Stochastic Estimation & Probabilistic Robotics

Probability Theory:

Joey Durham

April 2nd, 2008

Slides adapted from introduction.ppt, originally by Thrun, Burgard, and Fox at probabilistic-robotics.org

Reading Group Schedule

W1: Wed Apr 2	Probability Theory	Joey
W2: Wed Apr 9	Estimation Theory	Kurt
W3: Wed Apr 16	Kalman Filters	
W4: Wed Apr 23	Particle Filters	
W5: Wed Apr 30	Motion & Sensor Models	Paolo
W6: Wed May 7	Localization	
W7: Wed May 14	Mapping	
W8: Wed May 21	Simultaneous Localization & Mapping	
W9: Wed May 28	Markov Decision Processes	Alexandre
W10: Wed Jun 4	Data Association/Target Tracking	

Projector in room.

Presentation laptop available if needed, contact Joey.

Outline

- Probability theory
- Probability density functions
- Gaussian random variables
- Conditional probability
- Bayes formula
- Stochastic processes
- Markov processes and chains
- Bayes filters

Motivation

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

- Perception = state estimation
- Action = utility optimization

Axioms of Probability Theory

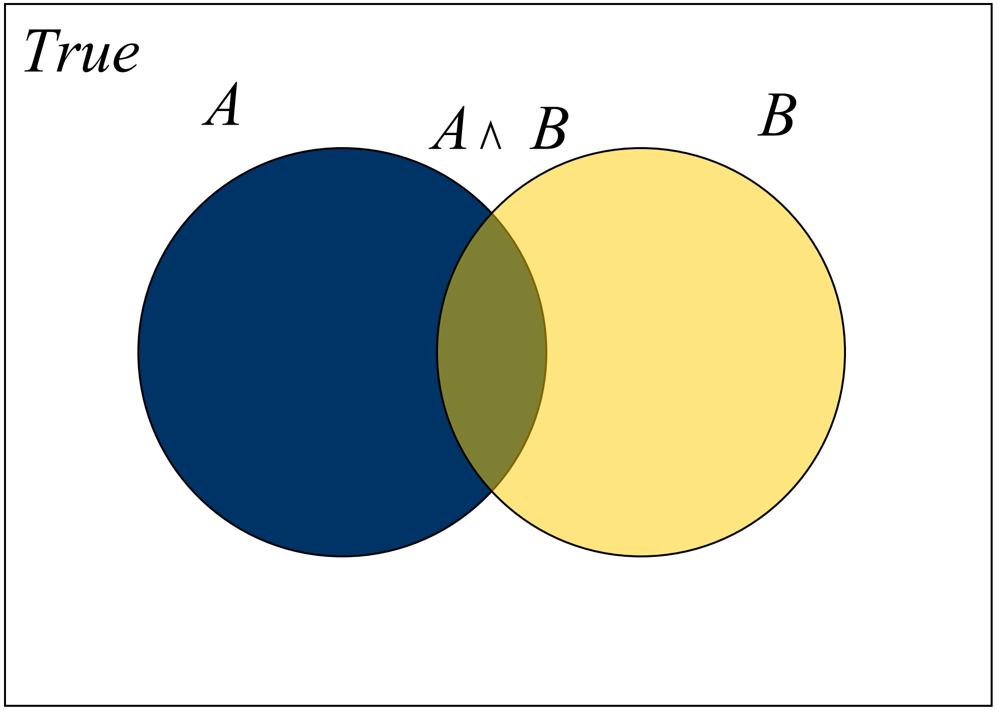
Pr*(A)* denotes probability that proposition *A* is true.

Let S be the set of all possible outcomes.

- $$0 \leq \Pr(A) \leq 1$$
- $$\Pr(S) = 1 \qquad \Pr(\emptyset) = 0$$
- $$\Pr(A \vee B) = \Pr(A) + \Pr(B) - \Pr(A \wedge B)$$

A Closer Look at Axiom 3

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



Using the Axioms

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

Discrete Random Variables

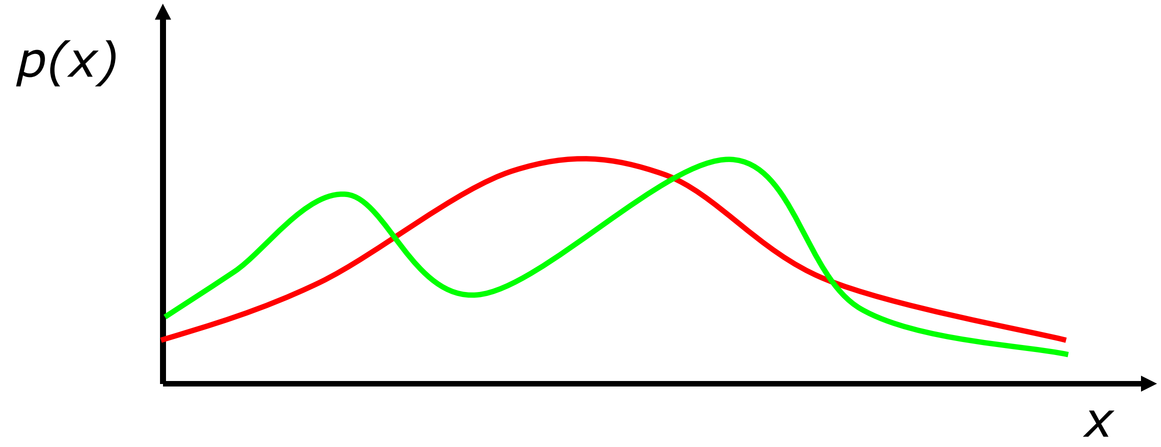
- *X* denotes a random variable.
- *X* can take on a countable number of values in $\{x_1, x_2, ..., x_n\}$.
- $P(X=x_i)$, or $P(x_i)$, is the probability that the random variable X takes on value x_i.
- $P(*)$ is called probability mass function.
- A proper pmf satisfies: $\sum_x P(x) = 1$

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.



Properties of PDFs

- Normalization property

$$\int_{-\infty}^{\infty} p(x)dx = 1$$

- Example: Uniform random variable

$$p(x)=\begin{cases} \frac{1}{(b-a)} & x\in[a,b] \\ 0 & elsewhere \end{cases}$$

Expectations and Moments

- Expectation value of a scalar random variable (aka mean or average):

$$E[x] = \int_{-\infty}^{\infty} x p(x) dx = \bar{x}$$

- *n*th moment:

$$E[x^n] = \int_{-\infty}^{\infty} x^n p(x) dx$$

Variance

- The 2nd central moment is also known as the variance:

$$\mathrm{var}(x) = E[(x - \bar{x})^2] = \int_{-\infty}^{\infty} (x - \bar{x})^2 p(x)dx$$

$$\mathrm{var}(x) = E[x^2] - (\bar{x})^2 = \sigma_x^2$$

- The square root of the variance, σ, is also called the standard deviation.

Joint Probability

- $P(X=x$ and $Y=y) = P(x,y)$

- If X and Y are independent then

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

Covariance

- The covariance of two scalar random variables *x* and *y*:

$$\mathrm{cov}(x, y) = E[(x - \bar{x})(y - \bar{y})] = \sigma_{xy}^{2}$$

Correlation

- The correlation coefficient between x and y:

$$\rho_{xy} = \frac{\sigma_{xy}^2}{\sigma_x \sigma_y}$$

- Because of normalization:

$$\left|\rho_{xy}\right| \leq 1$$

More on Correlation

- Uncorrelated:

$$\left|\rho_{xy}\right| = 0$$

$$E[xy] = E[x]E[y]$$

- Linearly dependent:

$$\left|\rho_{xy}\right| = 1$$

$$ax + by = 0$$

Joint and Marginal PDFs

- Marginal PDF for one random variable:

$$p(x) = \int_{-\infty}^{\infty} p(x, y)dy$$

- If a set of random variables are independent, their joint PDF satisfies:

$$p(x, y) = p(x)p(y)$$

Random Vectors

- Vector-valued random variable:

$$x = [x_1 \cdots x_n]$$

- Expectation value of *x*:

$$E[x] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x p(x) dx_1 \cdots dx_n = \bar{x}$$

- The covariance matrix of *x*:

$$\operatorname{cov}(x) = E[(x - \bar{x})(x - \bar{x})'] = P_{xx}$$

Characteristic Function

- The characteristic function of *x* is the n-dimensional Fourier transform of its PDF:

$$M_x(s) = E[e^{s'x}] = \int_{-\infty}^{\infty} e^{s'x} p(x) dx$$

- The moments of *x* can be found using gradients of M_x, eg:

$$E[x] = \nabla_s M_x(s) |_{s=0}$$

- Characteristic function = moment generating function.

Gaussian distributions

- The PDF of a Gaussian or normal random variable:

scalar: $$p(x) = N(x; \bar{x}, \sigma^2) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}}$$

vector: $$p(x) = \left|2\pi P\right|^{-1/2} e^{-(1/2)(x-\bar{x})' P^{-1} (x-\bar{x})}$$

- Has a mean of $E[x]$ and a variance of σ^2.

Joint Gaussians

- To variables x and z are jointly Gaussian if:

$$y = \begin{bmatrix} x \\ z \end{bmatrix} \qquad p(x,z) = p(y) = N(y; \bar{y}, P_{yy})$$

- The mean and covariance of *y*:

$$\bar{y} = \begin{bmatrix} \bar{x} \\ \bar{z} \end{bmatrix} \qquad P_{yy} = \begin{bmatrix} P_{xx} & P_{xz} \\ P_{zx} & P_{zz} \end{bmatrix}$$

Conditional Gaussians

- The conditional PDF for *x* given *z:*

$$p(x \mid z) = \frac{p(x,z)}{p(z)}$$

- The conditional mean and covariance of *x* given *z*:

$$E[x \mid z] = \bar{x} + P_{xz}P_{zz}^{-1}(z - \bar{z})$$

$$\operatorname{cov}(x \mid z) = P_{xx} - P_{xz}P_{zz}^{-1}P_{zx}$$

Mixture PDFs

- A mixture PDF is a weighted sum of PDFs:

$$p(x) = \sum_{j=1}^{n} a_j p_j(x)$$

- Mean and covariance of a mixture:

$$\bar{x} = \sum_{j=1}^{n} a_j \bar{x}_j$$

$$\mathrm{cov}(x) = \sum_{j=1}^{n} a_j P_j + \sum_{j=1}^{n} a_j \bar{x}_j \bar{x}_j' - \bar{x}\bar{x}'$$

Conditional Probability

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

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

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

- If X and Y are independent then

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

- The same rules hold for PDFs:

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

Conditional Expectation

- Conditional expectation, expectation with respect to a conditional PDF:

$$E[x \mid z] = \int_{-\infty}^{\infty} x p(x \mid z) dx$$

- Law of iterated expectations:

$$E[E[x \mid z]] = E[x]$$

Total Probability Theorem

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 Formula

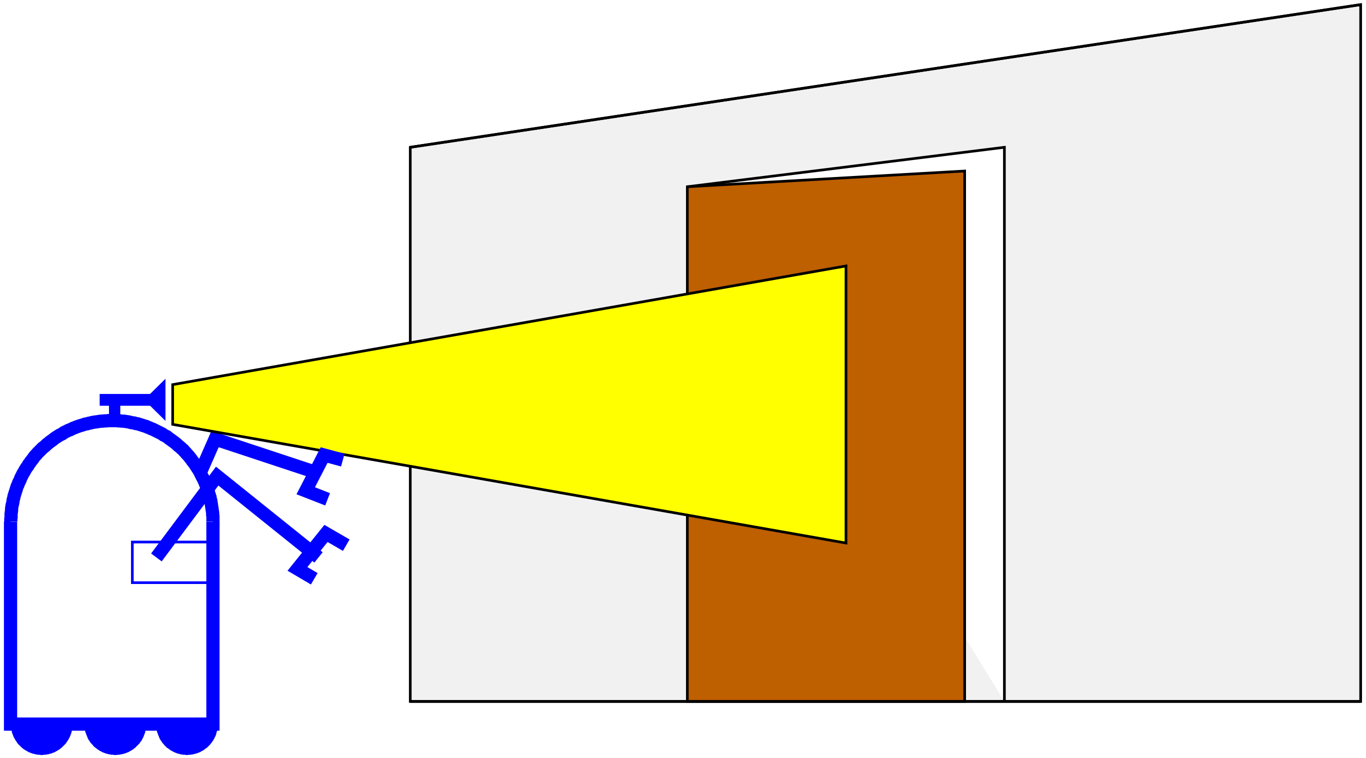
$$P(x, y) = P(x \mid y)P(y) = P(y \mid x)P(x)$$

$$\Rightarrow$$

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

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:

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

Example

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

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

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

- *z* raises the probability that the door is open.

Combining Evidence

- Suppose our robot obtains another observation z_2.

- How can we integrate this new information?

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

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

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

Bayes Rule with Background Knowledge

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

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

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

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.*

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

Example: Second Measurement

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

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

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

- 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**.

Stochastic Processes

- A function of time and some random experiment *w*:

$$x(t) = x(t, w)$$

- Mean of the stochastic process at *t*:

$$\bar{x}(t) = E[x(t)] = \int_{-\infty}^{\infty} \xi p_{x(t)}(\xi)d\xi$$

Properties of Stochastic Processes

- Autocorrelation:

$$R(t_1, t_2) = E[x(t_1)x(t_2)]$$

- Autocovariance:

$$V(t_1, t_2) = E\left[\left(x(t_1) - \bar{x}(t_1)\right)\left(x(t_2) - \bar{x}(t_2)\right)\right]$$

$$= R(t_1, t_2) - \bar{x}(t_1)\bar{x}(t_2)$$

More Properties

- Stationary if for all t_1 & t_2:

$$E[t_1] = E[t_2]$$

$$R(t_1, t_2) = R(t_1 - t_2)$$

- Ergodic if stationary and:

$$\lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} x(t)dt = \bar{x}$$

Random Walk

- Wiener-Levy or Brownian motion, steps of size s at intervals of Δ s.t.:

$$\frac{s}{\sqrt{\Delta}} \rightarrow \sqrt{\alpha}$$

- Produces stochastic process *w(t)* with a Gaussian PDF:

$$p(w(t)) = N(w(t);0,\alpha\, t)$$

Markov Processes

- "The future is independent of the past if the present is known"
- Brownian motion is a Markov process as:

$$w(t) = w(t_1) + \int_{t_1}^{t} n(\tau)\,d\tau$$

- Also, LTI excited by stationary white noise

$$\dot{x}(t) = Ax(t) + Bn(t)$$

is a stationary Markov process.

Random Sequences

- Time-indexed sequence of random variables:

$$X^k = \{x(j)\}_{j=1}^{k} \qquad k = 1,2,\ldots$$

- A sequence is Markov if:

$$p(x(k) \mid X^j) = p(x(k) \mid x(j))$$

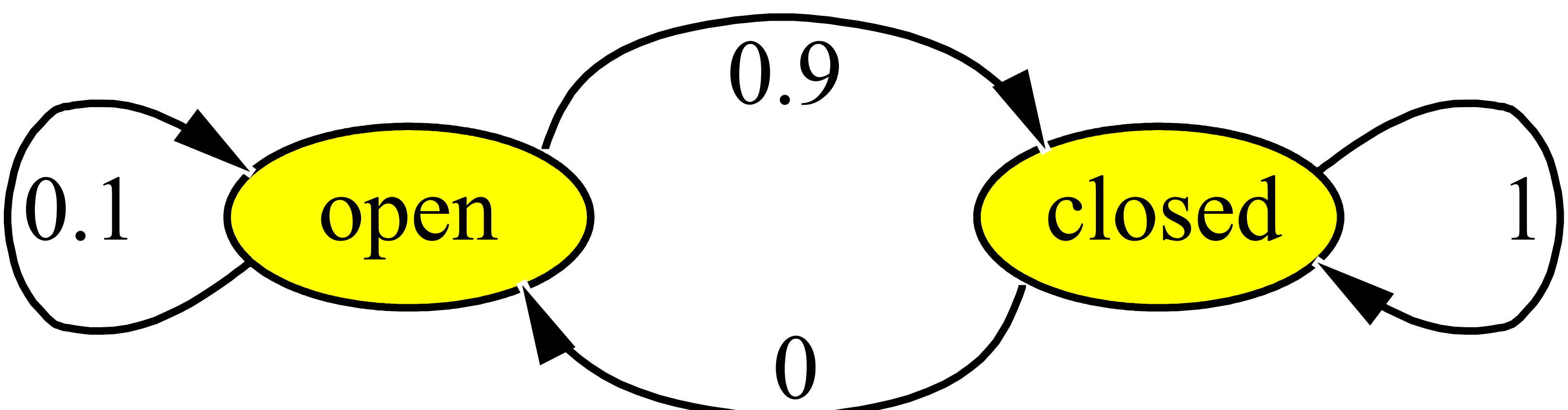
Markov Chains

- A Markov sequence in which state space is discrete and finite:

$$x(k) \in \{x_i, i = 1 \dots n\}$$

- With state transition probabilities:

$$P\{x(k) = x_j \mid x(k-1) = x_i\} = \pi_{ij}$$



More Markov Chains

- Vector of probabilities of being in each state:

$$u(k) = \left[u_1(k), \ldots, u_n(k)\right]$$

$$u_1(k) = P\left\{x(k) = x_i\right\}$$

- Time evolution given by:

$$u_i(k+1) = \sum_{j=1}^{n} \pi_{ij} u_j(k) \qquad i = 1 \ldots n$$

Law of Large Numbers

- Sum of a large number of sufficiently uncorrelated random variables tends towards the expected value
- Given stationary random sequence *x* with:

$$\lim_{|i-j|\to\infty} \rho(i-j) = 0$$

if correlation coefficients -> 0 "sufficiently fast", then

$$\lim_{n\to\infty}\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \bar{x}$$

Central Limit Theorem

- If a sequence consists of independent random variables, then the PDF of

$$z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_i$$

will tend towards a Gaussian.

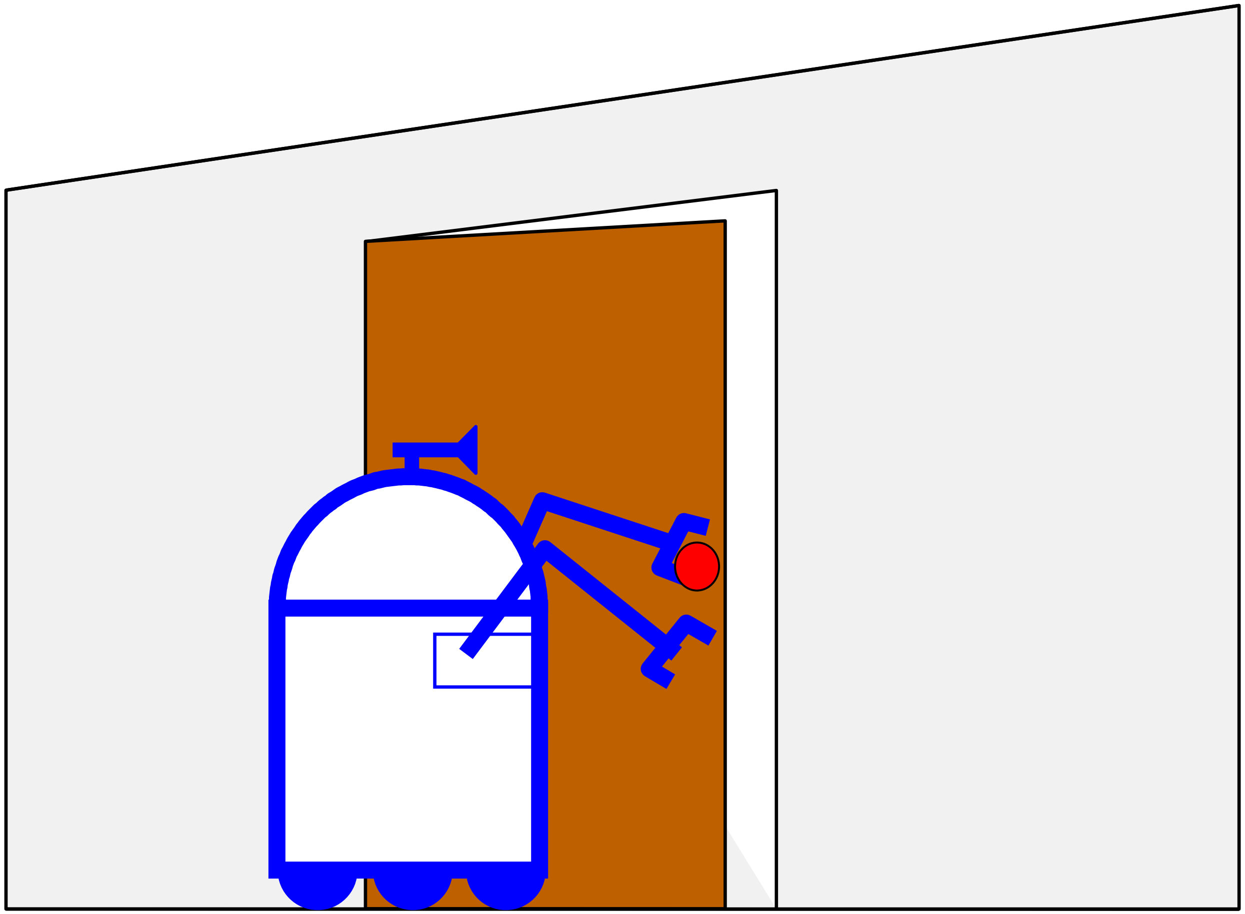
Modeling Actions

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

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

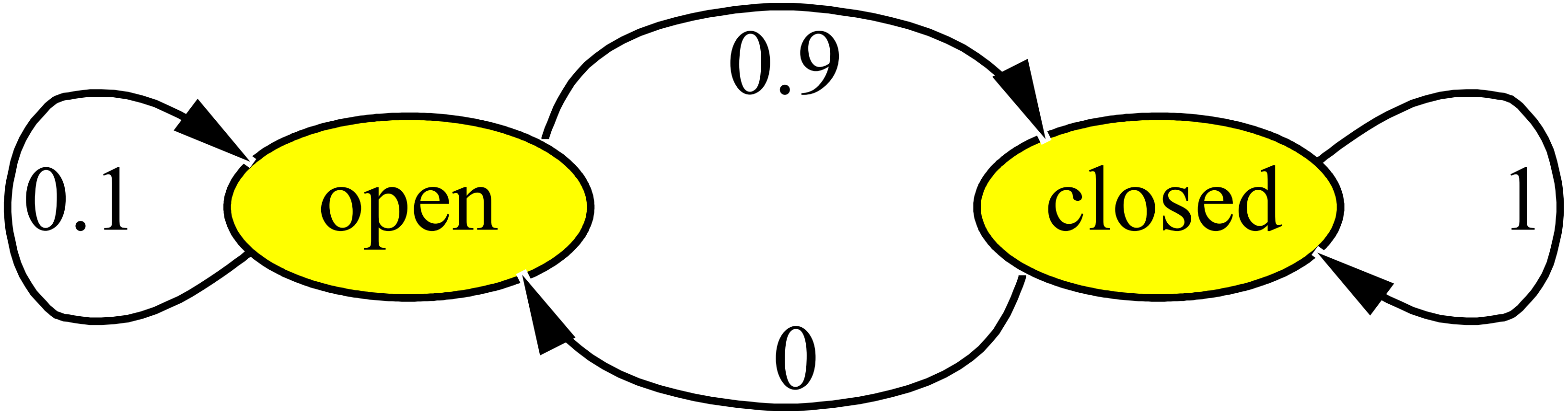
- This term specifies the pdf that **executing *u* changes the state from *x'* to *x*.**

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

$$P(closed \mid u) = \sum P(closed \mid u, x')P(x')$$

$$= P(closed \mid u, open)P(open)$$

$$+ P(closed \mid u, closed)P(closed)$$

$$= \frac{9}{10} * \frac{5}{8} + \frac{1}{1} * \frac{3}{8} = \frac{15}{16}$$

$$P(open \mid u) = \sum P(open \mid u, x')P(x')$$

$$= P(open \mid u, open)P(open)$$

$$+ P(open \mid u, closed)P(closed)$$

$$= \frac{1}{10} * \frac{5}{8} + \frac{0}{1} * \frac{3}{8} = \frac{1}{16}$$

$$= 1 - P(closed \mid u)$$

Bayes Filters: Framework

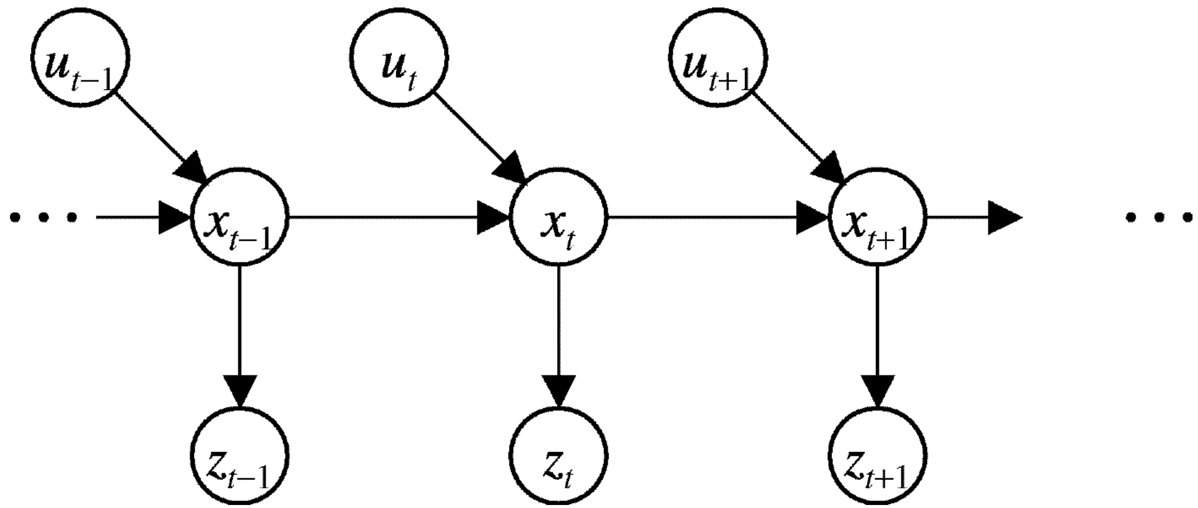
- **Given:**
  - Stream of observations *z* and action data *u:*

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

  - 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



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

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

Underlying Assumptions

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

Bayes Filters

z = observation
u = action
x = state

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

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

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

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

Algorithm **Bayes_filter**(*Bel(x),d*):

$\eta = 0$

If *d* is a perceptual data item *z* then

For all *x* do

$Bel'(x) = P(z \mid x) Bel(x)$

$\eta = \eta + Bel'(x)$

For all *x* do

$Bel'(x) = \eta^{-1} Bel'(x)$

Else if *d* is an action data item *u* then

For all *x* do

$Bel'(x) = \int P(x \mid u, x') \; Bel(x') \; dx'$

Return *Bel'(x)*

Bayes Filters are Common

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

- 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.