State Estimation-I

Lecture 5 – Thursday October 27, 2016

Objectives

When you have finished this lecture you should be able to:

- Recognize different **imperfection aspects** of data collected for situation awareness.
- Recognize different sources of **uncertainty** in autonomous systems and understand the problem of **state estimation**.
- Understand **Bayesian probability** as a commonly used and indispensable framework to handle quantitatively with uncertainty in autonomous systems.
- Understand **Bayes filter** and its role in state estimation.

Outline

- Uncertainty
- State Estimation
- Basic Concepts in Probability
- Bayesian Rule
- Environment Interaction
- Bayes Filter Algorithm
- Summary

Outline

- **Uncertainty**
- State Estimation
- Basic Concepts in Probability
- Bayesian Rule
- Environment Interaction
- Bayes Filter Algorithm
- Summary

Uncertainty

- **Situation Awareness**

Situation awareness is:

◇ the **perception** of environmental elements with respect to time and/or space,

◇ the **comprehension** of their meaning, and

◇ the **projection** of their status after some variable has changed, such as time, or some other variable, such as a predetermined event. [1]

Uncertainty

- **Situation Awareness**

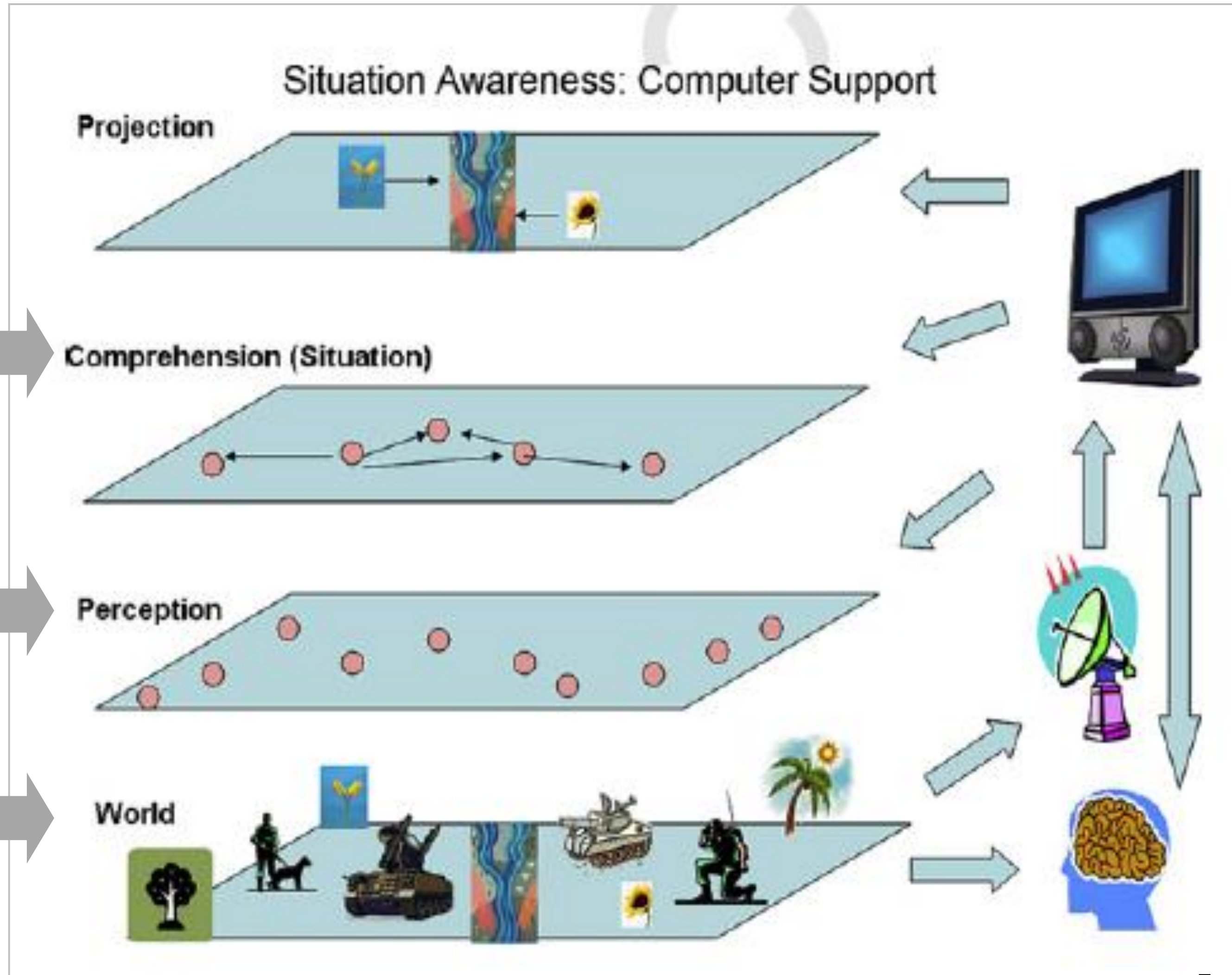
Semantic Analysis
- Thematic
- Spatio-Temporal
- Trust

Provenance

Relate Situation Entities

Identify Situation Entities

Collect Relevant Data

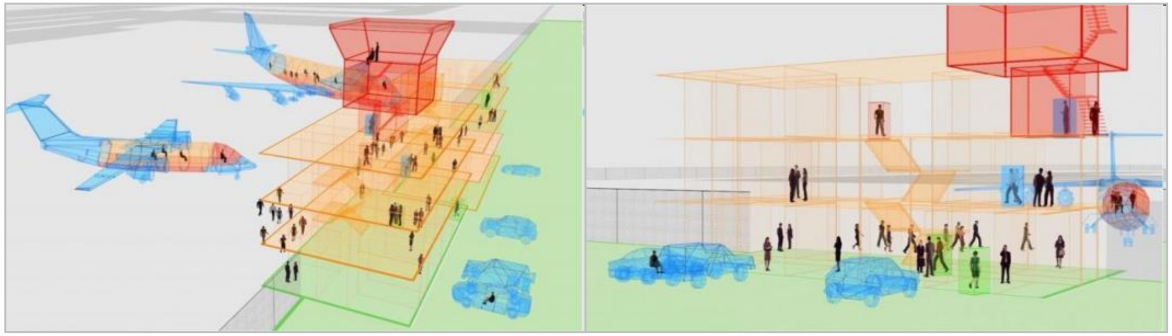


[2]

Uncertainty

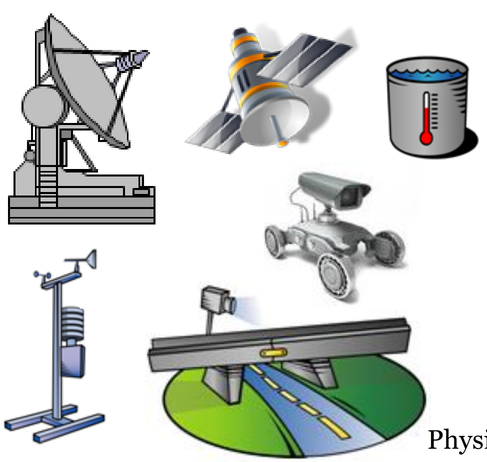
- **Situation Awareness**

Airport as Volume of Interest (VOI)



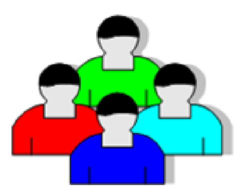
Sources of data

Sensor-Space



Physical sensors

Human-Space



Ad-hoc observers

Computer-Space

Archived documents and social media

Uncertainty

Data Imperfection

Uncertainty

Data is uncertain when the associated **confidence degree**, about what is stated by the data, is **less than 1**

Imprecision

Imprecise data is that data which **refers to several**, rather than only one, object(s)

Granularity

Data granularity refers to the ability to **distinguish among objects**, which are described by data, being dependent on the provided set of attributes

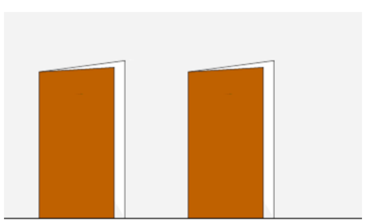
[3]

Uncertainty

Data Imperfection

Uncertainty

- "I think there are two doors in front of me".
- The number is exact but we are not sure.
- The associated confidence or belief degree<1.



Imprecision

- "There are at least two doors in front of me".
- The number of doors could be two or more.

Vagueness

- "The door is wide."
- The assigned attribute "wide" is not well-defined as it can be interpreted subjectively.

Ambiguity

- Door entrance is between 80 and 120 cm.

Incompleteness

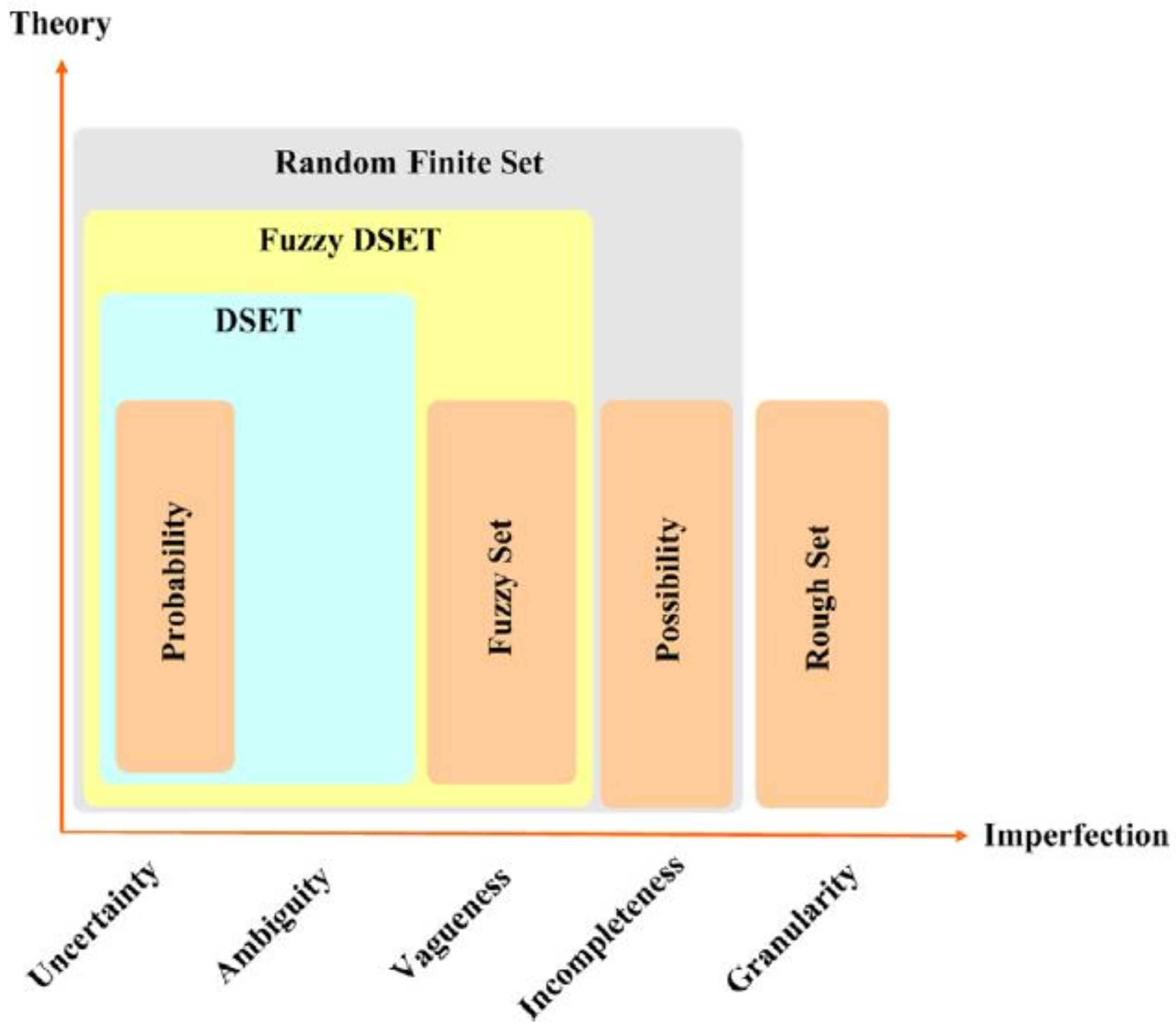
- "It is not possible to cross the door"
- Some information missing.

Granularity

- Coarse-grained: E5 6006
- Fine-grained: Building: E5, Floor: 6, Room: 006

[3]

Uncertainty

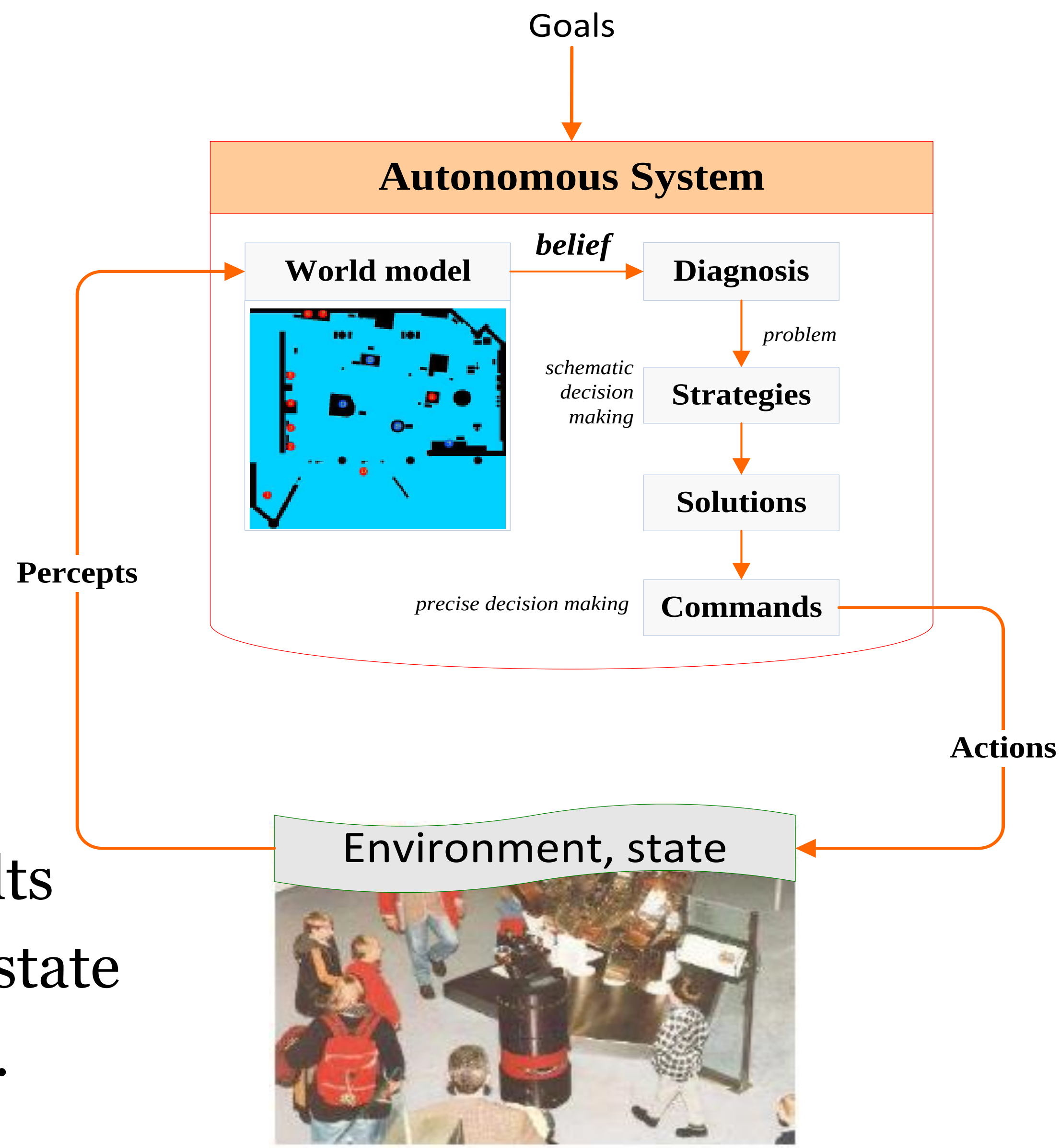


[3]

Uncertainty

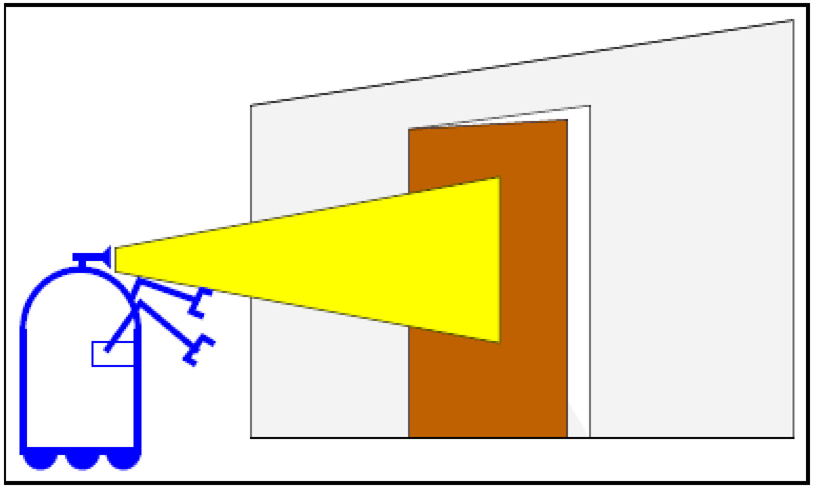
- **Autonomous systems** are able to perceive the physical world and physically interact with it through computer-controlled mechanical devices.

- A critical problem of autonomous systems is **uncertainty**, which results in wrong beliefs about its state and/or environment state.



Uncertainty

- ***Example:*** A robot is driving in front of a door.
- The robot is estimating the **state of a door** (open or closed) using its camera.



- Assume the robot's sensors are noisy.
- If **mistaking** a closed door for an open one incurs costs (e.g., the **robot crashes into a door**).

[4]

Uncertainty

- **Uncertainty** or lack of certainty arises if the system **lacks critical information** for carrying out its task.
- This **critical information** include, but is not limited to:
  - ◇ The robot pose.
  - ◇ The configuration of the robot's actuators, such as the joints of robotic manipulators.
  - ◇ The robot velocity and the velocities of its joints.
  - ◇ The location and features of surrounding objects in the environment.
  - ◇ The location and velocities of moving objects and people.
  - ◇ Others, sensor status, battery status…

Uncertainty

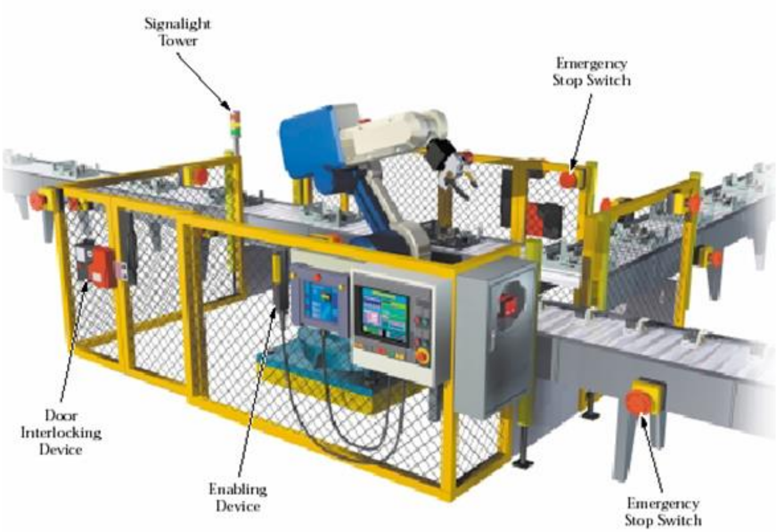
- In robotics for example, **uncertainty** arises from **five** different factors:

1. Environments
2. Sensors
3. Robots
4. Models
5. Computation

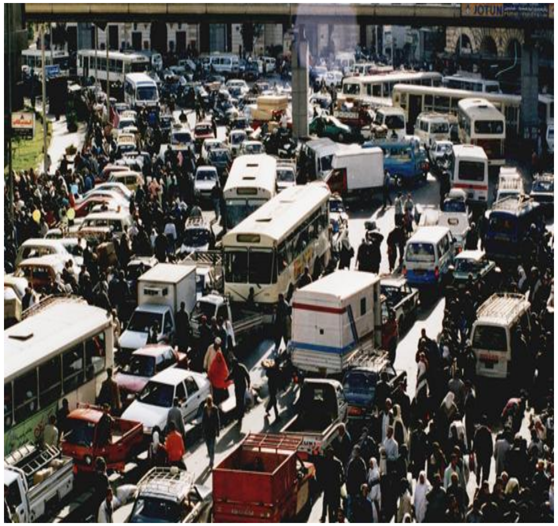
Uncertainty

- **Environment:**

  Physical worlds are inherently unpredictable. While the **degree of uncertainty** in well-structured environments such assembly lines is small, environments such as highways and private homes are **highly dynamic and unpredictable**.



Well-structured Static Environment



Unstructured Dynamic Environments

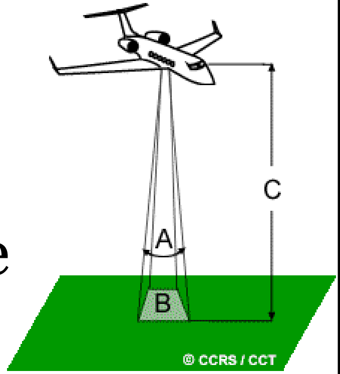
Uncertainty

- **Sensors:**

Sensors are **imperfect devices** with errors. They are inherently limited in what they can perceive.

Limitations arise from two primary factors:

1. **Range and resolution** of a sensor is subject to physical laws. For example, cameras can't see through walls, and even within the perceptual range the spatial resolution of camera images is limited.



2. Sensors are subject to **noise**, which perturbs sensor measurements in unpredictable ways and hence limits the information that can be extracted from sensor measurements.

Uncertainty

- **Sensors:**

Sensor Errors

Systematic Errors/ Deterministic Errors

Systematic errors are characterized by being **consistent and repeatable**. Such as Calibration Errors, Loading Errors (if the sensor is intrusive), Environmental Errors and Common Representation Format Errors.

Random Errors/ Non-deterministic Errors

Random errors are characterized by a **lack of repeatability** in the output of the sensor. They cause by **measurement noise**. For example, they may be due to fluctuations in the capacity of the resistance of electrical circuits in the sensor, or due to the limited resolution of the sensor.

lead to sensor uncertainty

Uncertainty

- **Robots:**

  Robot actuation involves motors that are, at least to some extent, unpredictable, due effects like **control noise and wear-and-tear**.

  Some actuators, like **low-cost mobile robots**, can be **extremely inaccurate**.

  Others, such as heavy-duty industrial robot arms and **research mobile robots**, are **quite accurate**.

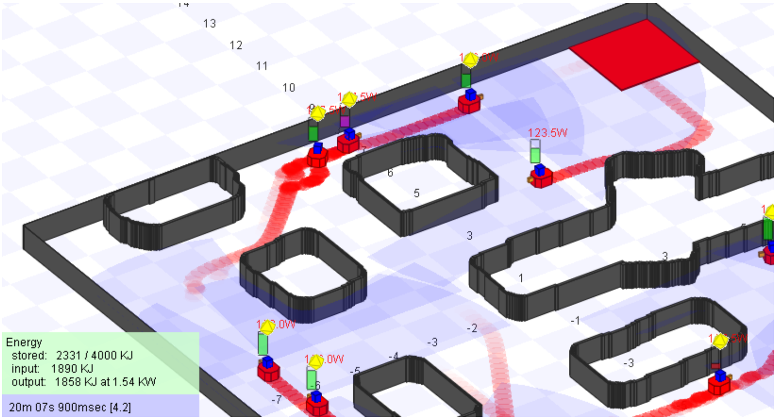
muRata Boy from muRata Manufacturing

Uncertainty

- **Models:**

Models are **inherently inaccurate**. Models are abstractions of the real world. As such, they only partially model the underlying physical processes of the robot and its environment.

Model errors are a **source of uncertainty** that has largely been ignored in robotics, despite the fact that most robotic models used in state-of-the-art robotic systems are rather crude.



Uncertainty

- **Computation:**

  Autonomous systems are **real-time systems**, which limits the amount of computation that can be carried out.

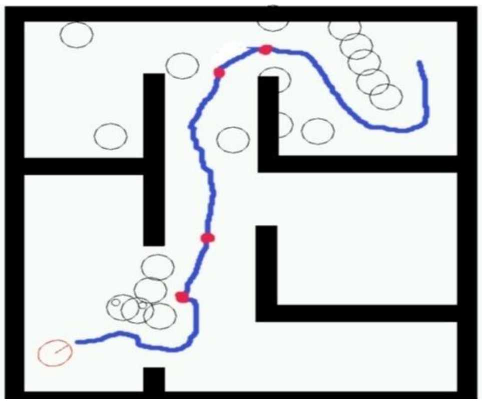
  Many state-of-the-art algorithms are **approximate**, achieving **timely response through sacrificing accuracy**.

Outline

- Uncertainty
- **<u>State Estimation</u>**
- Basic Concepts in Probability
- Bayesian Rule
- Environment Interaction
- Bayes Filter Algorithm
- Summary

State Estimation

- **State estimation** addresses the problem of estimating quantities from sensor data that are **not directly observable**, but that can be inferred.
- In most robotic applications, determining what to do is relatively easy if one only knew certain quantities.
- For example, moving a mobile robot is relatively easy if the exact **location of the robot** and **all nearby obstacles** are known.



- Unfortunately, these **variables** are **not directly measurable**.
- Instead, a robot has to rely on its **sensors to gather this information**.

State Estimation

- **Typical State variables:**
  - ◇ The robot pose.
  - ◇ The configuration of the robot's actuators, such as the joints of robotic manipulators.
  - ◇ The robot velocity and the velocities of its joints.
  - ◇ The location and features of surrounding objects in the environment.
  - ◇ The location and velocities of moving objects and people.
  - ◇ Others, sensor status, battery status…

State Estimation

- Sensors carry only **partial information** about those quantities, and their measurements are **corrupted by noise**.
- **Error of a sensor** is the difference between the sensor's output measurements and the true values being measured, within some specific operating context.

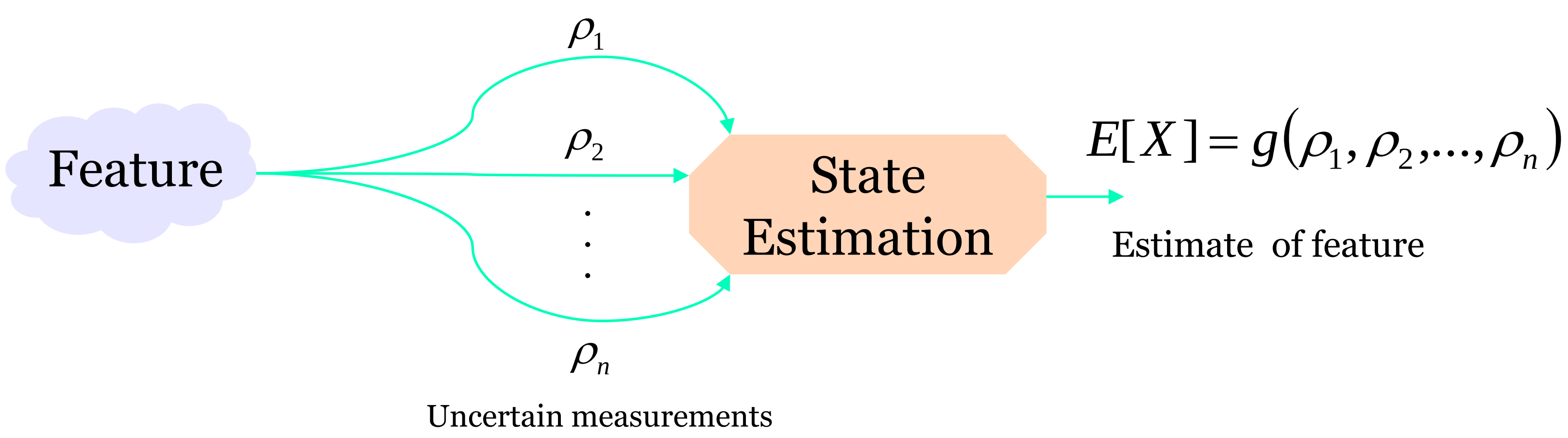
$$\boldsymbol{Error = m - v}$$

where
m is the measured value
v is the true value.

- From a **statistical point of view**, we wish to characterize the error of a sensor, not for one specific measurement but for **any measurement**.

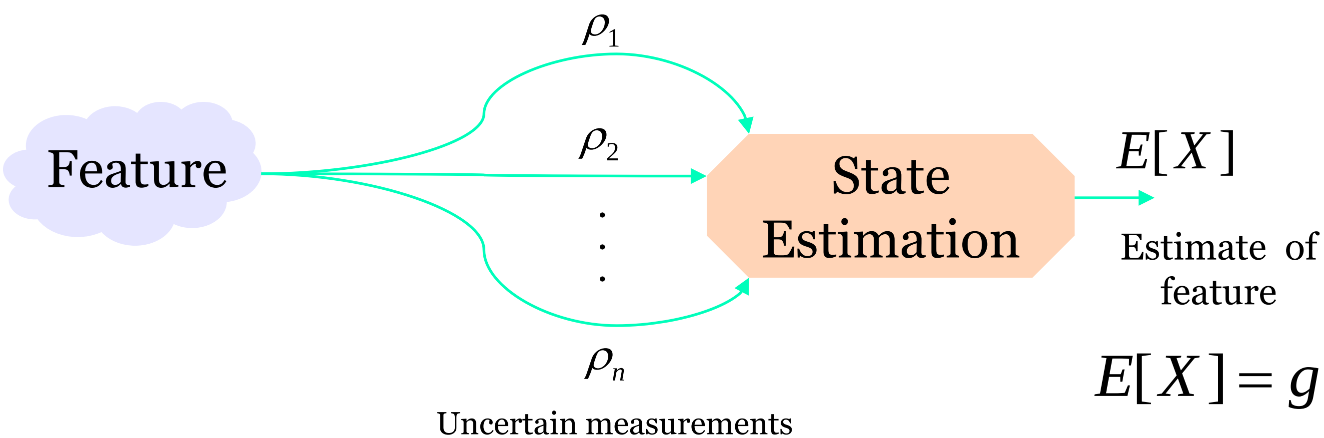
State Estimation

- Let us formulate the **problem of sensing** as an **estimation problem**.
- The sensor has taken a set of ***n*** measurements with values ρ_i.
- The goal is to characterize the **estimate** of the true value ***E[X]*** given these measurements:



State Estimation

- From this perspective, the **true value** is represented by a random (and therefore unknown) variable **X**.
- State estimation seeks to **recover state variables from the data**.



$$E[X] = g(\rho_1, \rho_2, \ldots, \rho_n)$$

State Estimation

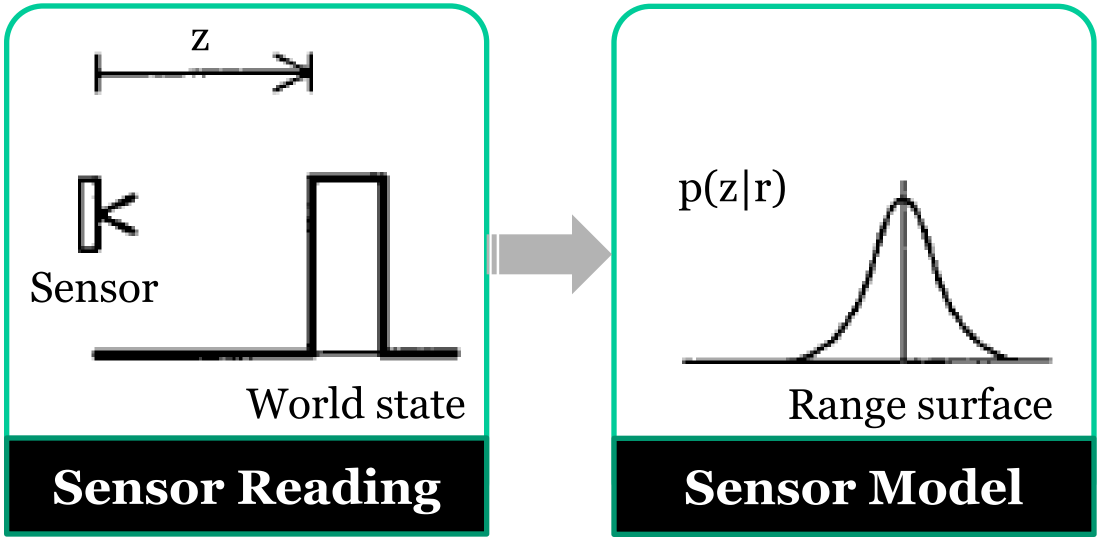
Modeling Dimensions

- **Uncertainty**
- **Partial Observability (PO)**
- **Time**
- **Decision**
- **Learning**
- **Multiagent**

The more dimensions you use, the less computational tractability you get

State Estimation

- **Modeling Uncertainty**



◇ To interpret the range data obtained from a given sensing device, **stochastic sensor model** is used. This model is defined by a probability density function (pdf).

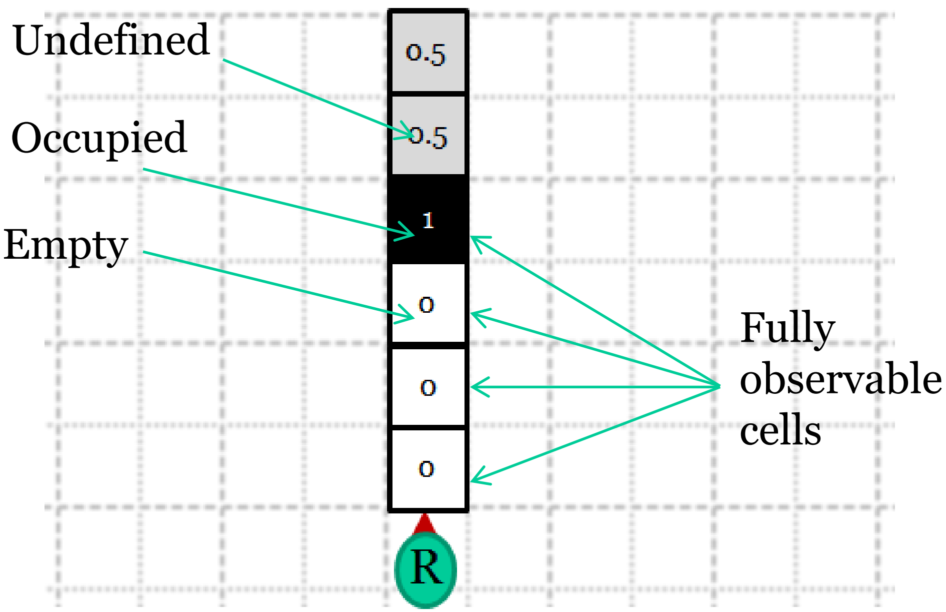
◇ This pdf is of the form **p(z|r)** and relates reading/observation of measurement **z** with the true parameter range value **r**.

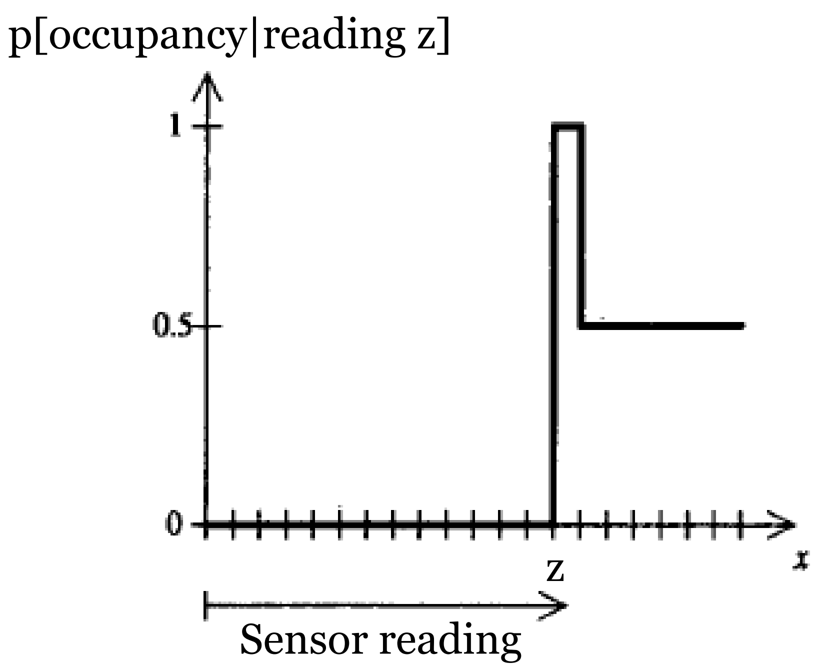
State Estimation

- **Partial Observability**

  A partially observable system is one in which the entire state of the system is **not fully visible** to an **external sensor**.

  In a partially observable system the observer may utilize a **memory** system in order to add information to the observer's understanding to the system.





State Estimation

- **Partial Observability**

Fully observable system

In **chess** (apart from the 'who is moving next' state) the full state of the system is observable at any point in time.

Partially observable system

In **card game**, some of the cards are discarded into a pile face down. In this case the observer is only able to view their own cards and potentially those of the dealer. They are not able to view the face-down (used) cards, and nor are they able to view the cards which will be dealt at some stage in the future. A memory system can be used to remember the previously dealt cards that are now on the used pile. This adds to the total sum of knowledge that the observer can use to make decisions.

Source: http://en.wikipedia.org/wiki/Partially_observable

State Estimation

Technique	Uncertainty	Time	Decision	Partial Observability	Learning	Multiagent
Bayes Filters	×	×				
Bayesian Networks (BN)	×		×			
Dynamic Bayesian Network (DBN)	×	×	×			
Hidden Markov Models (HMMs)	×	×		×		
Markov Decision Process (MDP)	×	×	×			
Reinforcement Learning (RL)	×	×	×		×	
Partial Observerable MDP (POMDP)	×	×	×	×		
Multiagent MDP	×	×	×			×
Multiagent RL	×	×	×		×	×
Partial Observerable RL (PO-RL)	×	×	×	×	×	
Stochastic Game (SG)	×	×	×	×		×
Partial Observerable Multiagent RL	×	×	×	×	×	×

Outline

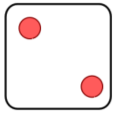
- Uncertainty
- State Estimation
- **Basic Concepts in Probability**
- Bayesian Rule
- Environment Interaction
- Bayes Filter Algorithm
- Summary

Basic Concepts in Probability

- **Random Variables**

A random variable (RV) represents **functional mapping** from a set of **experimental outcomes** (the domain) to **a set of real numbers** (the range).

Example-1: the roll of a die can be viewed as a RV if we map:

Experimental outcomes		A set of real number
Appearance of 	**Mapped to**	Output: 1
Appearance of 	**Mapped to**	Output: 2
Appearance of 	**Mapped to**	Output: 3

and so on…

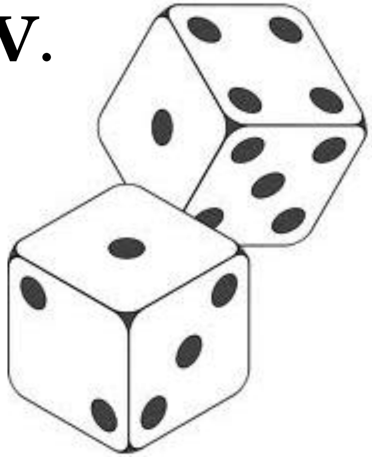
Basic Concepts in Probability

- **Random Variables**

Example-1 (cont'd):

Of course, **after** we throw the die, the value of the die is no longer a RV –it **becomes certain**.

The **outcome** of a particular experiment is **not a RV**.



Basic Concepts in Probability

- **Random Variables**

Example-1 (Cont'd): If we define X as a RV that represents the roll of a die, then the probability that X will be 4 is equal to 1/6.

If we then roll a four, the four is a **realization** of the RV X.

If we then roll the die again and get a three, the three is **another realization** of the RV X. However, the **RV X exists independently of any of its realizations**.

This distinction between a RV and its realization is important for understanding the concept of probability.

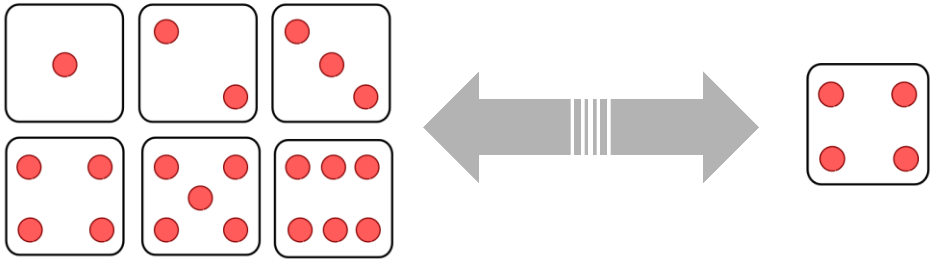
Realizations of a RV are not equal to the RV itself.

Basic Concepts in Probability

- **Random Variables**

Example-1 (Cont'd): When we say that

The probability of **X=4** is equal to **1/6**, that means that there is a **1 out of 6** chances that each **realization** of X will be equal to 4.



$$p(X = 4) = \frac{\text{Possible outcomes favoring event } (X = 4)}{\text{Total number of possible outcomes}} = \frac{1}{6}$$

However, the **RV X** will always be random and will never be equal a specific value.

Basic Concepts in Probability

- **Random Variables**

Example-2: Another standard example of a random variable is that of a **coin flip**, where X can take on the values **head or tail**.

If the space of all values that X can take on is discrete, as is the case if X is the outcome of a coin flip, we write ***p(X=x)*** to denote the probability that the random variable X has value x.

A fair coin is characterized by

p(X = head) = p(X = tail) = 0.5.

Discrete probabilities, sum to one, that is

$$\sum_{x} p(X = x) = 1$$

Basic Concepts in Probability

- **Random Variables:** is a **stochastic variable** that represents the **formal encoding** of one's **beliefs** about the various potential values of a quantity that is **not known with certainty**.

Random Variables

Discrete

- # of detected landmarks.
- A door state (open or closed)
- Space state (occupied, empty or undefined)
- Sensor state (on or off), etc…

Continuous

- Travelled distance by the robot.
- The time taken to reach a goal.
- Location and velocities of moving objects and people.
- Battery level, etc…

Basic Concepts in Probability

- **Probability Distribution Function**

The most fundamental property of a RV X is its probability distribution function (PDF) $F_X(x)$, defined as

$$F_X(x) = p(X \leq x)$$

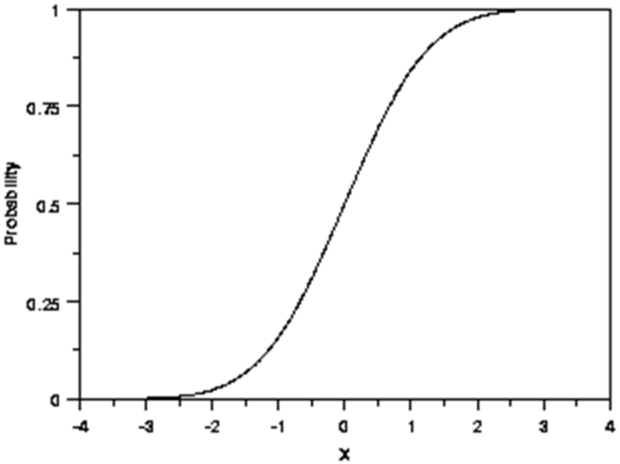
This probability distribution identifies either the probability of each value of a random variable (when the variable is discrete), or the probability of the value falling within a particular interval (when the variable is continuous).

$$F_X(x) \in [0,1]$$

$$F_X(-\infty) = 0 \qquad F_X(\infty) = 1$$

$$F_X(a) \leq F_X(b) \quad \text{if} \quad a \leq b$$

$$F_X(a \prec X \leq b) \leq F_X(b) - F_X(a)$$



Basic Concepts in Probability

- **Probability Density Function**

The probability density function (pdf) $f_X(x)$ is defined as the derivative of the PDF:

$$f_X(x) = \frac{dF_X(x)}{dx}$$

This density function is used to characterize the statistical properties of the value (or the **realization**) of X.

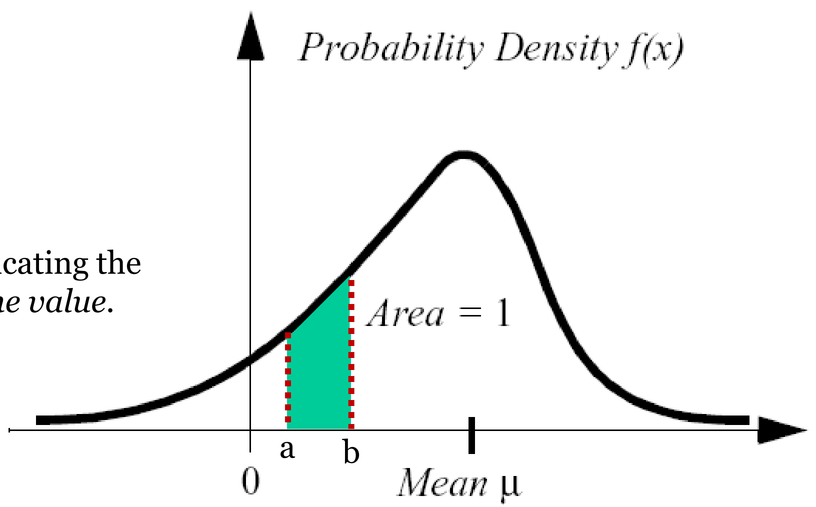
$$F_X(x) = \int_{-\infty}^{x} f_X(z)dz$$

$$f_X(x) \geq 0$$

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

The area under the curve is 1, indicating the complete chance of X having *some value*.

$$p(a \prec x \leq b) = \int_{a}^{b} f_X(x)dx$$



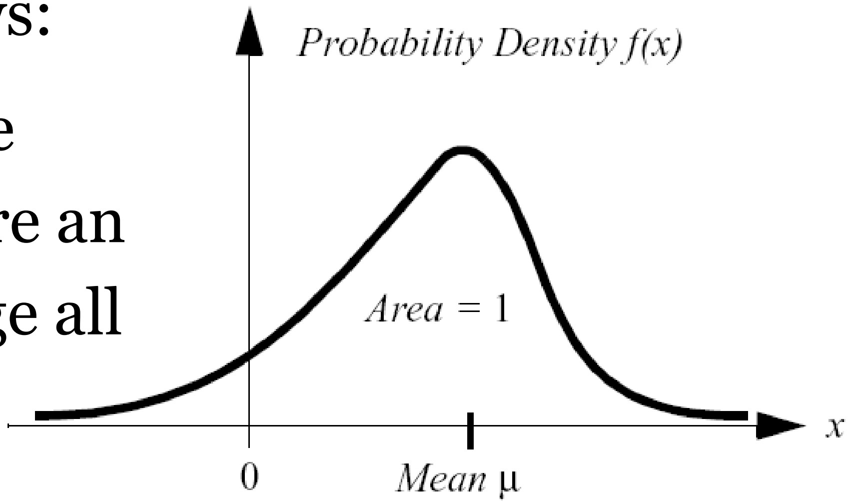
Basic Concepts in Probability

- **Probability Density Function**

The probability density function is a useful way to characterize the possible values of X because it not only captures the range of X but also the comparative probability of different values for X.

Using **pdf** we can quantitatively define the **mean, variance**, and **standard deviation** as follows:

Mean Value μ: is equivalent to the expected value if we were to measure an infinite number of times and average all of the resulting values.



$$\mu = E[X] = \int_{-\infty}^{\infty} x f_X(x) dx$$

→ **μ** represents the weighted average of all possible values of X.

Basic Concepts in Probability

- **Probability Density Function**

Mean Value μ (cont'd):

Example: For a discrete case, when a die is thrown, each of the possible faces 1, 2, 3, 4, 5, 6 (the x_i's) has a probability of 1/6 (the $P(x_i)$'s) of showing.

The **expected value** of the face showing is therefore

$$\mu = E(X) = (1 \times 1/6) + (2 \times 1/6) + (3 \times 1/6) + (4 \times 1/6) + (5 \times 1/6) + (6 \times 1/6) = 3.5$$

Notice that, in this case, E(X) is 3.5, which is not a possible value of X.

Expected value of a RV X is the average value over a large number of experiments. It is called **expectation**, the **mean**, or the **average** of RV.

Basic Concepts in Probability

- **Probability Density Function**

Variance σ²: is a measure of how much we expect the RV to vary from its mean.

$$Var(X) = \sigma^2 = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)dx$$

$$\sigma^2 = E[(X-\mu)^2] = E[X^2 - 2X\mu + \mu^2]$$

$$= E(X^2) - 2\mu^2 + \mu^2$$

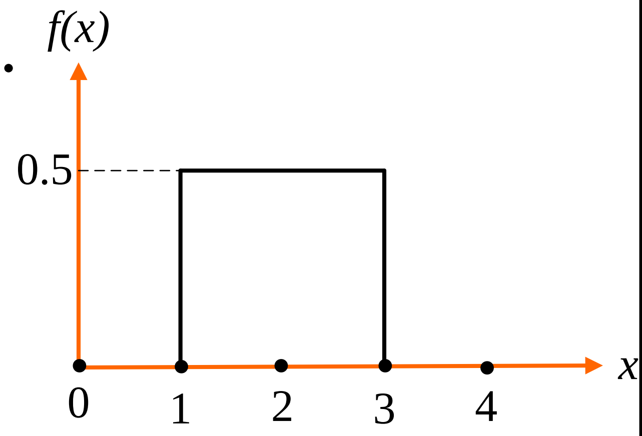
$$= E(X^2) - \mu^2$$

Standard deviation σ: is square root of the variance. Variance and standard deviation will play important roles in our characterization of the error of a single sensor as well as the error of a model generated by combining multiple sensor readings.

Basic Concepts in Probability

- **Probability Density Function**

Example: Given the following **pdf** of a RV, calculate the expected value (the mean) and the variance.



$$f_x(x) = \begin{cases} 1/2 & x \in [1,3] \\ 0 & \text{otherwise} \end{cases}$$

The mean is computed as follows:

$$\mu = \int_{-\infty}^{\infty} x f(x) dx = \int_{1}^{3} \frac{1}{2} x dx = 2$$

The variance is computed as follows:

$$Var(X) = \sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x) dx$$

$$= \int_{1}^{3} \frac{1}{2} (x-2)^2 dx = \frac{1}{3}$$

Basic Concepts in Probability

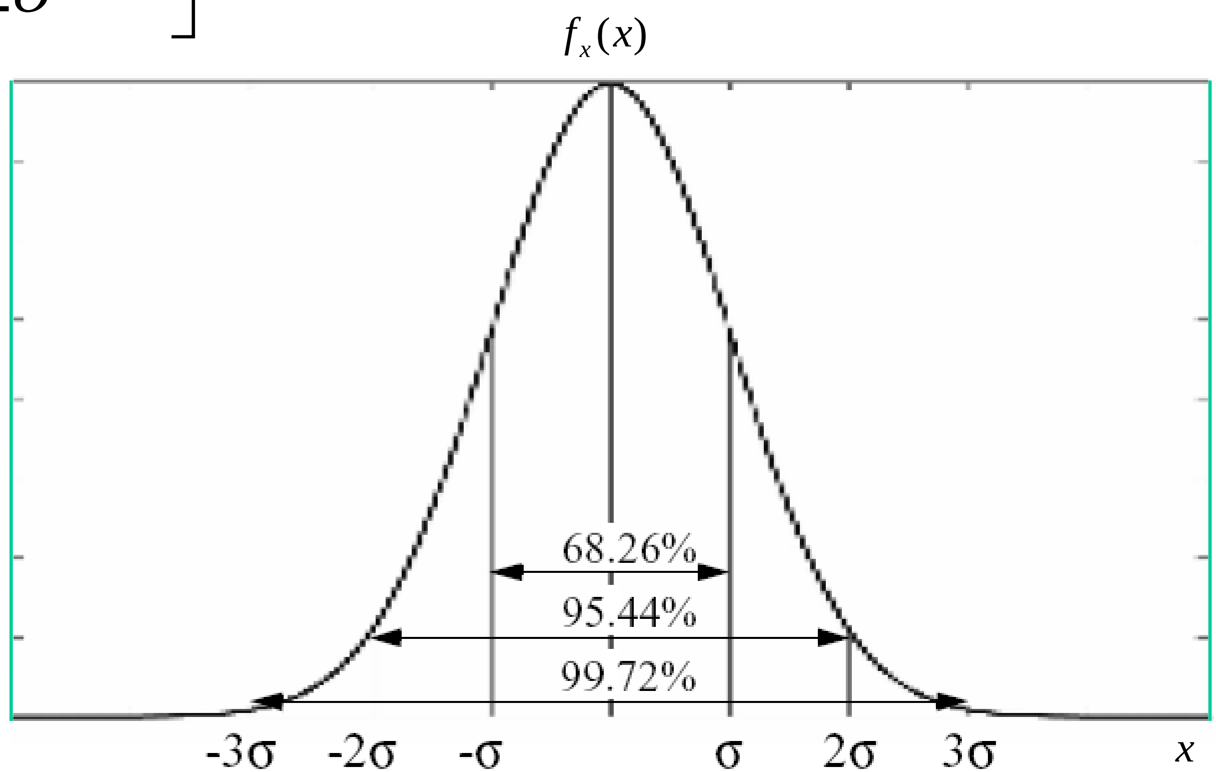
- **Gaussian Distribution**

A RV is called Gaussian or normal if its **pdf** is given by

$$f_x(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right]$$

We use the following notation to indicate that X is a Gaussian RV

$$X \sim N(\mu, \sigma^2)$$



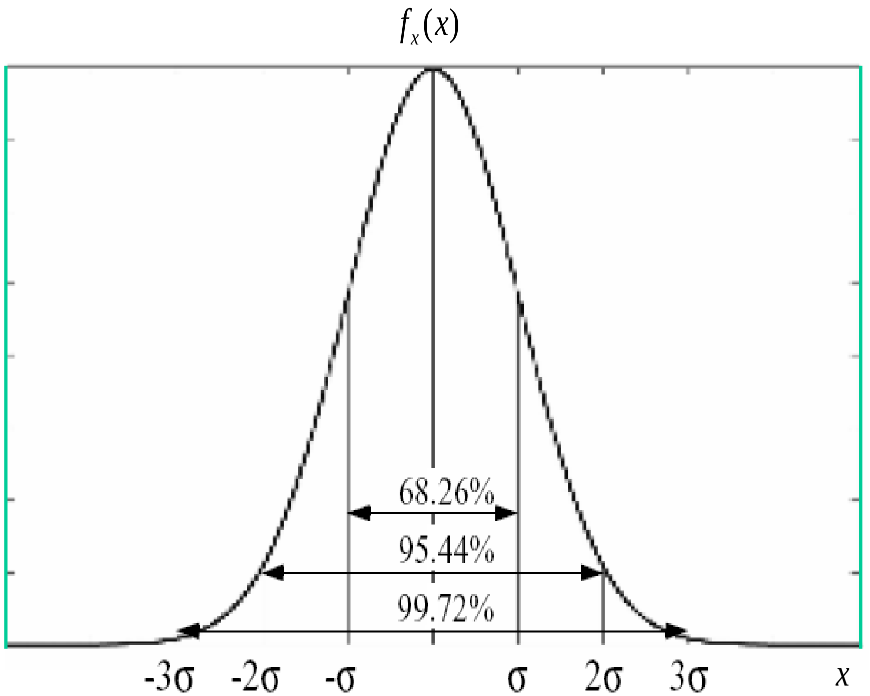
The Gaussian function with μ=0 and σ=1. The value **2σ** is often refereed to as the **signal quality**; 95.44% of the values are falling within ± 2σ.

Basic Concepts in Probability

- **Gaussian Distribution**

$$f_x(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right]$$

Only two parameters required to fully specify a particular Gaussian are its mean μ, and its standard deviation σ.

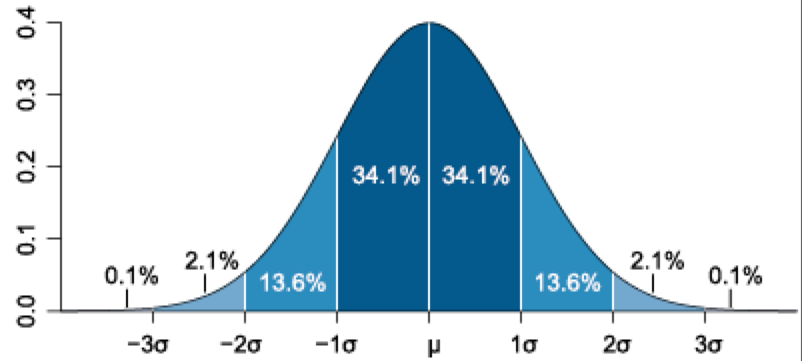


Gaussian distribution is also **unimodal**, with a single peak that reaches a maximum at μ (necessary for any symmetric, unimodal distribution).

Basic Concepts in Probability

- **Gaussian Distribution**

$$f_x(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right]$$

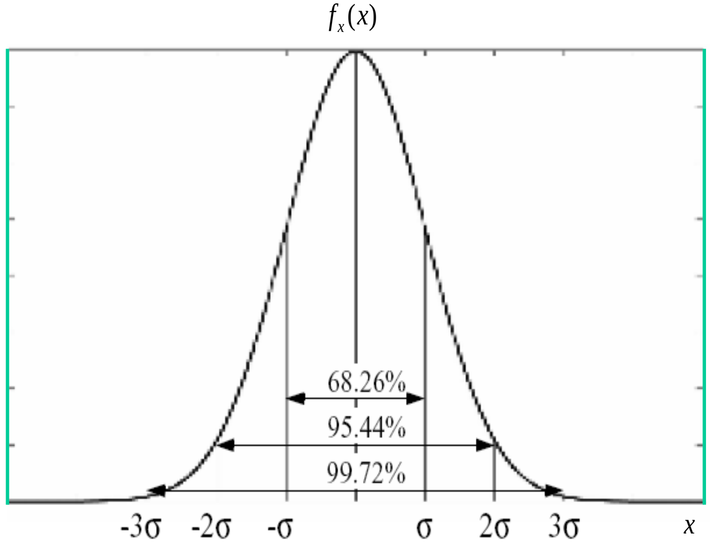


This Gaussian distribution, also called the normal distribution, is used across engineering disciplines when a **well-behaved error model** is required for a random variable for which no error model of greater felicity has been discovered.

Basic Concepts in Probability

- **Gaussian Distribution**

Q: Suppose that a RV X is modeled as a Gaussian. How does one identify the chance that the value of X is within one standard deviation of μ?



A: In practice, this requires integration of , the Gaussian function to compute the area under a portion of the curve:

$$p(-\sigma \prec x \leq \sigma) = \int_{-\sigma}^{\sigma} f_X(x)dx$$

Unfortunately, there is **no closed-form solution for this integral**, and so the common technique is to use a **Gaussian cumulative probability table**.

Basic Concepts in Probability

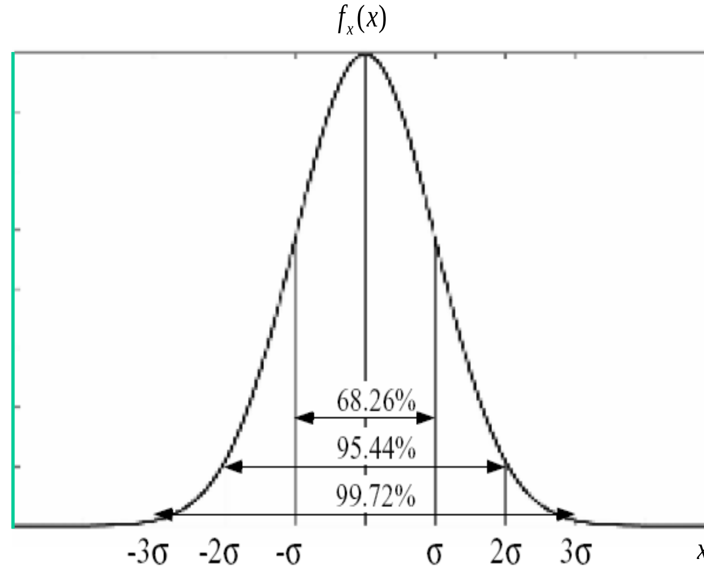
- **Gaussian Distribution**

Using **Gaussian cumulative probability table**, one can compute the area under the curve for various values ranges of X:

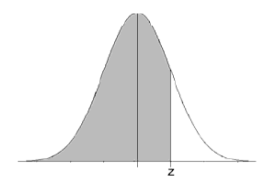
$$p(\mu - \sigma \prec X \leq \mu + \sigma) = 0.68;$$

$$p(\mu - 2\sigma \prec X \leq \mu + 2\sigma) = 0.95;$$

$$p(\mu - 3\sigma \prec X \leq \mu + 3\sigma) = 0.997;$$



Standard Normal Cumulative Probability Table



Cumulative probabilities for POSITIVE z-values are shown in the following table:

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319

Basic Concepts in Probability

- **Multivariate Normal Distribution**

Often, x will be a **multi-dimensional vector**. Normal distributions over vectors are called multivariate.

Multivariate normal distributions are characterized by density functions of the following form:

$$p(x) = \det(2\pi\Sigma)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right\}$$

Here μ is the mean vector and a (positive semidefinite) symmetric matrix called Σ or **covariance matrix**.

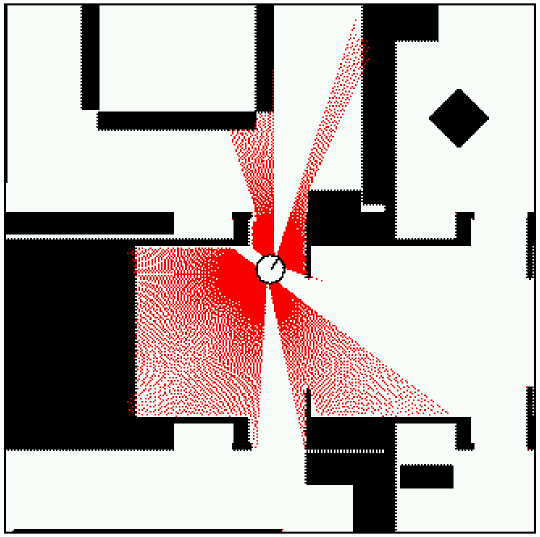
Basic Concepts in Probability

- **Joint Distribution**

The joint distribution of two random variables X and Y is given by: $p(x,y) = p(x \cap y) = p(X = x \text{ and } Y = y)$

If X and Y are **independent** (the particular value of one has no bearing on the particular value of the other), we have $p(x,y) = p(x)\,p(y)$

Example: A mobile robot's laser rangefinder may be used to measure the position of a feature on the robot's right and, later, another feature on the robot's left. The position of each feature in the real world may be treated as random variables, X and Y.



Basic Concepts in Probability

- **Conditional Probability**

Often, random variables **carry information** about other random variables (**dependent** random variables).

Suppose we already know that Y 's value is y, and we would like to know the probability that X's value is x conditioned on that fact.

Such a **conditional probability** will be denoted:

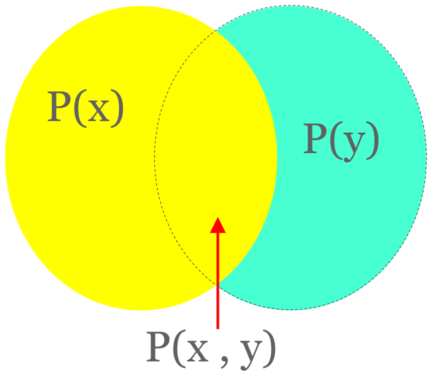
$$p(x \mid y) = p(X = x \mid Y = y)$$

Basic Concepts in Probability

- **Conditional Probability**

If p(y)>0, then the conditional probability is defined by:

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



Venn Diagram

If X and Y are **independent**, we have

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

In other words, if X and Y are independent, **Y tells us nothing about the value of X**. There is no advantage of knowing Y if our interest pertains to knowing X.

Independence, and its generalization known as **conditional independence**.

Basic Concepts in Probability

- **Conditional Probability**

Example-1: The probability that it is Thursday and that a student is absent is 0.03.

Since there are 5 university days in a week, the probability that it is Thursday is 0.2.

What is the probability that a student is absent given that today is Thursday?

$$p(\text{Absent} \mid \text{Thursday}) = \frac{p(\text{Thursday} \cap \text{Absent})}{p(\text{Thursday})}$$

$$= \frac{0.03}{0.2} = 0.15 = 15\%$$

Basic Concepts in Probability

- **Conditional Probability**

Example-2: An inspection robot is used to inspect some work pieces using two tests.

25% of the pieces passed both tests and 42% of the pieces passed the first test.

What percent of those which passed the first test also passed the second test?

Solution:

$$p(\text{SecondTest} \mid \text{FirstTest}) = \frac{p(\text{FirstTest and SecondTest})}{p(\text{FirstTest})}$$

$$= \frac{0.25}{0.42} = 0.60 = 60\%$$

Basic Concepts in Probability

- **Theorem of Total Probability**

$$p(x) = \sum_y p(x \mid y)\, p(y) \qquad \text{- Discrete case}$$

$$p(x) = \int p(x \mid y)\, p(y)\, dy \qquad \text{- Continous case}$$

Basic Concepts in Probability

- **Theorem of Total Probability**

Example: An insurance company rents 40% of the cars for its customers from agency-I and 60% from agency-II.

If 6% of the cars from agency-I and 5% of the cars from agency-II break down.

What is the probability that a car rented by this company breaks down?

Solution:

◇ Let x represents the event "car breaks down"

◇ Let y_1, y_2 represents the event "agency-I" and "agency-II" resp.

Basic Concepts in Probability

- **Theorem of Total Probability**

Example (cont'd):

- ◇ $P(y_1)=0.4$
- ◇ $P(y_2)=0.6$
- ◇ $P(x|y_1)=0.06$
- ◇ $P(x|y_2)=0.05$

$$p(\mathrm{x}) = \sum_y p(x \mid y)\, p(y)$$

$$= p(x \mid y_1)\, p(y_1) + p(x \mid y_2)\, p(y_2)$$

$$= (0.06)(0.4) + (.05)(0.6)$$

$$= 0.054 \text{ or } 5.4\%$$

Outline

- Uncertainty
- State Estimation
- Basic Concepts in Probability
- **Bayesian Rule**
- Environment Interaction
- Bayes Filter Algorithm
- Summary

Bayesian Rule

- **Conditional probability**

Assume that there are two dependent events:

Event x	Event z	
	True	False
True	0.1	0.3
False	0.4	0.2

The probability of both events occurring can be expressed as:

$$p(x,z) = p(z \mid x)\,p(x)$$

Or

$$p(x,z) = p(x \mid z)\,p(z)$$

Equating the RHS

$$p(x \mid z) = \frac{p(z \mid x)\,p(x)}{p(z)} = \frac{\text{likelihood} \times \text{prior}}{\text{normalization factor}}$$

Bayesian Rule

Thomas Bayes
English mathematician
(1702 – 1761)

Bayes rule is a rule that relates conditionals of the type p(x | z) to their "**inverse**," p(z | x).

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

where **x**=state,

z=observation or sensor measurement or data,

p(z) is the evidence

p(x) is the prior estimate or belief of the state (before taking the measurement)

p(z|x) is the likelihood or measurement probability

p(x|z) is the posterior estimate or belief of the state (after the measurement has been taken).

Bayesian Rule

- **A priori**
  - ◊ In philosophy, a priori is knowledge, justifications or arguments that is independent of experience.
  - ◊ In statistics, a priori knowledge refers to prior knowledge about a population, rather than that estimated by recent observation.
  - ◊ A priori is derived by logic, without observed facts .
  - ◊ A priori is deductive reasoning - involving inferences from general principles.

Deductive Reasoning

All men are mortal. (major premise)

Socrates is a man. (minor premise)

Socrates is mortal. (conclusion)

[Wikipedia]

Bayesian Rule

- **A Posteriori**
  - ◇ In philosophy, a posteriori is knowledge, justifications or arguments that is dependent on experience or empirical evidence.
  - ◇ In Logic, a posteriori is the process of reasoning from effect to cause, based upon observation.
  - ◇ A posteriori is inductive reasoning; proceeding from particular facts to a general conclusion.

Inductive Reasoning

3+5=8 and eight is an even number.

Therefore, an odd number added to another odd number will result in an even number.

[Wikipedia]

Bayesian Rule

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

$$\text{posterior belief} = \frac{\text{likelihood.prior belief}}{\text{evidence}}$$

- It is common to think of Bayes rule in terms of updating our belief about a **hypothesis *x*** in the light of new evidence ***z***.
- Specifically, our ***posterior* belief p(*x*|*z*)** is calculated by multiplying our ***prior* belief p(*x*)** by the ***likelihood* p(*z*|*x*)** that ***z*** will occur if ***x*** is true.

Bayesian Rule

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

$$\text{posterior belief} = \frac{\text{likelihood.prior belief}}{\text{evidence}}$$

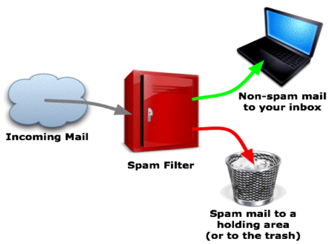
- The power of Bayes' rule is that in many situations where we want to compute **p(x|z)** it turns out that it is difficult to do so directly, yet we might have direct information about **P(z|x)**.
- Bayes' rule enables us to **compute p(*x*|*z*) in terms of p(*z*|*x*)**.

Bayesian Rule

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

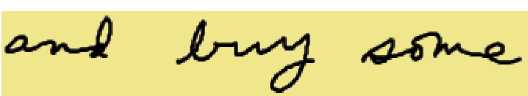
Bayes' rule enables us to **compute p(*x*|*z*) in terms of p(*z*|*x*)**.

- **Spam Filtering:**

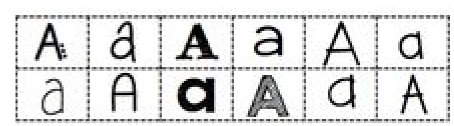


$$p(spam \mid word) \Rightarrow p(word \mid spam)$$

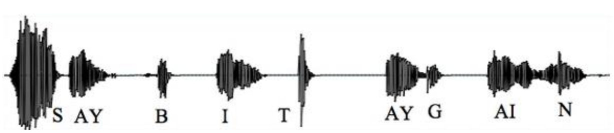
- **Handwriting Recognition:**



$$p(letter \mid handwriting) \Rightarrow p(handwriting \mid letter)$$



- **Speech Recognition:**



$$p(sound \mid phoneme) \Rightarrow p(phoneme \mid sound)$$

Bayesian Rule

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

$$\text{posterior belief} = \frac{\text{likelihood.prior belief}}{\text{evidence}}$$

- An important observation is that the denominator of Bayes rule, **p(z)**, **does not depend on x**. For this reason, **p(z)$^{-1}$** is often written as a **normalizer** variable, and generically denoted η.

$$p(x \mid z) = \eta . p(z \mid x)\, p(x)$$

Bayesian Rule

- **Existence of God**
  - ◇ Here is Bayes' rule applied to the probability that God exists, given some data we observe about the universe.
  - ◇ This data could be the fact that the physical world seems to obey an orderly set of rules, it could be a miraculous experience, or anything else that might affect our belief in God.
  - ◇ The probability that God exists, given some data about the universe is equal to the probability of observing that data if God did exist, multiplied by our prior probability that God exists (our belief before observing the data), divided by the overall probability of observing the data.

Bayesian Rule

- **Existence of God**
  - ◇ In mathematical form:

$$p(\text{God exists given data}) = \frac{p(\text{data given God exists}) \times p(\text{God exists})}{p(\text{data})}$$

  - ◇ **So what does this mean?**
  - ◇ It means the higher somebody's prior probability that God exists, the higher the probability they will assign to God existing even after observing the data, and vice versa, somebody with a very low prior probability of God existing will tend to give a lower probability to God existing, even after observing the same data.

Bayesian Rule

- **Existence of God**

$$p(\text{God exists given data}) = \frac{p(\text{data given God exists}) \times p(\text{God exists})}{p(\text{data})}$$

◇ In fact, if you look carefully, somebody who has a zero prior probability of God existing (an **Atheist**) will continue to be an atheist, no matter how much evidence he or she is confronted with, since zero times anything is … zero!

Bayesian Rule

- **Example-1: Spam Filtering**
  - ◇ Event A: The message is spam.
  - ◇ Test X: The message contains certain words (X)

$$p(\mathrm{A}\,|\,\mathrm{X}) = \frac{p(\mathrm{X}\,|\,\mathrm{A})\,p(\mathrm{A})}{p(\mathrm{X})}$$

  - ◇ Bayesian filtering allows us to predict the chance a message is really spam given the "test results" (the presence of certain words). Clearly, words like "**Lucky you**" have a higher chance of appearing in spam messages than in normal ones.
  - ◇ Spam filtering based on a **blacklist** is flawed — it's too restrictive and false positives are too great. But Bayesian filtering gives us a middle ground — we use *probabilities*.

Bayesian Rule

- **Example-2: Cancer Test**

Given:

◇ Probability of having cancer p(C)=0.01

◇ p(Pos|C)=0.9, i.e. 90% test is positive if you have C [**Sensitivity**]

◇ p(Neg|¬C]=0.9, i.e. 90% test is negative if you don't have C [**Specificity**]

◇ Test is positive

Required:

What is the probability of having cancer?

Bayesian Rule

- **Example-2: Cancer Test**
  - **Prior**
    - $p(C)=0.01$, $p(\neg C)=0.99$
    - $p(Pos|C)=0.9$, $p(Pos|\neg C)=0.1$
    - $p(Neg|\neg C)=0.9$, $p(Neg|C)=0.1$
  - **Joint Probabilities**
    - $p(C,Pos)=p(C).\ p(Pos|C)=0.009$
    - $p(\neg C,Pos)=p(\neg C).\ p(Pos|\neg C)=0.099$

      } Don't add to 1
    - $p(Pos)=p(C,Pos)+ p(\neg C,Pos)=0.0108$

Bayesian Rule

- **Example-2: Cancer Test**
  - **Posterior Probabilities**

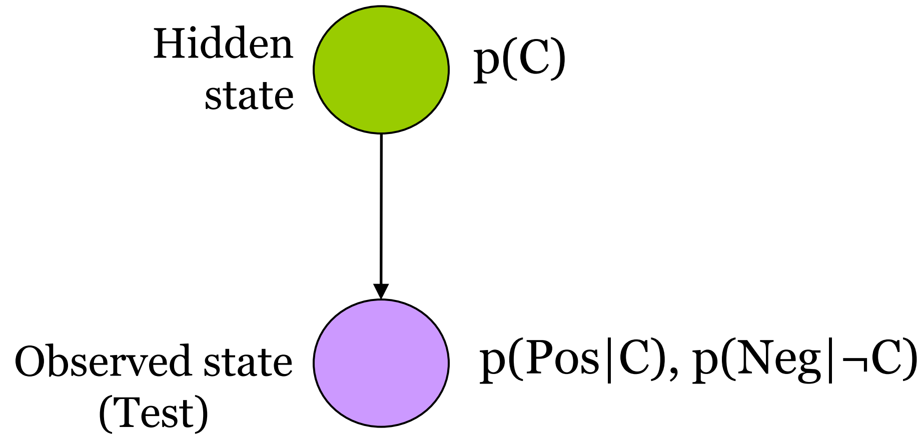
$$p(\mathrm{C} \mid \mathrm{Pos}) = \frac{p(\mathrm{P}os \mid \mathrm{C})\, p(\mathrm{C})}{p(\mathrm{P}os)} = 0.083$$

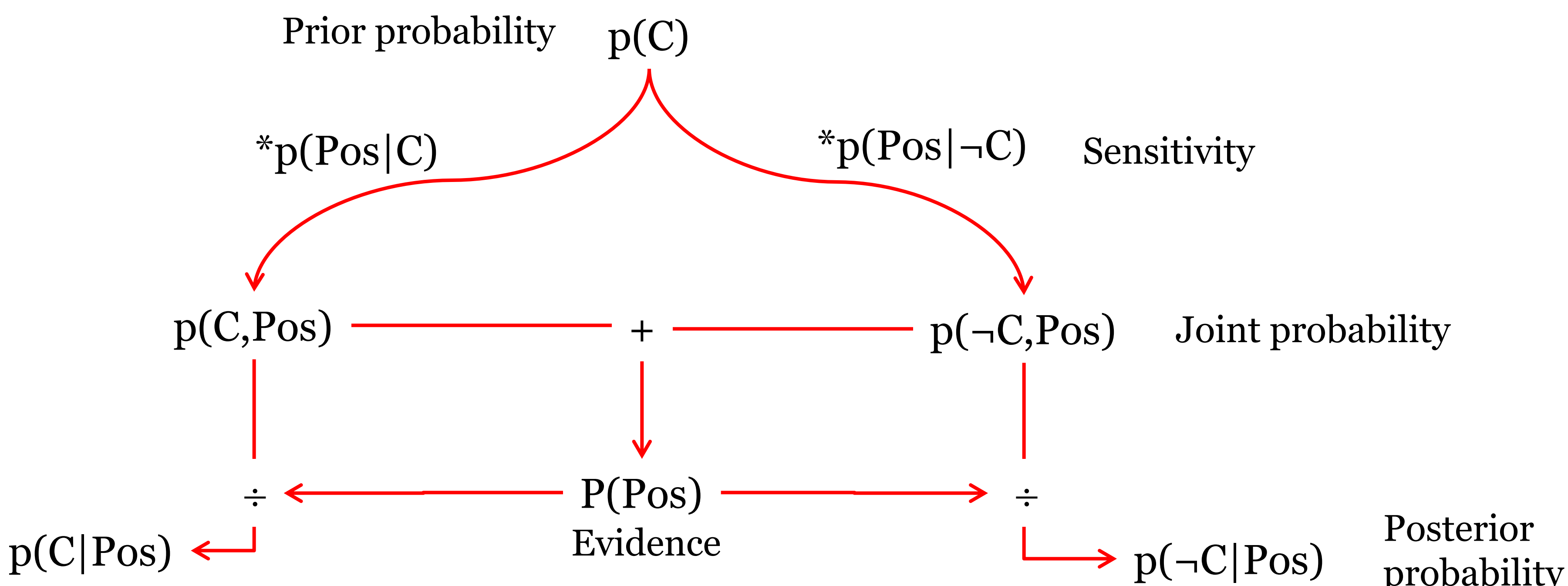
$$p(\neg\mathrm{C} \mid \mathrm{Pos}) = \frac{p(\mathrm{P}os \mid \neg\mathrm{C})\, p(\neg\mathrm{C})}{p(\mathrm{P}os)} = 0.9167$$

Add to 1

Bayesian Rule

- **Example-2: Cancer Test**
  - ◇ **Prior:** p(C)
  - ◇ **Sensitivity:** p(Pos|C)
  - ◇ **Specificity:** p(Neg|¬C)





Bayesian Rule

- **Example-3: Disease Diagnosis**
  - ◇ Disease ∈ {malaria, cold, flu}; Symptom = fever.
  - ◇ Must compute **p(D | fever)** to prescribe treatment
  - ◇ Why not assess this quantity directly?
    - **p(mal | fever)** is not natural to assess;

      p(fever | mal) reflects the underlying "causal" mechanism
    - **p(mal | fever)** is not "stable": a malaria epidemic changes this quantity (for example).
  - ◇ So we use **Bayes rule**:
    - **p(mal | fever) = p(fever | mal) p(mal) / p(fever)**

Bayesian Rule

- **Example-3: Disease Diagnosis**

  p(mal | fever) = p(fever | mal) p(mal) / p(fever)

  ◊ What about **p(mal)**?

  - This is the **prior probability** of Malaria, i.e., before you exhibited a fever, and

    with it we can account for other factors, e.g., a **malaria epidemic**, or recent travel to a **malaria risk zone**.

Bayesian Rule

- **Example-3: Disease Diagnosis**

  p(mal | fever) = p(fever | mal) p(mal) / p(fever)

  ◊ What about **p(fever)**?

  - We compute **p** of each disease given fever, i.e., solve the same problem for each possible clause of fever:

    $$p(fev) = p(mal \cap fev) + p(cold \cap fev) + p(flu \cap fev)$$

    P(fev) is called **evidence** and it's a **normalizing factor** to guarantee that:

    $$p(mal|fever) + p(cold|fever) + p(flu|fever) = 1.$$

Bayesian Rule

- **Example-4: Test Analysis**

Given:

◇ 1% of the population is ill: $p(i) = 0.01$

◇ Given an ill person, the test is positive in 90% of the cases:

$p(t \mid i) = 0.9$

◇ Given a person that is not ill, the test is positive in 20% of the cases: $p(t \mid \neg i) = 0.2$

Required:

What is the probability of being ill given a positive test $p(i|t)$?

Bayesian Rule

- **Example-4: Test Analysis**

Solution:

$$p(i) = 0.01$$

$$p(\neg i) = 0.99$$

$$p(t \mid i) = 0.9$$

$$p(t \mid \neg i) = 0.2$$

$$p(i \mid t) = \frac{p(t \mid i)\, p(i)}{p(t)} = \frac{p(t \mid i)\, p(i)}{\sum_i p(t \mid i)\, p(i)} = \frac{p(t \mid i)\, p(i)}{p(t \mid i)\, p(i) + p(t \mid \neg i)\, p(\neg i)}$$

$$= \frac{0.9 \times 0.01}{0.9 \times 0.01 + 0.2 \times 0.99} = \frac{0.009}{0.207} \approx 0.0043 \approx 4\%$$

The probability of being ill given a positive test is only 4%!

Bayesian Rule

- **Example-5: Cancer Diagnosis**
  - ◇ Suppose that we are interested in diagnosing cancer in patients who visit a chest clinic.
  - ◇ Let **x** represents the event "**Person has cancer**"
  - ◇ Let **z** represents the event "**Person is a smoker**"
  - ◇ We know the probability of the prior event **p(*x*)=0.1** on the basis of past data (**10% of patients entering the clinic turn out to have cancer**).
  - ◇ We want to compute the probability of the **posterior event p(*x*|*z*)**. It is difficult to find this out directly.

Bayesian Rule

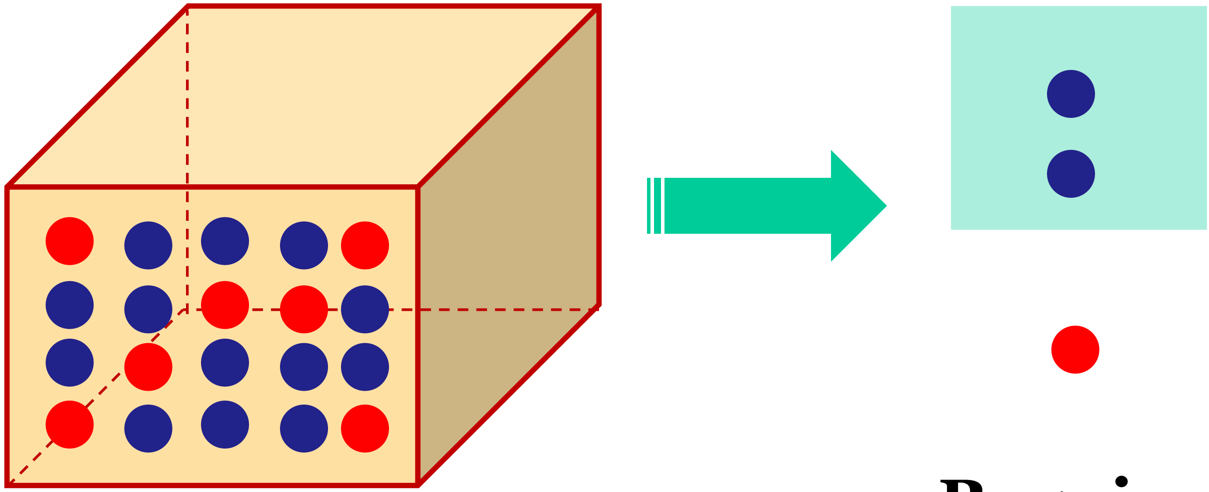
- **Example-5: Cancer Diagnosis**
  - ◇ We are likely to know p(*z*) by considering the percentage of patients who smoke – suppose **p(*z*)=0.5**.
  - ◇ We are also likely to know **p(*z*|*x*)** by checking from our record the proportion of smokers among those diagnosed. Suppose **p(*z*|*x*)=0.8**.
  - ◇ We can now use **Bayes' rule** to compute:

$$p(x \mid z) = \frac{p(z \mid x)\, p(x)}{p(z)} = \frac{0.8 \times 0.1}{0.5} = 0.16$$

  - ◇ Thus, in the light of evidence that the person is a smoker we revise our prior probability from 0.1 to a posterior probability of 0.16. This is a significance increase, but it is still unlikely that the person has cancer.

Bayesian Rule

- **Example-6: Ball Game**
  - A box contains seven red and thirteen blue balls. Two balls are selected at random and are discarded without their colors being seen. If a third ball is drawn randomly and observed to be red, what is the probability that both of the discarded balls were blue?



Required: p(BB|R)

Bayesian Rule

- **Example-6: Ball Game**

$$p(BB \mid R) = \frac{p(R \mid BB)\, p(BB)}{p(R)}$$

$$= \frac{p(R \mid BB)\, p(BB)}{p(R \mid BB)\, p(BB) + p(R \mid BR)\, p(BR) + p(R \mid RR)\, p(RR)}$$

$$p(R \mid BB) = \frac{7}{18} \qquad p(R \mid BR) = \frac{6}{18}$$

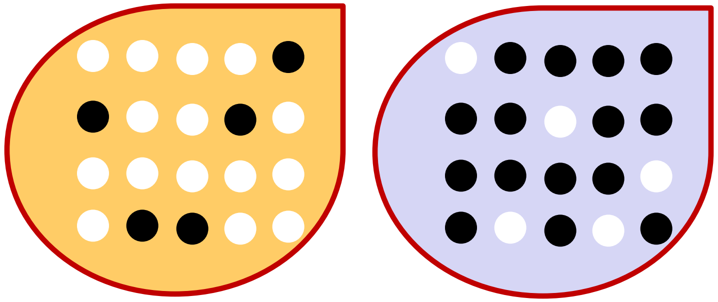
$$p(BB) = \frac{13}{20} \times \frac{12}{19} = \frac{39}{95} \qquad p(BR) = \frac{13}{20} \times \frac{7}{19} = \frac{91}{380}$$

$$p(R \mid RR) = \frac{5}{18}$$

$$p(RR) = \frac{7}{20} \times \frac{6}{19} = \frac{21}{190} \qquad p(BB \mid R) = \frac{0.159}{0.269} \approx 0.59 = 59\%$$

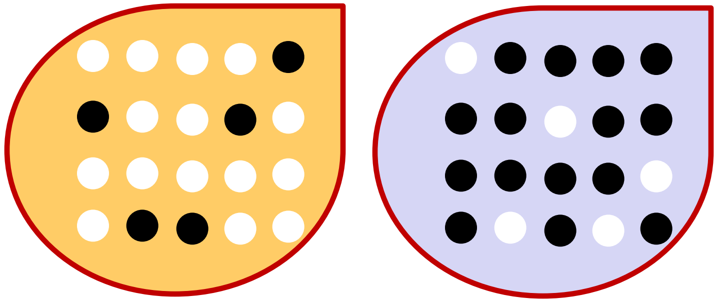
Bayesian Rule

- **Example-7: Two Bags Game**



  - ◇ Suppose that we have two bags each containing black and white balls.
  - ◇ One bag contains three times as many white balls as blacks. The other bag contains three times as many black balls as white.
  - ◇ Suppose we choose one of these bags at random.
  - ◇ For this bag we select five balls at random, replacing each ball after it has been selected.
  - ◇ The result is that we find 4 white balls and one black.
  - ◇ What is the probability that we were using the bag with mainly white balls?

Bayesian Rule



- **Example-7: Two Bags Game**
  - ◇ **Given:**

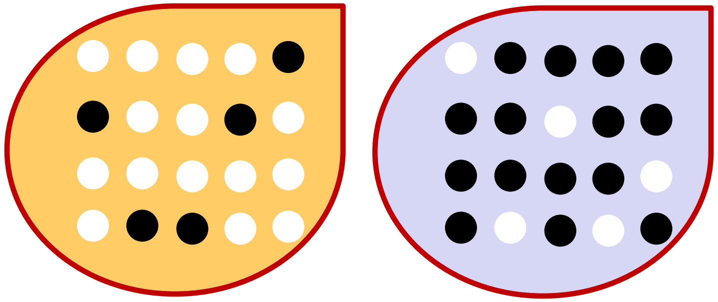
    A: the random variable "**bag chosen**"

    then **$A=\{a_1,a_2\}$** where **a_1** represents "**bag with mostly white balls**" and **a_2** represents "**bag with mostly black balls**" .
  - ◇ We know that **$p(a_1)=p(a_2)=1/2$** since we choose the bag at random.
  - ◇ Let **B** be the event "**4 white balls and one black ball chosen from 5 selections**".
  - ◇ **Required:**

    Calculate $p(a_1|B)$

Bayesian Rule

- **Example-7: Two Bags Game**



◇ From Bayes rule

$$p(a_1 \mid B) = \frac{p(B \mid a_1)\, p(a_1)}{p(B)}$$

$$p(B) = p(B \mid a_1).p(a_1) + p(B \mid a_2).p(a_2)$$

◇ Now, for the bag with **mostly white balls $\mathbf{a_1}$:**

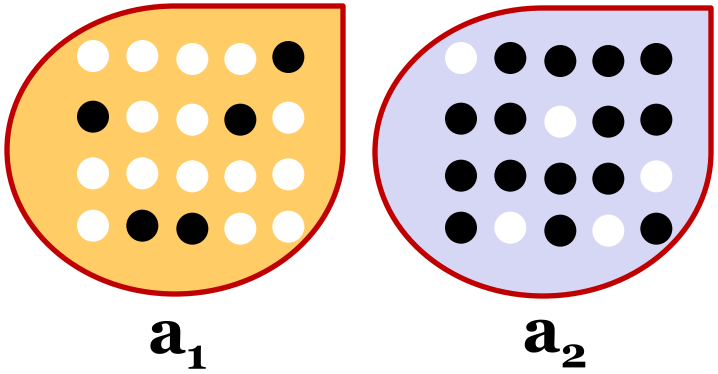
The probability of a ball being white is **¾** and the probability of a ball being black is **¼**.

◇ Now, for the bag with **mostly black balls $\mathbf{a_2}$:**

The probability of a ball being white is **¼** and the probability of a ball being black is **¾**.

Bayesian Rule

- **Example-7: Two Bags Game**
  - ◇ Using **Binomial probability**



The probability of getting exactly ***k*** **successes** in ***n*** **trials** is

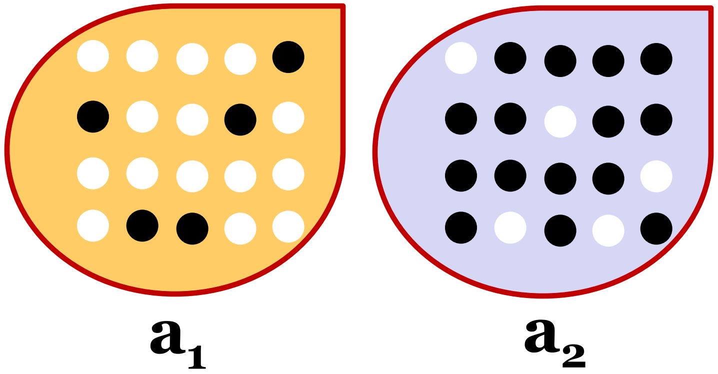
$$p(X=k)=\begin{bmatrix} n \\ k \end{bmatrix} p^k (1-p)^{(n-k)}$$

$$p(B\,|\,a_1)=\begin{bmatrix} 5 \\ 4 \end{bmatrix}\left(\frac{3}{4}\right)^4\left(\frac{1}{4}\right)^1=\frac{5!}{4!(5-4)!}.\left(\frac{3}{4}\right)^4.\left(\frac{1}{4}\right)^1=\frac{405}{1024}$$

Bayesian Rule

- **Example-7: Two Bags Game**
  - ◇ Similarly

$$p(B \mid a_2) = \begin{bmatrix} 5 \\ 4 \end{bmatrix} \left(\frac{1}{4}\right)^4 \left(\frac{3}{4}\right)^1 = \frac{15}{1024}$$



  - ◇ hence

$$p(a_1 \mid B) = \frac{405/1024}{405/1024 + 15/1024} = \frac{405}{420} = 0.964$$

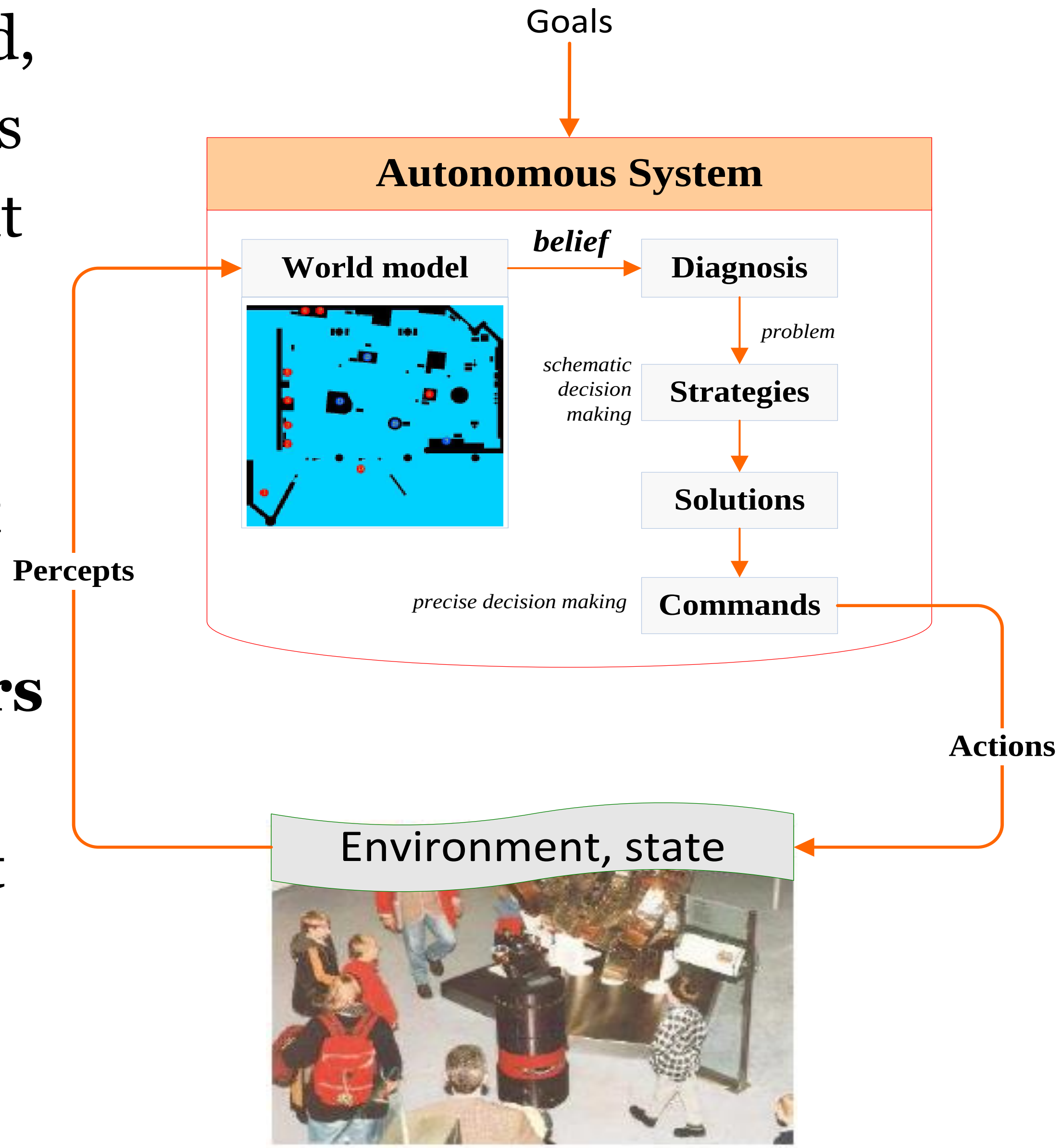
For more see, the big essay on Bayes' Theorem: http://yudkowsky.net/rational/bayes

Outline

- Uncertainty
- State Estimation
- Basic Concepts in Probability
- Bayesian Rule
- **Environment Interaction**
- Bayes Filter Algorithm
- Summary

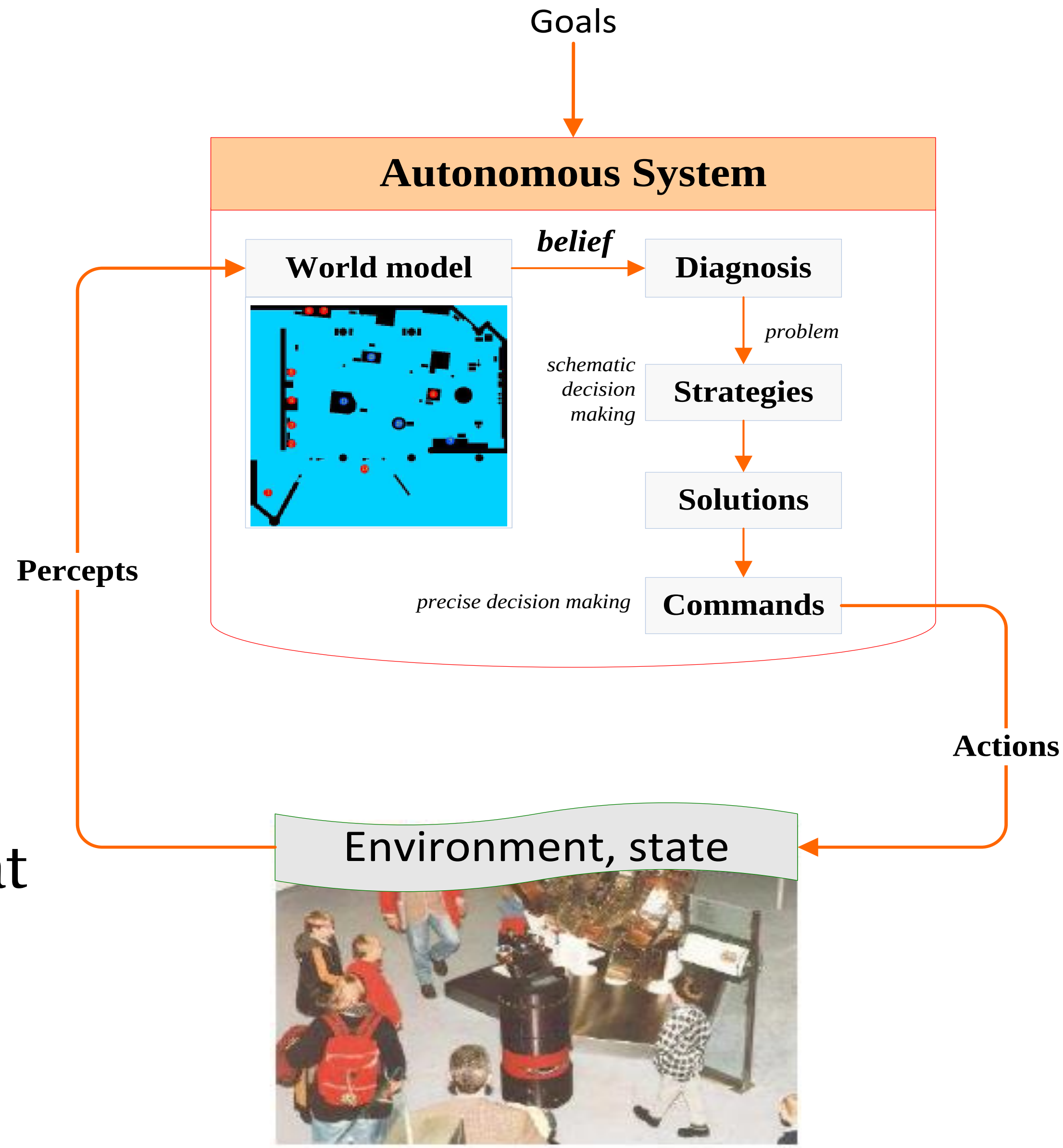
Environment Interaction

- The environment, or world, of an intelligent machine is a **dynamical system** that possesses internal state.
- Intelligent machine can acquire information about its environment using its sensors. However, **sensors are noisy**, and there are usually many **things that cannot be sensed directly**.



Environment Interaction

- As a consequence, the robot maintains an **internal belief** with regards to the state of its environment.
- The robot can also influence its environment through its **actuators**. However, the effect of doing so is often somewhat **unpredictable**.



Environment Interaction

- **Sensor Measurements or observations or percepts**
  - ◇ **Measurement data:** provides information about a **momentary state** of the environment. Examples of measurement data include **camera images**, **range scans**, and so on.
  - ◇ Typically, sensor **measurements arrive with some delay**. Hence they provide information about the state a few moments ago.
  - ◇ For most parts, we will simply ignore small **timing effects** (e.g., most ladar sensors scan environments sequentially at very high speeds, but we will simply assume the measurement corresponds to a specific point in time).

Environment Interaction

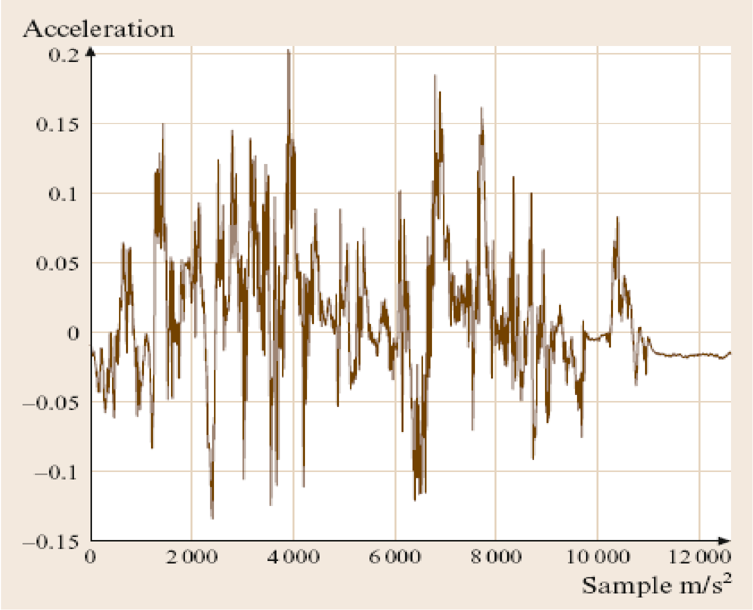
- **Sensor Measurements or observations or percepts**
  - ◇ **Measurement data:**

    The measurement data at time t will be denoted $\mathbf{z_t}$

    Assuming that the robot takes exactly one measurement at a time. The set of all measurements acquired from time t_1 to time t_2 for $t_1 \leq t_2$ is:

$$Z_{t_1:t_2} = Z_{t_1}, Z_{t_1+1}, Z_{t_1+2}, \ldots, Z_{t_2}$$

[5]



Example data from an IMU unit for driving on an unpaved road

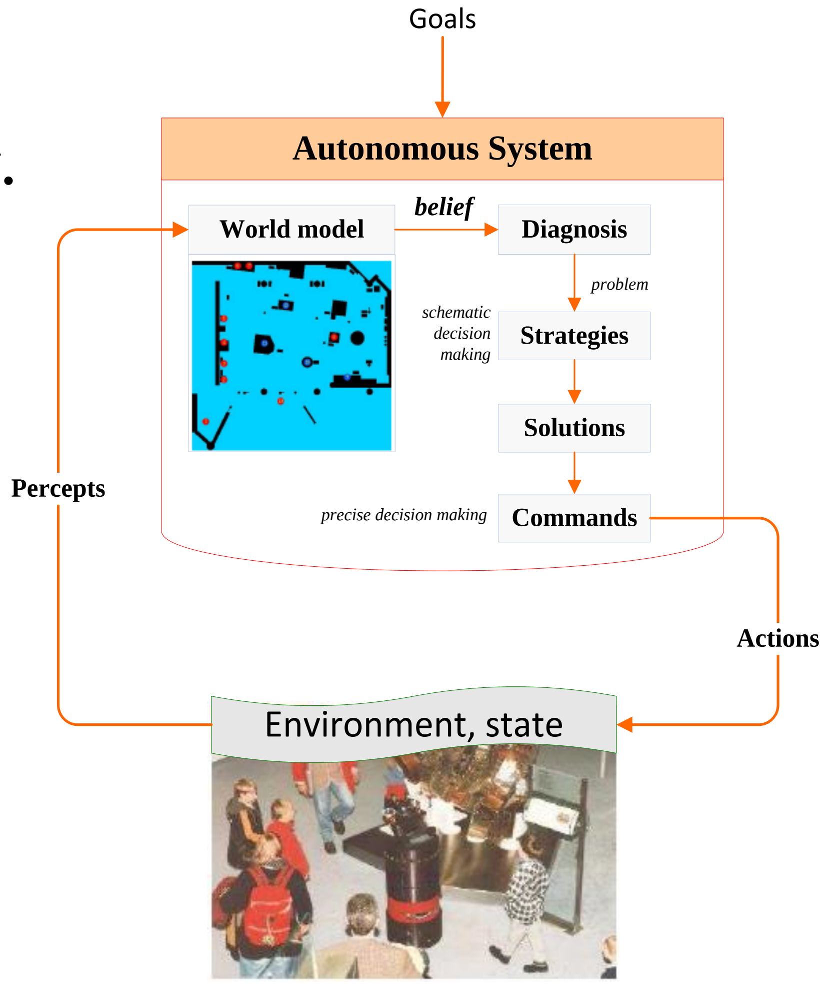
Environment Interaction

- **Control Actions**
  - ◇ Control actions **change the state of the world**. They do so by actively asserting forces on the robot's environment.
  - ◇ Examples of control actions include **robot motion** and the **manipulation of objects**.
  - ◇ In practice, the robot continuously executes controls and measurements are made concurrently.
  - ◇ Control data will be denotes u_t.
  - ◇ As before, we will denote sequences of control data by:

$$u_{t_1:t_2} = u_{t_1}, u_{t_1+1}, u_{t_1+2}, \ldots, u_{t_2}$$

Environment Interaction

- **Belief Distributions**
  - ◇ A **belief** reflects the robot's **internal knowledge** about the state of the environment.
  - ◇ State cannot be measured directly. For example, a robot pose is not measurable directly (not even with GPS!).
  - ◇ Instead, the robot must **infer its pose from data**. We therefore distinguish the **true state** from its **internal belief**.



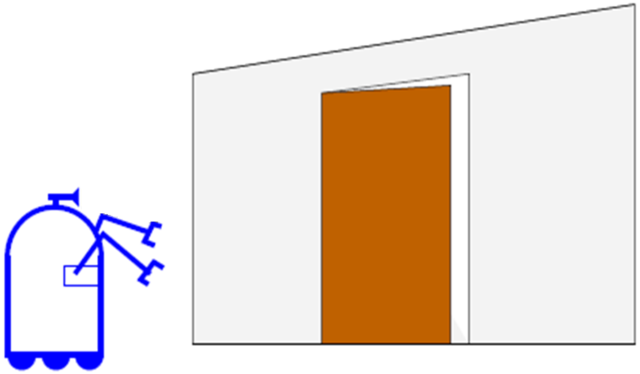
Environment Interaction

- **Belief Distributions**
  - ◊ Intelligent systems represents beliefs through **conditional probability distributions**.
  - ◊ A belief distribution assigns a probability (or density value) to each **possible hypothesis** with regards to the true state.
  - ◊ Belief distributions are **posterior probabilities** over state variables conditioned on the available data.

Environment Interaction

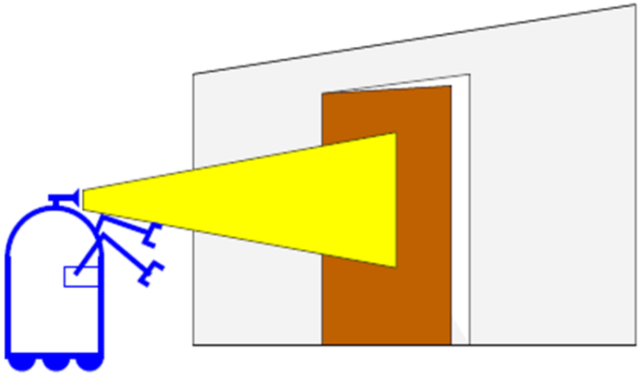
- **Belief Distributions**

  ◇ The belief is taken **before** incorporating the measurement z_t is



$$\overline{bel}(x_t) = p(x_t \mid z_{1:t-1}, u_{1:t})$$ → Prediction

  ◇ The belief is taken **after** incorporating the measurement z_t is



$$bel(x_t) = p(x_t \mid z_{1:t}, u_{1:t})$$ → Update

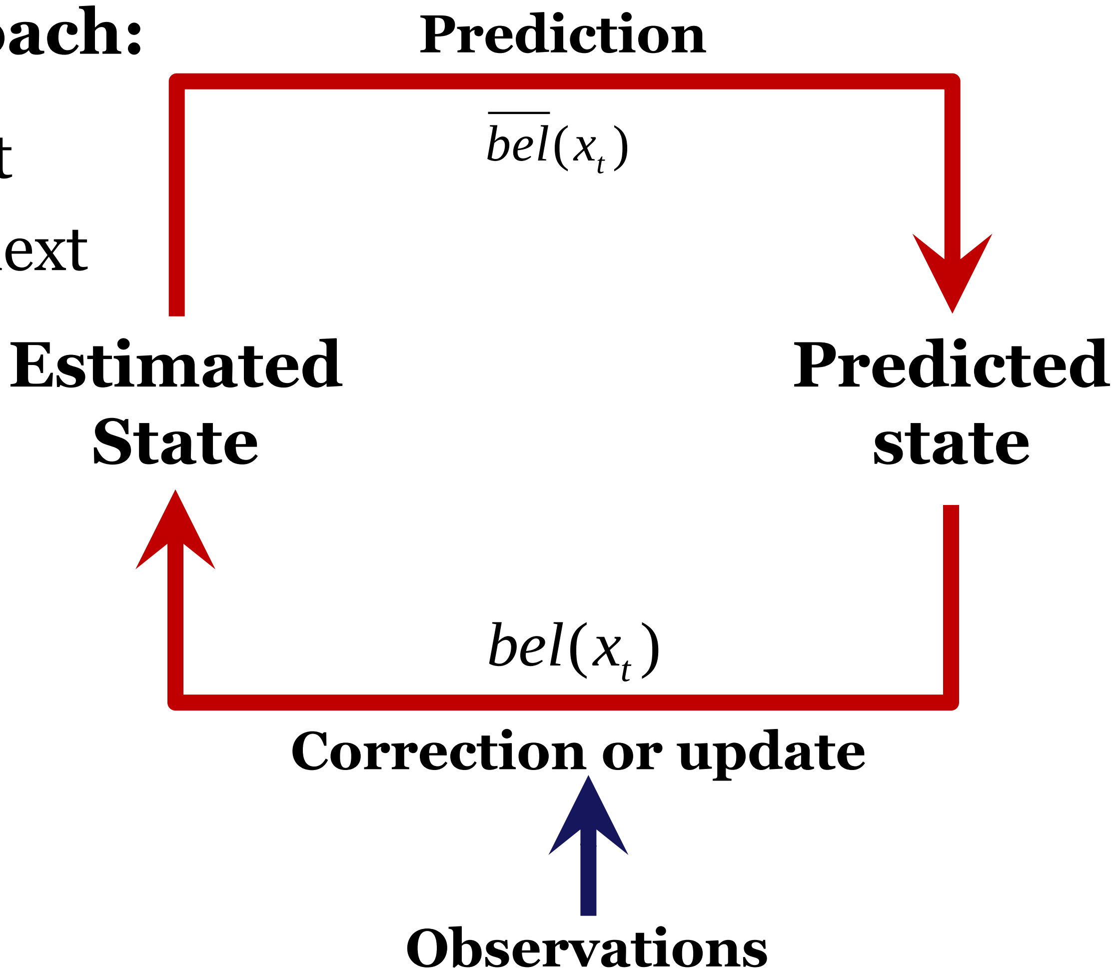
Outline

- Uncertainty
- State Estimation
- Basic Concepts in Probability
- Bayesian Rule
- Environment Interaction
- **Bayes Filter Algorithm**
- Summary

Bayes Filter Algorithm

- A **statistically optimal recursive estimator** of the state of a system given a set of observations.

- **2-step iterative approach:**
  - ◇ **Predict:** Take current state and predict the next state using the system model.
  - ◇ **Update:** Adjust the predicted state using sensor observations.



Bayes Filter Algorithm

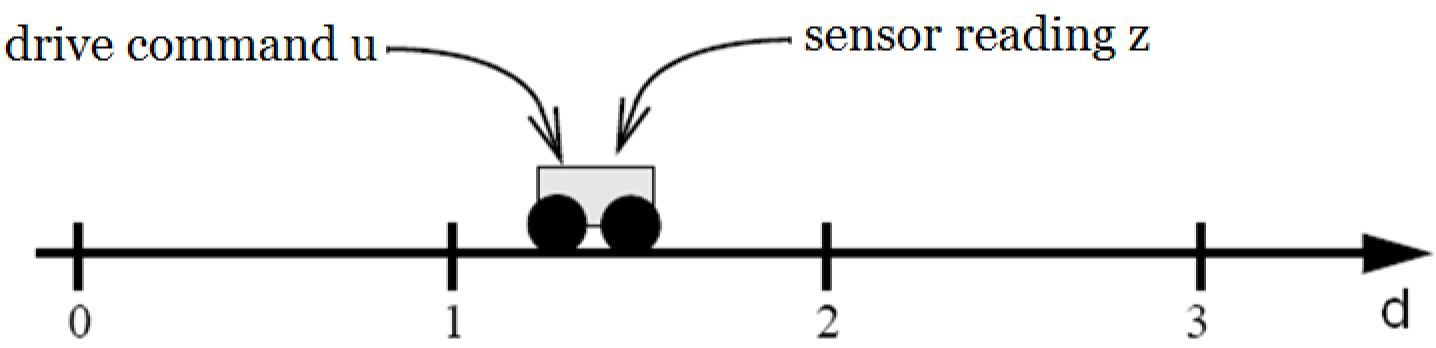
1. **Algorithm** $Bayes\_filter(bel(x_{t-1}), u_t, z_t)$: ← Initial belief, Control, Observation
2. for all x_t do
3. $\overline{bel}(x_t) = \int p(x_t \mid u_t, x_{t-1}) bel(x_{t-1}) dx$ ← Prediction
4. $bel(x_t) = \eta p(z_t \mid x_t) \overline{bel}(x_t)$ ← Update
5. endfor
6. Return $bel(x_t)$ ← Final belief

$\overline{bel}(x_t)$ → Prior belief (before taking the measurement)

$bel(x_t)$ → Posterior belief (after taking the measurement)

Bayes Filter Algorithm

- **Case Study-1**
  - ◇ Assume a robot is driving in a straight line along the d axis, starting at the true position **d=0**.
  - ◇ The robot executes **driving commands** with distance u, where u is an integer, and
  - ◇ It receives **sensor data** from its on-board global (absolute) positioning system z (e.g. a GPS receiver), where z is also an integer.



Bayes Filter Algorithm

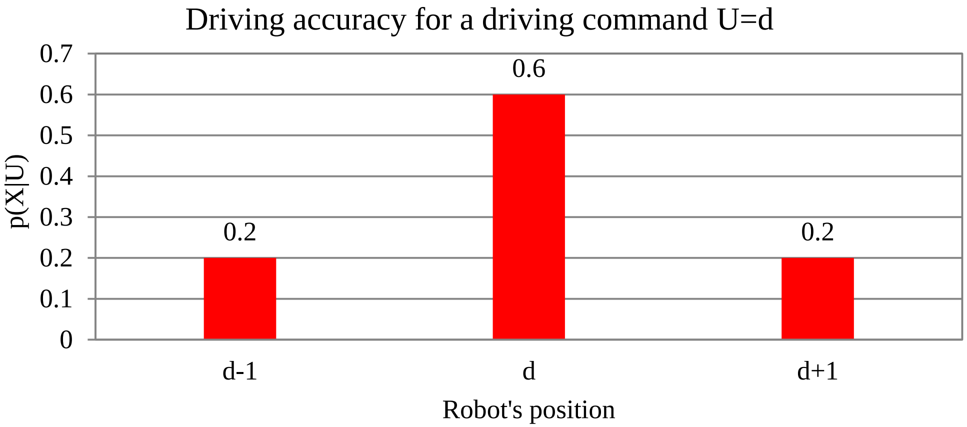
- **Case Study-1**
  - ◇ The robot's **driving accuracy** from an arbitrary starting position has to be established by **extensive experimental measurements** and can then be expressed by:

$$p(x = d - 1 | U = d) = 0.2$$

$$p(x = d | U = d) = 0.6$$

$$p(x = d + 1 | U = d) = 0.2$$

Note that in this example, the robot's true position can only **deviate by plus or minus one** unit (e.g. cm); all position data are discrete.



Bayes Filter Algorithm

- **Case Study-1**
  - ◇ the accuracy of the robot's position **sensor** has to be established by **measurements**, before it can be expressed as following.

$$p(Z = z \mid x = z - 1) = 0.1$$

$$p(Z = z \mid x = z) = 0.8$$

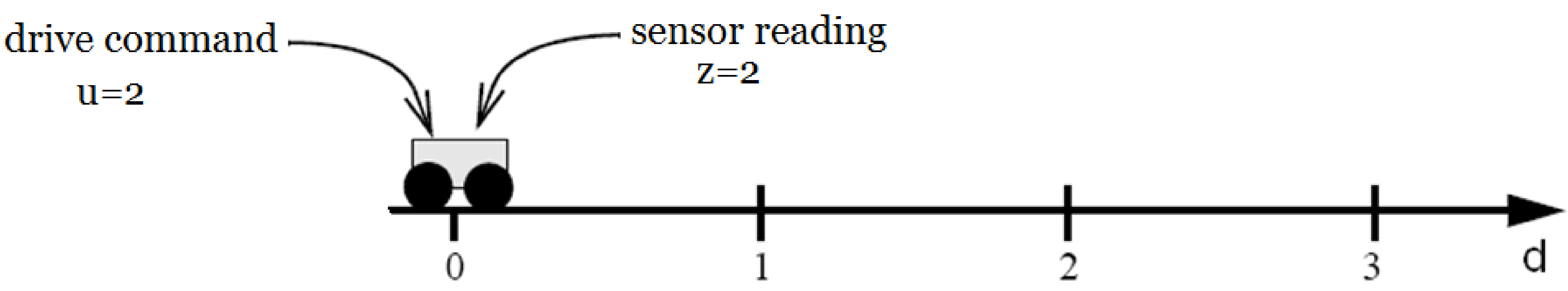
$$p(Z = z \mid x = z + 1) = 0.1$$

Note that in this example, there will again only be a **possible deviation** from the true position by plus or minus one unit.

Bayes Filter Algorithm

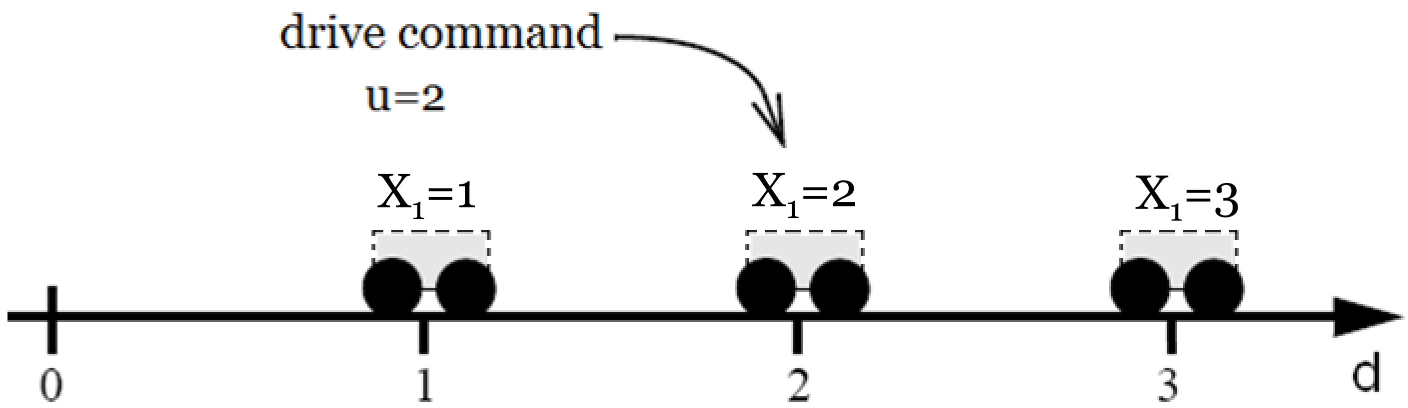
- **Case Study-1**
  - **Initial Condition**

    $bel(X_0 = 0) = 1$ → starting at the true position **d=0.**
  - **First Iteration**
    - Assuming the robot has executed a driving command with **u=2** and after completion of this command, its local sensor reports its position as **z=2**.



Bayes Filter Algorithm

- **Case Study-1: First Iteration (cont'd)**

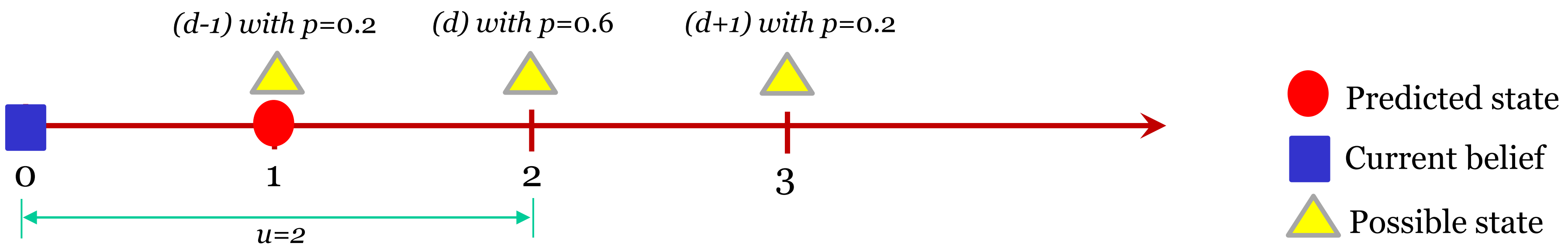


$$\overline{bel}(X_1 = 1) = \sum_{X_o} p(X_1 | U_1, X_0) bel(X_o)$$

Predict

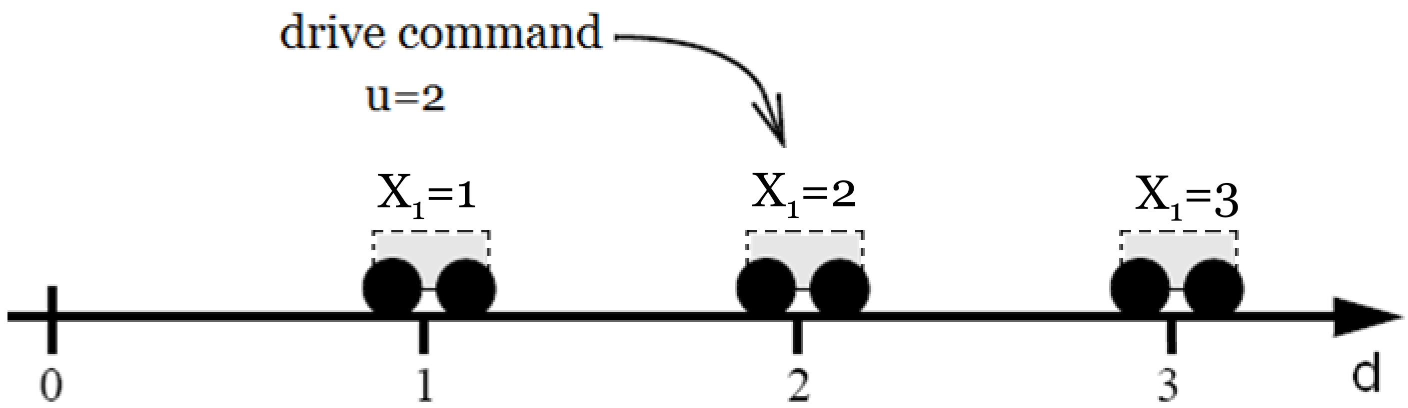
$$= p(X_1 = 1 | U_1 = 2, X_0 = 0) bel(X_o = 0)$$

$$= 0.2 \times 1 = 0.2$$



Bayes Filter Algorithm

- **Case Study-1: First Iteration (cont'd)**



$$\overline{bel}(X_1 = 1) = \sum_{X_o} p(X_1 | U_1, X_0) bel(X_o)$$ Predict

$$= p(X_1 = 1 | U_1 = 2, X_0 = 0) bel(X_o = 0)$$

$$= 0.2 \times 1 = 0.2$$

$$bel(X_1 = 1) = \eta p(Z_1 = 2 | X_1 = 1) \overline{bel}(X_1 = 1)$$ Update

$$= \eta \times 0.1 \times 0.2 = 0.02\eta$$

Bayes Filter Algorithm

- **Case Study-1: First Iteration (cont'd)**

$$\overline{bel}(X_1=2)=\sum_{X_o} p(X_1|U_1,X_0)bel(X_o)$$ → Predict

$$= p(X_1=2|U_1=2,X_0=0)bel(X_o=0)$$

$$=0.6\times 1=0.6$$

$$bel(X_1=2)=\eta p(Z_1=2|X_1=2)\overline{bel}(X_1=2)$$ → Update

$$=\eta\times 0.8\times 0.6=0.48\eta$$

$$\overline{bel}(X_1=3)=\sum_{X_o} p(X_1|U_1,X_0)bel(X_o)$$ → Predict

$$= p(X_1=3|U_1=2,X_0=0)bel(X_o=0)$$

$$=0.2\times 1=0.2$$

$$bel(X_1=3)=\eta p(Z_1=2|X_1=3)\overline{bel}(X_1=3)$$ → Update

$$=\eta\times 0.1\times 0.2=0.02\eta$$

Bayes Filter Algorithm

- **Case Study-1: First Iteration (cont'd)**

$$bel(X_1 = 1) = 0.02\eta$$

$$bel(X_1 = 2) = 0.48\eta$$

$$bel(X_1 = 3) = 0.02\eta$$

$$0.02\eta + 0.48\eta + 0.02\eta = 1 \Rightarrow \eta = 1.92$$

$$bel(X_1 = 1) = 0.04;$$

$$bel(X_1 = 2) = 0.92;$$

$$bel(X_1 = 3) = 0.04$$

So the robot is **most likely to be in position 2**, but it remembers all probabilities at this stage.

Bayes Filter Algorithm

- **Case Study-1: Second Iteration**

  ◇ Let us assume the robot executes a second driving command, this time with **u=1**, but after execution its sensor still reports **z=2**.

  ◇ The robot will now recalculate its position belief according to the conditional probabilities, with $\boldsymbol{x}$ denoting the robot's true position after driving and $\boldsymbol{x_o}$ before driving:

Bayes Filter Algorithm

- **Case Study-1: Second Iteration (cont'd)**

$$\overline{bel}(X_2=1)=\sum_{X_1} p(X_2 \mid U_1, X_1)bel(X_1)$$

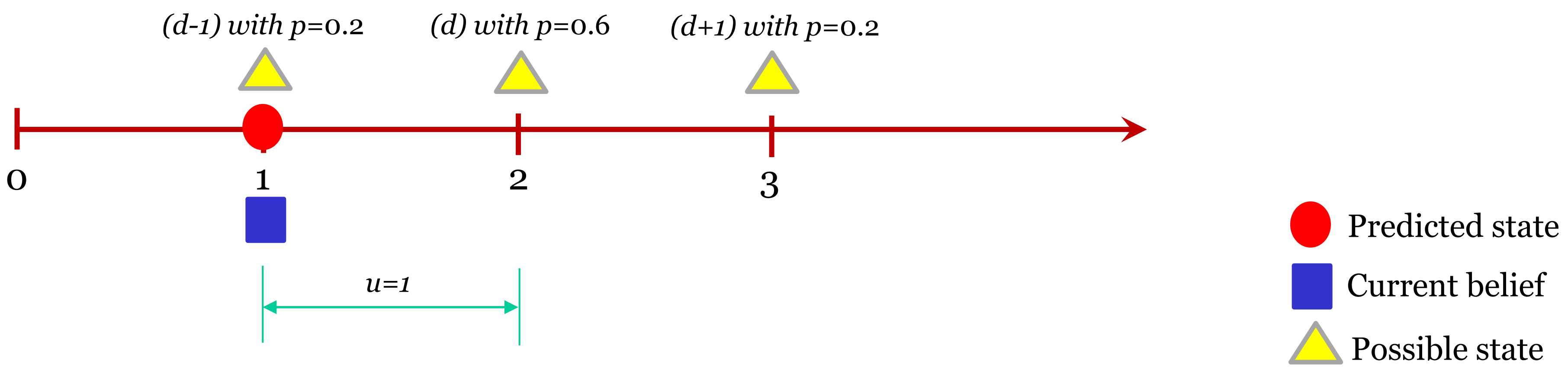
$$= p(X_2=1 \mid U_1=1, X_1=1)bel(X_1=1)$$

$$+ p(X_2=1 \mid U_1=1, X_1=2)bel(X_1=2)$$

$$+ p(X_2=1 \mid U_1=1, X_1=3)bel(X_1=3)$$

Predict

$$p(X_2=1 \mid U_1=1, X_1=1)bel(X_1=1)=0.2\times 0.04=0.008$$



Bayes Filter Algorithm

- **Case Study-1: Second Iteration (cont'd)**

$$\overline{bel}(X_2=1)=\sum_{X_1} p(X_2\,|\,U_1,X_1)bel(X_1)$$

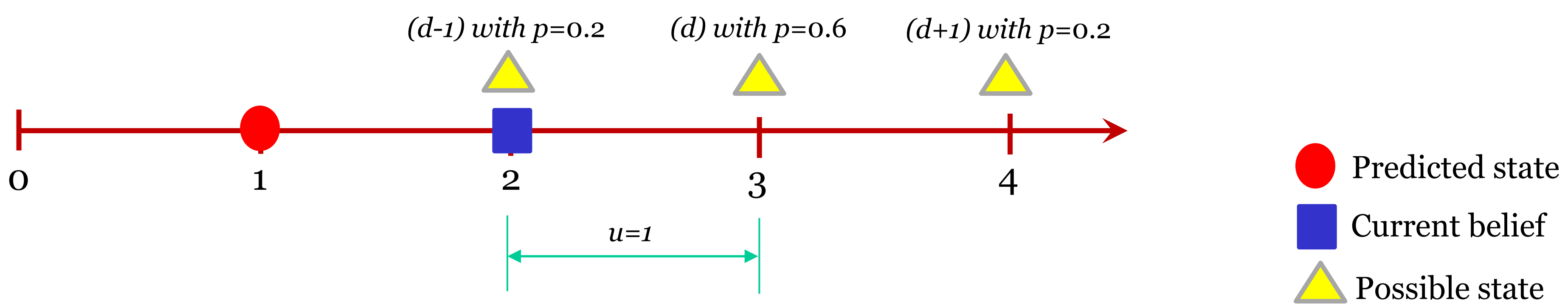
$$= p(X_2=1|U_1=1,X_1=1)bel(X_1=1)$$

$$+ p(X_2=1|U_1=1,X_1=2)bel(X_1=2)$$

$$+ p(X_2=1|U_1=1,X_1=3)bel(X_1=3)$$

Predict

$$p(X_2=1|U_1=1,X_1=2)bel(X_1=2)=0\times 0.92=0$$



Bayes Filter Algorithm

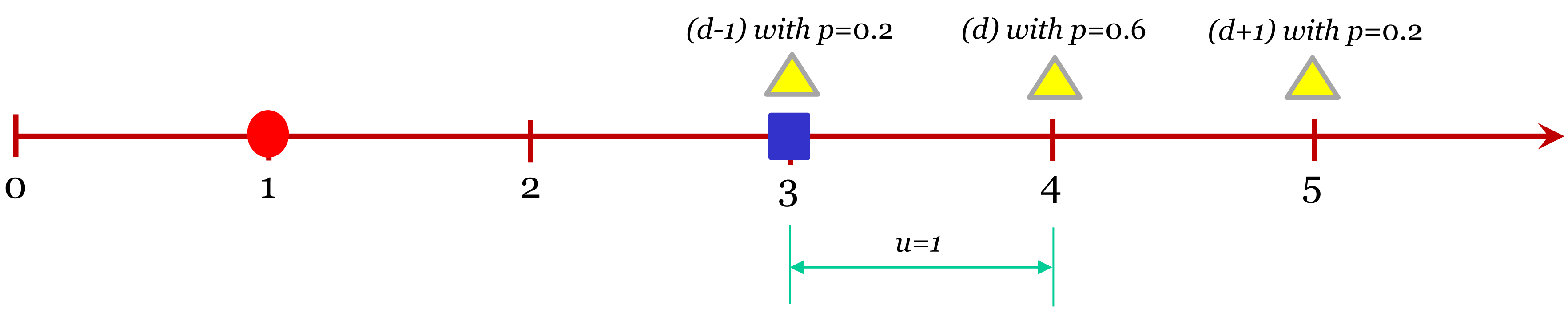
- **Case Study-1: Second Iteration (cont'd)**

$$\overline{bel}(X_2=1)=\sum_{X_1} p(X_2\,|\,U_1,X_1)bel(X_1)$$

Predict

$$= p(X_2=1|U_1=1,X_1=1)bel(X_1=1)$$
$$+ p(X_2=1|U_1=1,X_1=2)bel(X_1=2)$$
$$+ p(X_2=1|U_1=1,X_1=3)bel(X_1=3)$$

$$p(X_2=1|U_1=1,X_1=3)bel(X_1=3)=0\times 0.04=0$$



Bayes Filter Algorithm

- **Case Study-1: Second Iteration (cont'd)**

$$\overline{bel}(X_2=1)=\sum_{X_1} p(X_2 \mid U_1, X_1)bel(X_1)$$

Predict

$$\begin{aligned} &= p(X_2=1 \mid U_1=1, X_1=1)bel(X_1=1) \\ &+ p(X_2=1 \mid U_1=1, X_1=2)bel(X_1=2) \\ &+ p(X_2=1 \mid U_1=1, X_1=3)bel(X_1=3) \\ &= 0.008+0+0=0.008 \end{aligned}$$

$$bel(X_2=1)=\eta p(Z_2=2 \mid X_2=1)\overline{bel}(X_2=1)$$

Update

$$=\eta \times 0.1 \times 0.008 = 0.0008\,\eta$$

Bayes Filter Algorithm

- **Case Study-1: Second Iteration (cont'd)**
  - ◇ Similarly:

$$\overline{bel}(X_2=2)=\sum_{X_1} p(X_2 \mid U_1, X_1)bel(X_1)$$ → Predict

$$=p(X_2=2 \mid U_1=1, X_1=1)bel(X_1=1)$$
$$+p(X_2=2 \mid U_1=1, X_1=2)bel(X_1=2)$$
$$+p(X_2=2 \mid U_1=1, X_1=3)bel(X_1=3)$$
$$=0.6\times 0.04+0.2\times 0.92+0\times 0.04=0.208$$

$$bel(X_2=2)=\eta p(Z_2=2 \mid X_2=2)\overline{bel}(X_2=2)$$ → Update

$$=\eta\times 0.8\times 0.208=0.1664\,\eta$$

Bayes Filter Algorithm

- **Case Study-1: Second Iteration (cont'd)**

$$\overline{bel}(X_2=3)=\sum_{X_1} p(X_2\,|\,U_1,X_1)bel(X_1)$$ → Predict

$$= p(X_2=3|U_1=1,X_1=1)bel(X_1=1)$$
$$+ p(X_2=3|U_1=1,X_1=2)bel(X_1=2)$$
$$+ p(X_2=3|U_1=1,X_1=3)bel(X_1=3)$$
$$=0.2\times0.04+0.6\times0.92+0.2\times0.04=0.568$$

$$bel(X_2=3)=\eta p(Z_2=2\,|\,X_2=3)\overline{bel}(X_2=3)$$ → Update

$$=\eta\times0.1\times0.568=0.0568\,\eta$$

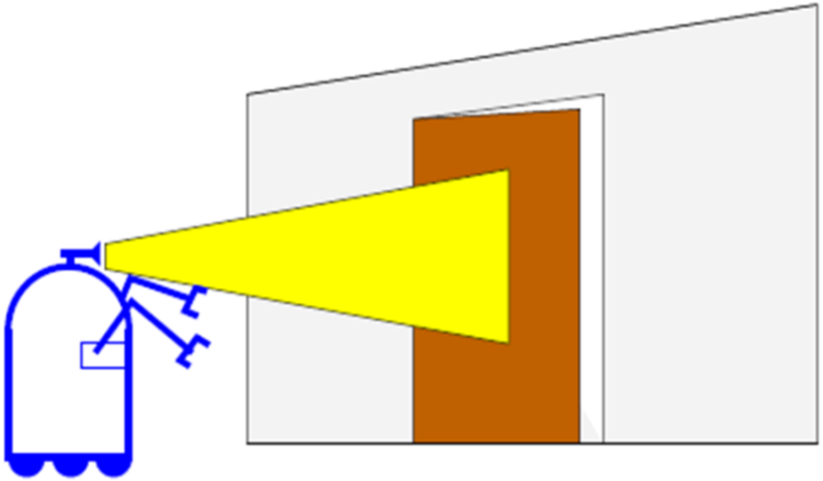
$$0.0008\eta+0.1664\eta+0.0568\eta=1\Rightarrow\eta=4.464$$

$$bel(X_2=1)=0.0036;$$
$$bel(X_2=2)=0.7428;$$
$$bel(X_2=3)=0.254.$$

The robot will be at X_2=2 with higher probability after the new control action u_1=1 in accordance with the data reported by the GPS

Bayes Filter Algorithm

- **Case Study-2**
  - ◇ A robot is driving in front of a door. The robot is estimating the **state of a door** (open or closed) using its camera.



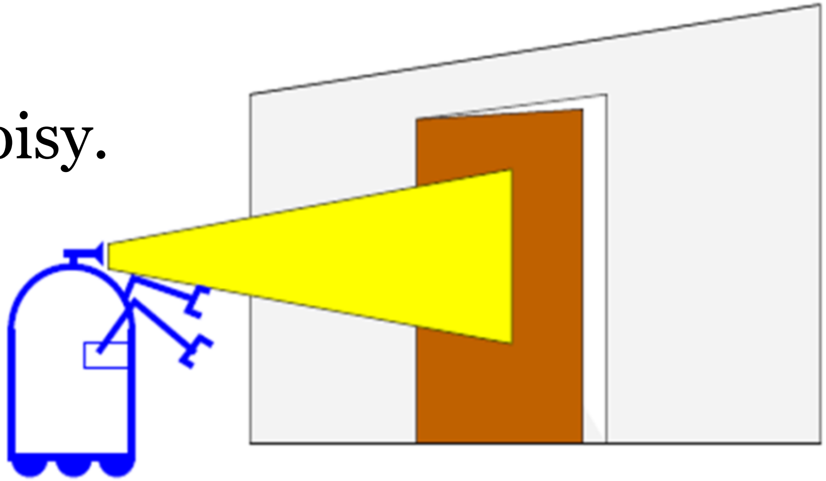
  - ◇ To make this problem simple, let us assume that the door can be in one of two possible states, **open or closed**, and that only the robot can change the state of the door.
  - ◇ Let us furthermore assume that the robot **does not know** the **state of the door initially**. Instead, it assigns equal prior probability to the two possible door states:

$$bel(X_o = \text{open}) = 0.5, \qquad bel(X_o = \text{closed}) = 0.5$$

[4]

Bayes Filter Algorithm

- **Case Study-2**
  - ◇ Assume the robot's sensors are noisy.
  - ◇ The noise is characterized by the following conditional probabilities:



$$p(Z_t = \text{sense\_open} \mid X_t = \text{is\_open}) = 0.6$$
$$p(Z_t = \text{sense\_closed} \mid X_t = \text{is\_open}) = 0.4$$

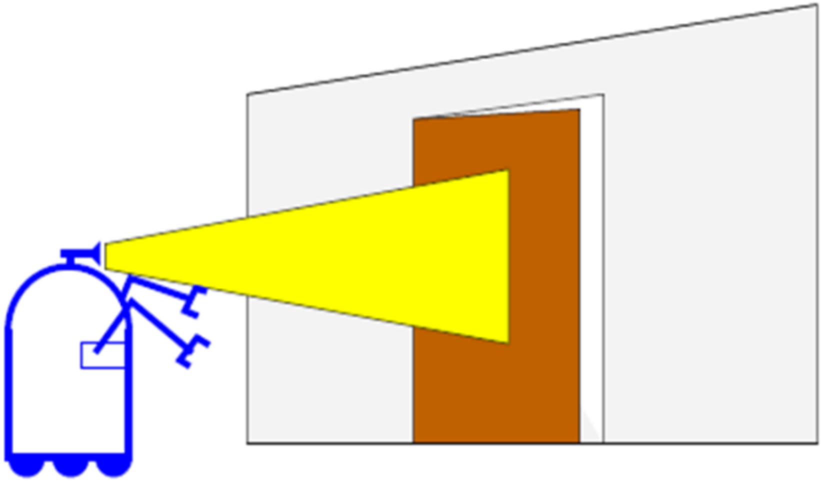
$$p(Z_t = \text{sense\_open} \mid X_t = \text{is\_closed}) = 0.2$$
$$p(Z_t = \text{sense\_closed} \mid X_t = \text{is\_closed}) = 0.8$$

These probabilities suggest that the robot's sensors are **relatively reliable** in detecting a **closed door**.

[4]

Bayes Filter Algorithm

- **Case Study-2**
  - ◇ Finally, let us assume the robot uses its **manipulator** to push the door open.



  - ◇ If the door is already open, it will remain open. If it is closed, the robot has a 0.8 chance that it will be open afterwards:

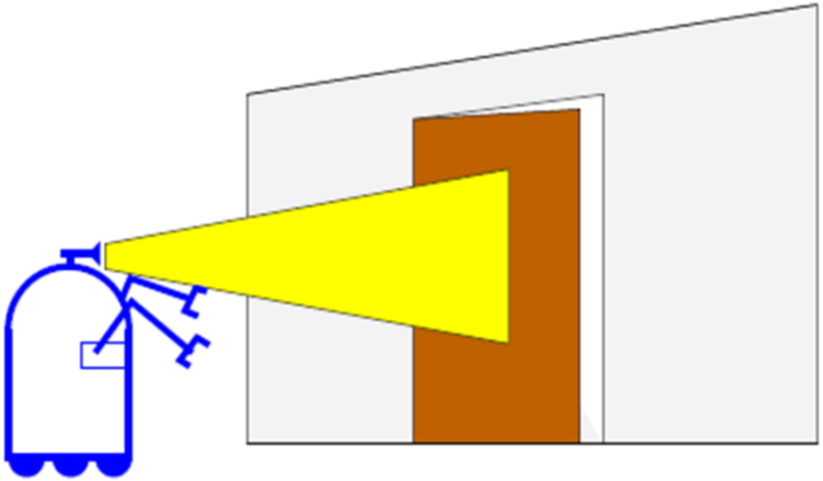
$$p(X_t = \text{is\_open} \mid U_t = \text{push}, X_{t-1} = \text{is\_open}) = 1$$
$$p(X_t = \text{is\_closed} \mid U_t = \text{push}, X_{t-1} = \text{is\_open}) = 0$$

$$p(X_t = \text{is\_open} \mid U_t = \text{push}, X_{t-1} = \text{is\_closed}) = 0.8$$
$$p(X_t = \text{is\_closed} \mid U_t = \text{push}, X_{t-1} = \text{is\_closed}) = 0.2$$

[4]

Bayes Filter Algorithm

- **Case Study-2**
  - ◇ It can also choose **not to use its manipulator**, in which case the state of the world does not change.



  - ◇ This is stated by the following conditional probabilities:

$$p(X_t = \text{is\_open} \mid U_t = \text{do\_nothing}, X_{t-1} = \text{is\_open}) = 1$$
$$p(X_t = \text{is\_closed} \mid U_t = \text{do\_nothing}, X_{t-1} = \text{is\_open}) = 0$$

$$p(X_t = \text{is\_open} \mid U_t = \text{do\_nothing}, X_{t-1} = \text{is\_closed}) = 0$$
$$p(X_t = \text{is\_closed} \mid U_t = \text{do\_nothing}, X_{t-1} = \text{is\_closed}) = 1$$

[4]

Bayes Filter Algorithm

- **Case Study-2**
  - ◇ Suppose at time t, the robot takes **no control action** but it senses an **open door**.
  - ◇ The resulting **posterior belief** is calculated by the Bayes filter using the prior belief **bel(Xo)**, the control u1 = **do_nothing**, and the measurement **sense_open** as input.

$$\overline{bel}(x_1) = \int p(x_1 \mid u_1, x_0) bel(x_0) dx_0$$

  - ◇ Since the state space is finite, the integral turns into a finite sum: $\overline{bel}(x_1) = \sum_{x_o} p(x_1 \mid u_1, x_0) bel(x_0)$

$$= p(x_1 \mid U_1 = \text{do\_nothing}, X_0 = \text{is\_open}) bel(X_0 = \text{is\_open})$$
$$+ p(x_1 \mid U_1 = \text{do\_nothing}, X_0 = \text{is\_closed}) bel(X_0 = \text{is\_closed})$$

[4]

Bayes Filter Algorithm

- **Case Study-2**
  - ◇ For the hypothesis $\mathbf{X_1}$ **= is_open**, we obtain

$$\overline{bel}(X_1 = \text{is\_open}) = p(X_1 = \text{is\_open} \mid U_1 = \text{do\_nothing}, X_0 = \text{is\_open})bel(X_0 = \text{is\_open})$$
$$+ p(X_1 = \text{is\_open} \mid U_1 = \text{do\_nothing}, X_0 = \text{is\_closed})bel(X_0 = \text{is\_closed})$$
$$= 1 \times 0.5 + 0 \times 0.5 = 0.5$$

  - ◇ Likewise, for the hypothesis $\mathbf{X_1}$ **= is_closed**, we obtain

$$\overline{bel}(X_1 = \text{is\_closed}) = p(X_1 = \text{is\_closed} \mid U_1 = \text{do\_nothing}, X_0 = \text{is\_open})bel(X_0 = \text{is\_open})$$
$$+ p(X_1 = \text{is\_closed} \mid U_1 = \text{do\_nothing}, X_0 = \text{is\_closed})bel(X_0 = \text{is\_closed})$$
$$= 0 \times 0.5 + 1 \times 0.5 = 0.5$$

The fact that the belief $\overline{bel}(x_1)$ equals our prior belief $bel(x_o)$ should not surprise, as the action **do_nothing** does not affect the state of the world; neither does the world change over time by itself in this example.

[4]

Bayes Filter Algorithm

- **Case Study-2**
  - ◇ Incorporating the measurement, however, changes the belief.
  - ◇ Line 4 of the Bayes filter algorithm implies

$$bel(x_1) = \eta p(Z_1 = \text{sense\_open} \mid x_1)\overline{bel}(x_1)$$

  - ◇ For the two possible cases **X_1=is_open** and **X_1=is_closed**, we get

$$bel(X_1 = \text{is\_open}) = \eta p(Z_1 = \text{sense\_open} \mid X_1 = \text{is\_open})\overline{bel}(X_1 = \text{is\_open})$$
$$= \eta \times 0.6 \times 0.5 = 0.3\eta$$

$$bel(X_1 = \text{is\_closed}) = \eta p(Z_1 = \text{sense\_open} \mid X_1 = \text{is\_closed})\overline{bel}(X_1 = \text{is\_closed})$$
$$= \eta \times 0.2 \times 0.5 = 0.1\eta$$

The normalizer η is now easily calculated: $\eta = (0.3 + 0.1)^{-1} = 2.5$

[4]

Bayes Filter Algorithm

- **Case Study-2**
  - ◇ Hence, we have

$$bel(X_1 = \text{is\_open}) = 0.75$$
$$bel(X_1 = \text{is\_closed}) = 0.25$$

  - ◇ **In the second iteration:**

For **u_2 = push** and **z_2 = sense_open** we get

$$\overline{bel}(X_2 = \text{is\_open}) = 1 \times 0.75 + 0.8 \times 0.25 = 0.95$$
$$\overline{bel}(X_1 = \text{is\_closed}) = 0 \times 0.75 + 0.2 \times 0.25 = 0.05$$

and

$$bel(X_2 = \text{is\_open}) = \eta \times 0.6 \times 0.95$$
$$bel(X_2 = \text{is\_closed}) = \eta \times 0.2 \times 0.05$$
$$\eta = 1.72 \Rightarrow bel(X_2 = \text{is\_open}) \approx 0.983, \qquad bel(X_2 = \text{is\_closed}) \approx 0.017$$

Outline

- Uncertainty
- State Estimation
- Basic Concepts in Probability
- Bayesian Rule
- Environment Interaction
- Bayes Filter Algorithm
- **Summary**

Summary

- **Probabilistic descriptions** are indispensable when representing and dealing **quantitatively with uncertainty**.
- In most of autonomous systems, we often assume the **independence** of random variables even when this assumption is not strictly true.
- The simplification that results makes a number of the existing mapping and navigation **algorithms tenable**.
- A further simplification, revolves around one specific probability density function used more often than any other when modeling error: the **Gaussian distribution**.

Summary

- The interaction of a robot and its environment is modeled as a coupled dynamical system, in which the robot can manipulate its environment by choosing controls, and in which it can perceive its environment through sensor measurements.
- Bayesian probability is a formalism that allows us to **reason about beliefs** under conditions of **uncertainty**.
- The belief of a robot is the posterior distribution over the state of the environment (including the robot state), given all past sensor measurements and all past controls. The Bayes filter is the principal algorithm for calculating the belief in robotics. The Bayes filter is recursive; the belief at time t is calculated from the belief at time t-1.

Summary

- The **Bayes filter** makes a **Markov assumption** that specifies that **the state is a complete summary of the past**. This assumption implies the belief is sufficient to represent the past history of the robot. In robotics, the Markov assumption is usually only an approximation.

References

1. M. Endsley, “Toward a theory of situation awareness in dynamic systems,” Human Factors 37(1), 32–64, 1995.
2. Mieczyslaw M. Kokar, Christopher J. Matheus and Kenneth Baclawski , “Ontology-based situation awareness,” Information Fusion, Volume 10, Issue 1 (January 2009) Pages 83-98, ISSN:1566-2535, Elsevier Science Publishers, 2009.
3. Bahador Khaleghi, Alaa Khamis, Fakhreddine O. Karray and Saiedeh N. Razavi, “Multisensor data fusion: A review of the state-of-the-art,” Information Fusion, 14(1):28–44, January 2013.
4. Sebastian Thrun, Wolfram Burgard and Dieter Fox. *Probabilistic Robotics*. MIT Press, ISBN-13:978-0-262-20162-9, 2005.
5. Henrik I. Christensen, Gregory D. Hager. Sensing and Estimation, Ch.4 Springer Handbook of Robotics.