities of propositions

  e.g., P(Cavity = true) = 0.1 and P(Weather = sunny) = 0.72
  correspond to belief prior to arrival of any (new) evidence
- Probability distribution gives values for all possible assignments:

  P(*Weather*) = <0.72,0.1,0.08,0.1> (*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(*Weather,Cavity*) = *a* 4×2 *matrix of values:*

Weather =	sunny	rainy	cloudy	snow
Cavity = true	0.144	0.02	0.016	0.02
Cavity = false	0.576	0.08	0.064	0.08

Joint Distribution

Weather =	sunny	rainy	cloud	snow
Cavity = true	0.144	0.02	0.016	0.02
Cavity = false	0.576	0.08	0.064	0.08

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

Joint distribution

- Example of joint probability distribution:

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

Conditional probability

- Definition of conditional probability:

$$P(a \mid b) = \frac{P(a \wedge b)}{P(b)}$$

- Product rule gives an alternative formulation:

$$P(a \wedge b) = P(a \mid b)P(b) = P(b \mid a)P(a)$$

- A general version holds for whole distributions, e.g.,

  $\mathbf{P}(Weather, Cavity) = \mathbf{P}(Weather \mid Cavity)\ \mathbf{P}(Cavity)$
- (View as a set of 4 × 2 equations, not matrix multiplication)
- Chain rule is derived by successive product rule

$$\begin{aligned}\mathbf{P}(X_1, \ldots, X_n) &= \mathbf{P}(X_1,\ldots,X_{n-1})\ \mathbf{P}(X_n \mid X_1,\ldots,X_{n-1}) \\ &= \mathbf{P}(X_1,\ldots,X_{n-2})\ \mathbf{P}(X_{n-1} \mid X_1,\ldots,X_{n-2})\ \mathbf{P}(X_n \mid X_1,\ldots,X_{n-1}) = \ldots \\ &= \prod_{i=1}^{n} P(X_i \mid X_1,\ldots,X_{i-1})\end{aligned}$$

Inference by enumeration

- Start with the joint probability distribution:

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

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

Inference by enumeration

- Start with the joint probability distribution:

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

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

- P(*toothache*) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2

Inference by enumeration

- Start with the joint probability distribution:

	toothache		$\neg$ *toothache*	
	catch	$\neg$ *catch*	*catch*	$\neg$ *catch*
cavity	**.108**	**.012**	**.072**	**.008**
$\neg$ *cavity*	**.016**	**.064**	**.144**	**.576**

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

- $P(toothache) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2$

Inference by enumeration

- Start with the joint probability distribution:

	toothache		¬ *toothache*	
	catch	¬ *catch*	*catch*	¬ *catch*
cavity	.108	.012	.072	.008
¬ *cavity*	.016	.064	.144	.576

- Can also compute conditional probabilities:

$$P(\neg cavity \mid toothache) = \frac{P(\neg cavity \wedge toothache)}{P(toothache)} = \frac{(0.016 + 0.064)}{0.108 + 0.012 + 0.016 + 0.064} = 0.4$$

Discrete Random Variables

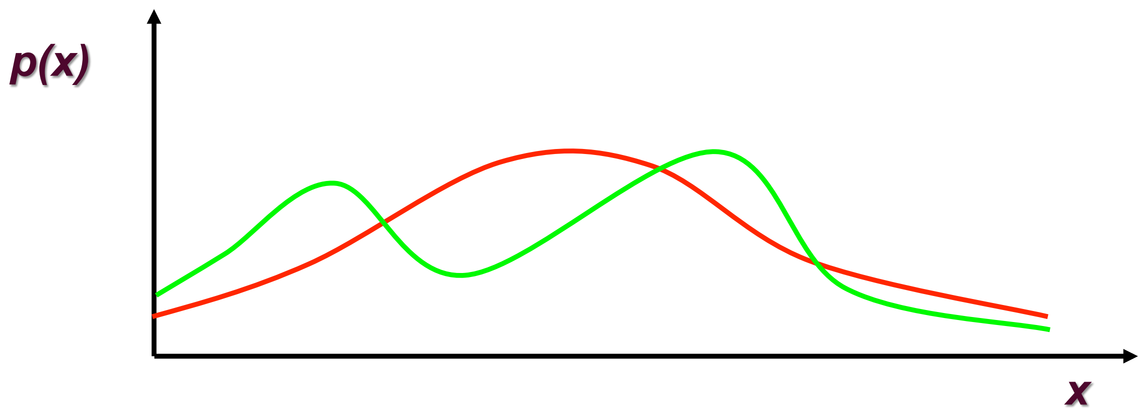
- X denotes a random variable.
- X can take on a countable number of values in $\{x_1, x_2, \ldots, x_n\}$.
- $P(X=x_i)$, or $P(x_i)$, is the probability that the random variable X takes on value x_i.
- $P(\cdot)$ is called probability mass function.
- E.g. $$P(Room) = \langle 0.7, 0.2, 0.08, 0.02 \rangle$$
- This is just shorthand for *P(Room = office), P(Room = kitchen), P(Room = bedroom), P(Room = corridor)*

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*

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

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

- If X and Y are independent then

$$P(x \mid y) = P(x)$$ *(verify using definitions of conditional probability and independence)*

Law of Total Probability, Marginals

Discrete case

Continuous case

Law of total probability

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

$$\int p(x)\, dx = 1$$

Marginalization

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

$$p(x) = \int p(x, y)\, dy$$

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

$$p(x) = \int p(x \mid y) p(y)\, dy$$

Inference by enumeration

- Start with the joint probability distribution:

	toothache		¬ *toothache*	
	catch	¬ *catch*	*catch*	¬ *catch*
cavity	.108	.012	.072	.008
¬ *cavity*	.016	.064	.144	.576

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

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:

$$\forall x : \mathrm{aux}_{x|y} = P(y \mid x)\ P(x)$$

$$\eta = \frac{1}{\sum_x \mathrm{aux}_{x|y}}$$

$$\forall x : P(x \mid y) = \eta\ \mathrm{aux}_{x|y}$$

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

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

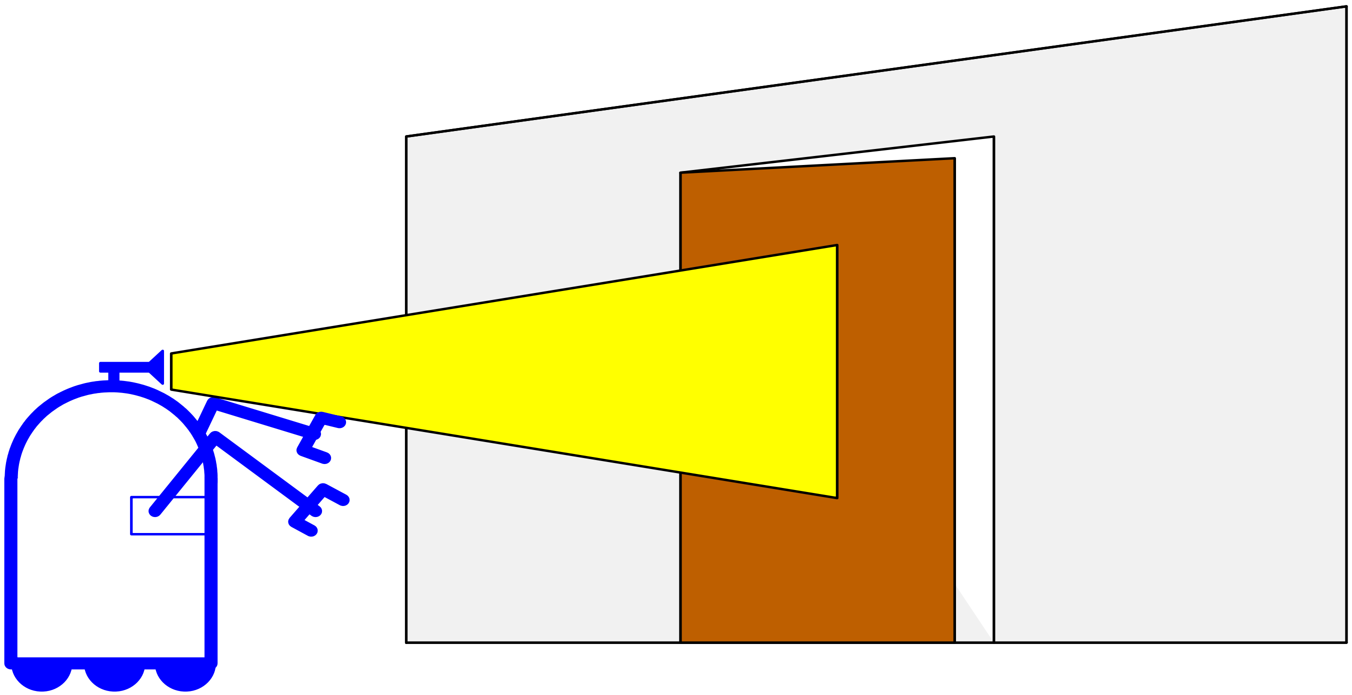
- But this does not necessarily mean

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

(independence/marginal independence)

Simple Example of State Estimation

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



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: count frequencies!

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

Example

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

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

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

- ***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 $P(x| z_1...z_n)$?

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

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

$$= \eta\, P(z_n \mid x)\, P(x \mid z_1, \ldots, z_{n-1})$$

$$= \eta_{1\ldots n} \prod_{i=1\ldots n} P(z_i \mid x)\, P(x)$$

Example: Second Measurement

- $P(z_2|open) = 0.5$ $\quad$ $P(z_2|\neg open) = 0.6$
- $P(open|z_1)=2/3$

$$P(open \mid z_2, z_1) = \frac{P(z_2 \mid open)\, P(open \mid z_1)}{P(z_2 \mid open)\, P(open \mid z_1) + P(z_2 \mid \neg open)\, P(\neg open \mid z_1)}$$

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

- ***z_2 lowers the probability that the door is open***

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*

  change the world

- 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**

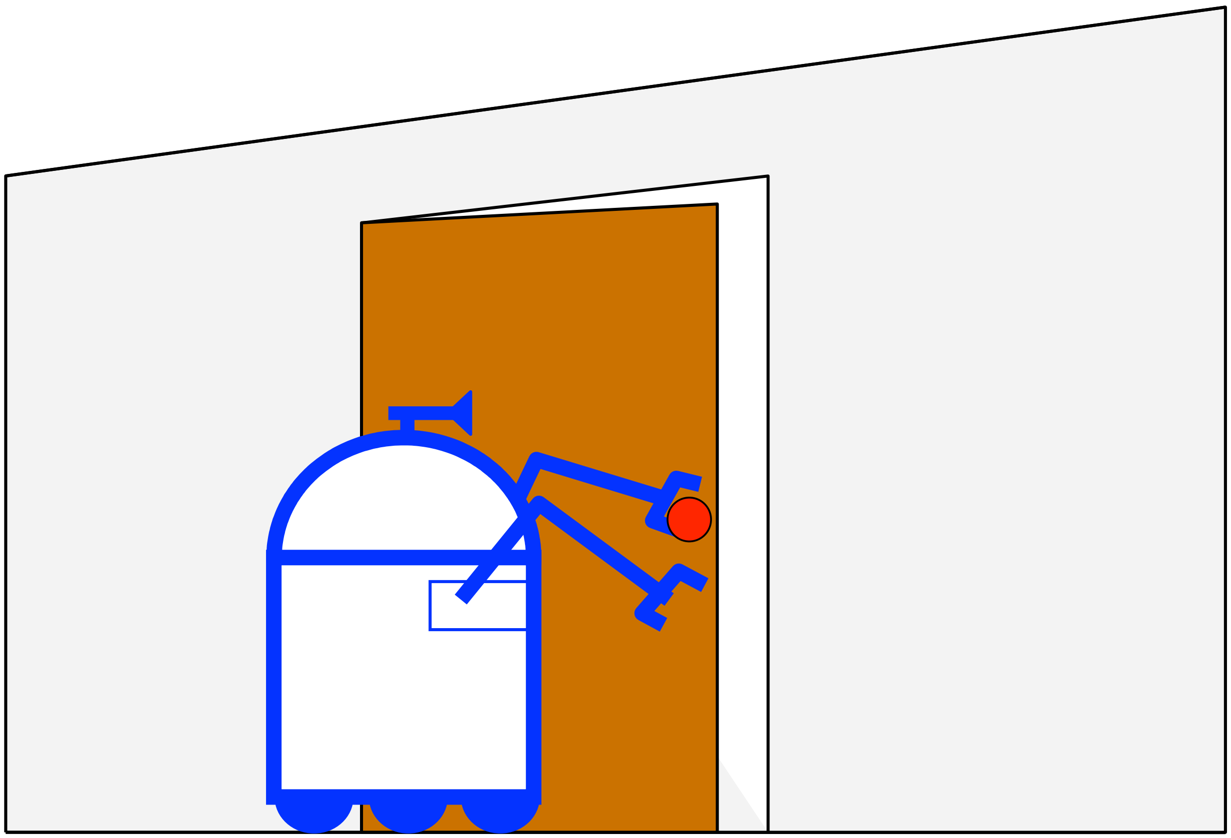
Modeling Actions

- To incorporate the outcome of an action *u* into the current "belief", we use the conditional pdf

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

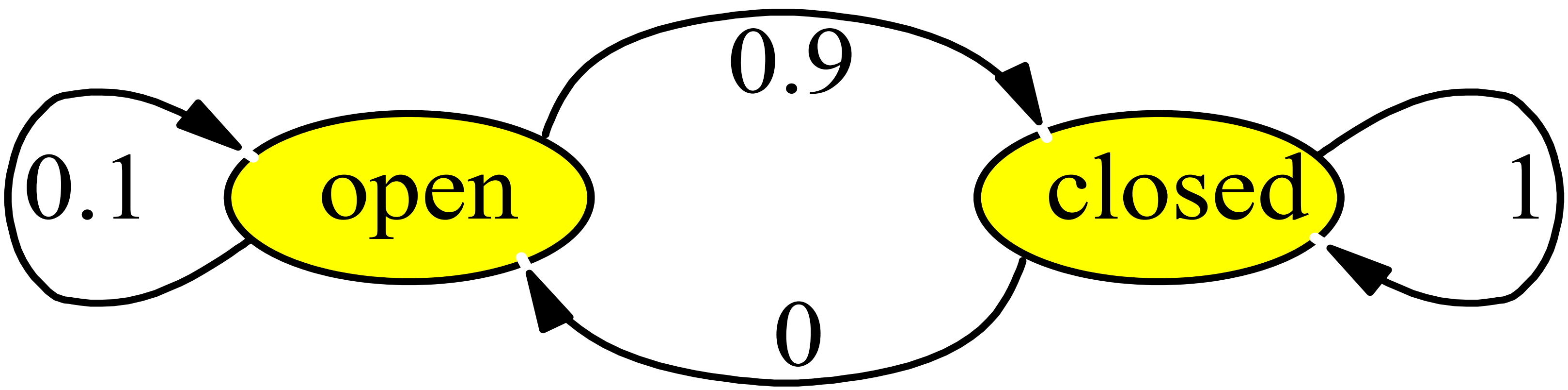
- This term specifies the pdf that **executing *u* changes the state from *x' to x***.

Example: Closing the door



State Transitions

$P(x|u,x')$ for u = "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

$$
\begin{aligned}
P(closed \mid u) &= \sum P(closed \mid u, x')P(x') \\
&= P(closed \mid u, open)P(open) \\
&\quad + P(closed \mid u, closed)P(closed) \\
&= \frac{9}{10} * \frac{5}{8} + \frac{1}{1} * \frac{3}{8} = \frac{15}{16}
\end{aligned}
$$

$$
\begin{aligned}
P(open \mid u) &= \sum P(open \mid u, x')P(x') \\
&= P(open \mid u, open)P(open) \\
&\quad + P(open \mid u, closed)P(closed) \\
&= \frac{1}{10} * \frac{5}{8} + \frac{0}{1} * \frac{3}{8} = \frac{1}{16} \\
&= 1 - P(closed \mid u)
\end{aligned}
$$

Bayes Filters: Framework

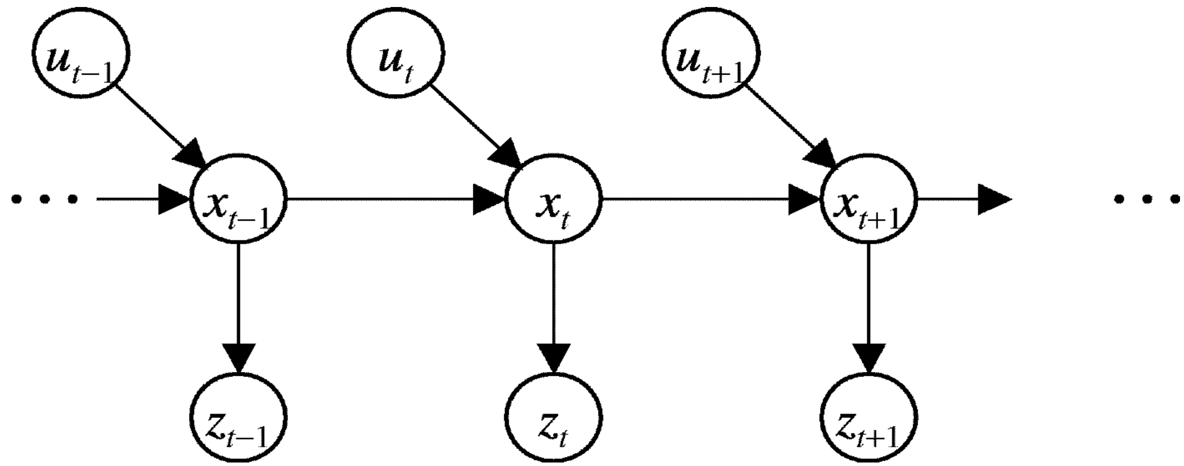
- **Given:**
  - *Stream of observations z and action data u:*

$$d_t = \{u_1, z_1 \ldots, u_t, z_t\}$$

  - *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



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

$$p(x_t \mid x_{1:t-1}, z_{1:t}, u_{1:t}) = p(x_t \mid x_{t-1}, u_t)$$

Underlying Assumptions

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

Bayes Filters

z = observation
u = action
x = state

$$\boxed{Bel(x_t)} = P(x_t \mid u_1, z_1 \ldots, u_t, z_t)$$

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

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

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

```
Algorithm Bayes_filter( Bel(x),d ):
η=0
If d is a perceptual data item z then
    For all x do
        Bel'(x) = P(z | x) Bel(x)
        η = η + Bel'(x)
    For all x do
        Bel'(x) = η^{-1} Bel'(x)
Else if d is an action data item u then
    For all x do
        Bel'(x) = ∫ P(x | u, x') Bel(x') dx'
Return 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.