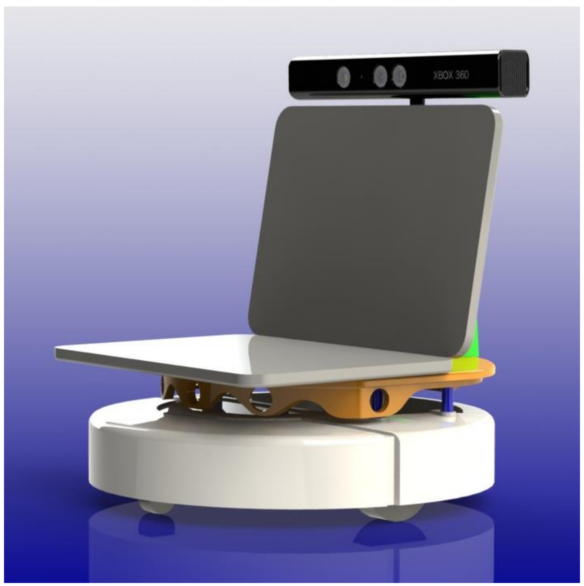
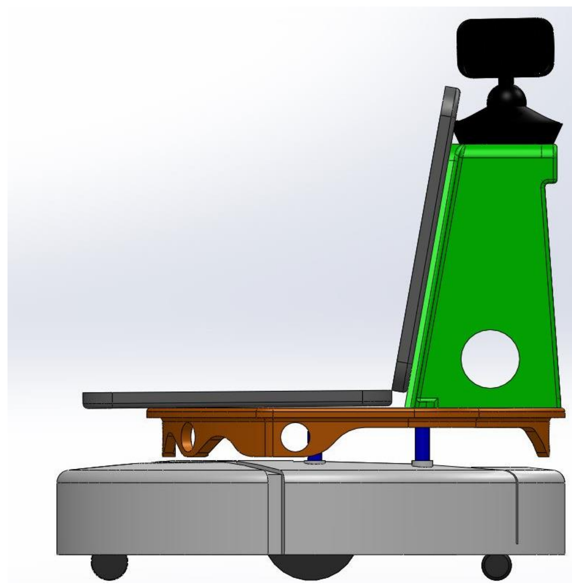
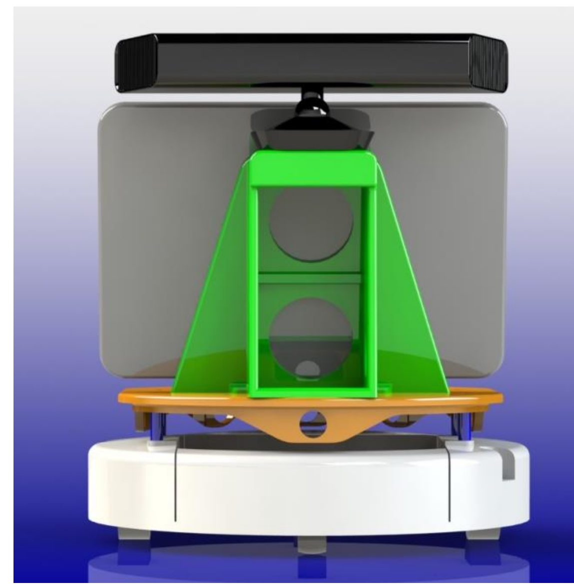
MAE 493G, CpE 493M, Mobile Robotics

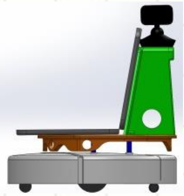
6. Basic Probability







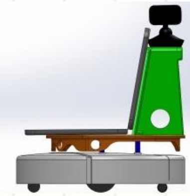
Instructor: Yu Gu, Fall 2013



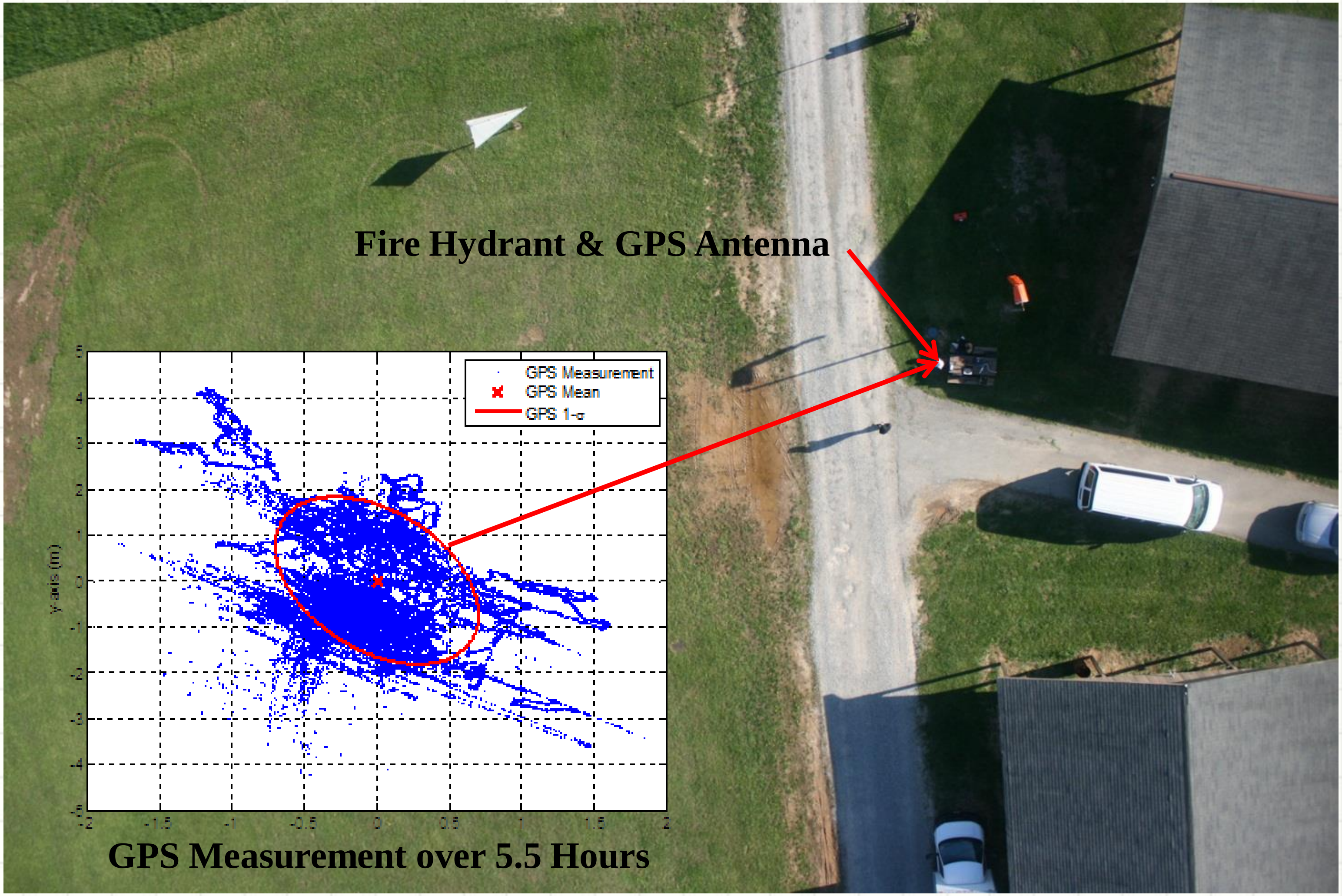
Uncertainties in Robotics

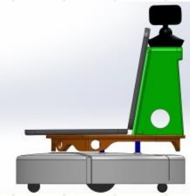
- Robot environments are inherently unpredictable;
- Sensors and data acquisition systems are never perfect;
- Robot actuators are not perfect either;
- The internal models of the environment and robot itself are often inadequate, inaccurate, or totally wrong;
- Hardware components can break down and bugs exist in the robot software;

- **The ability to recognize ones limitations and to make decisions accordingly is a sign of <u>intelligence</u>**;
- Therefore, we need to learn some probability and statistics!

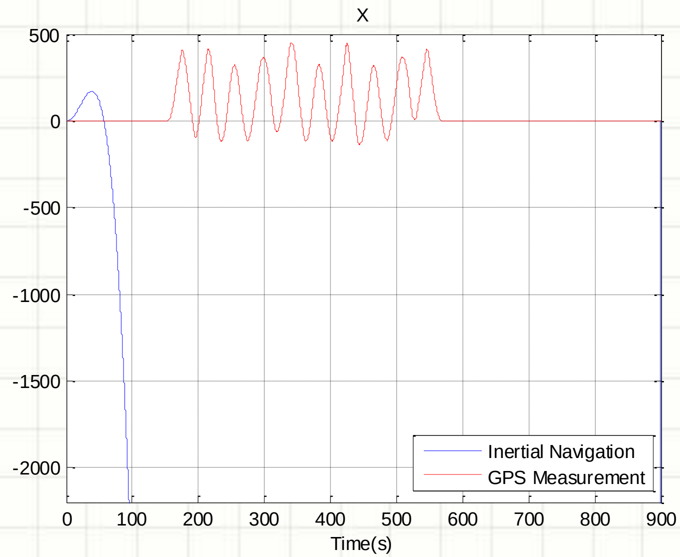


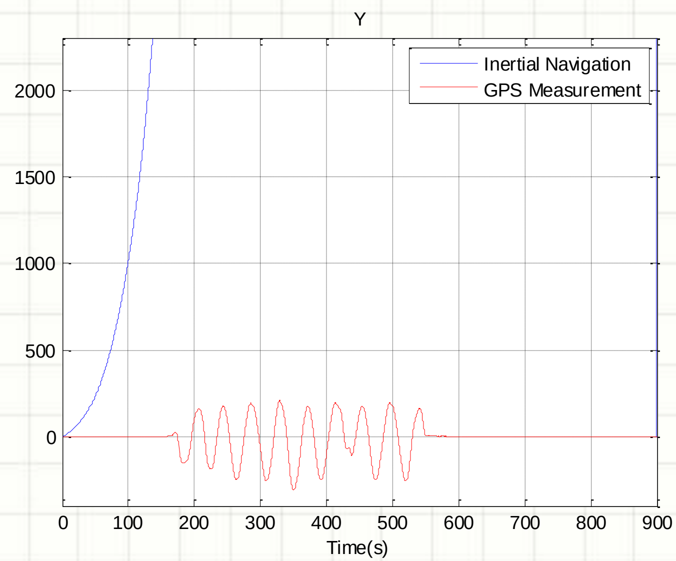
A GPS Example

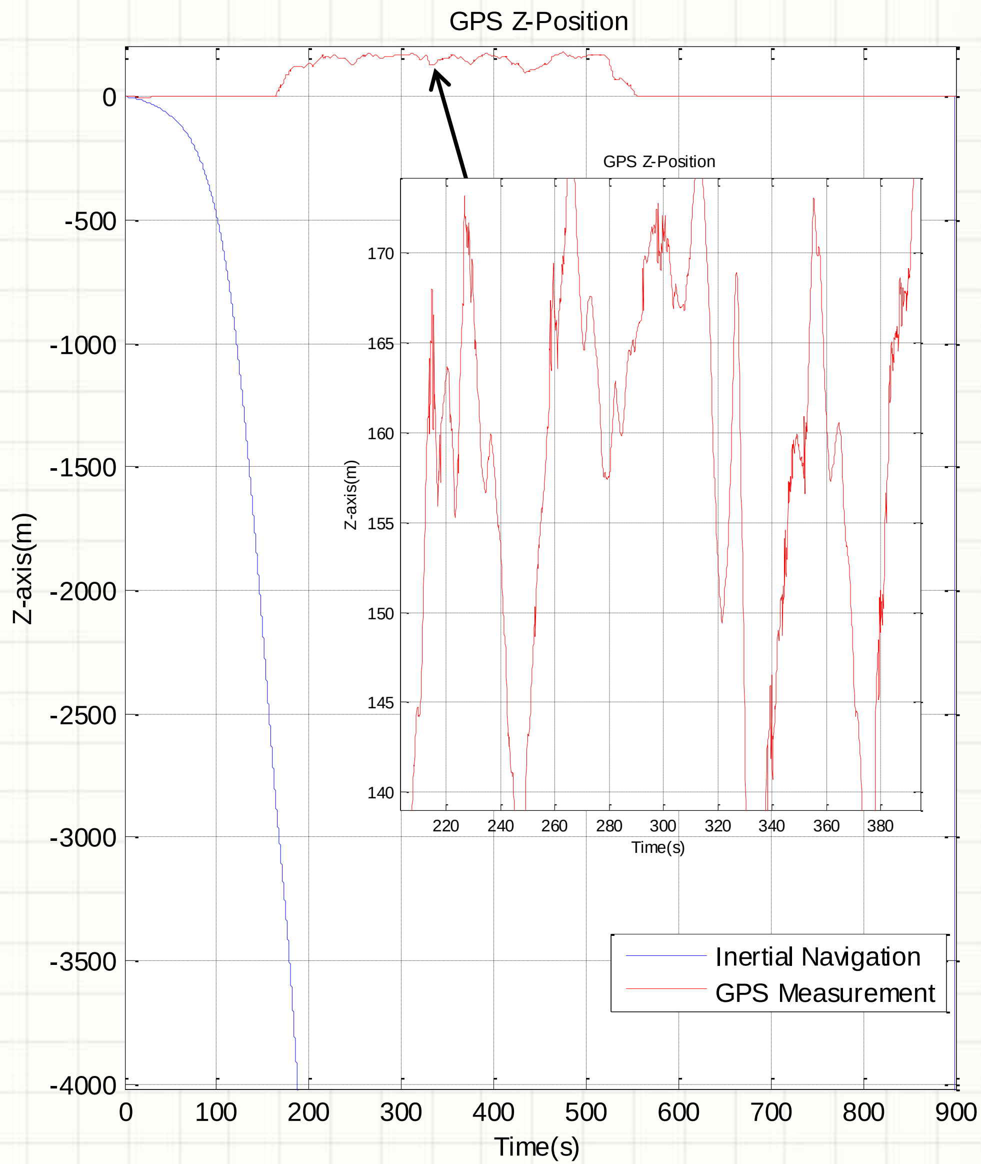




GPS & Inertial Navigation Example







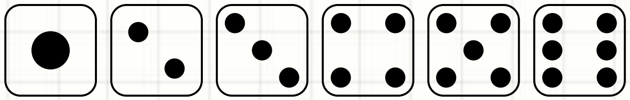
Probability

- Probability is a measure or estimation of how likely an event will happen or that a statement is true.

$$\text{Probability} = \frac{\text{Occurence}}{\text{A Large Number of Tries}}$$

50%

50%



1/6 +1/6 +1/6 +1/6 +1/6 +1/6 = 1

Probability (In Other Words)

- Let S denotes the set of all possible outcomes for a given experiment, the sample space, and let E be an event, i.e., $E \subset S$
- The probability of the event E occurring when the experiment is conducted is demoted Pr(E);
- The probability maps S to the interval [0,1]. It has the following basic properties:

$$0 \le \Pr(E) \le 1 \text{ \textit{for all} } E \subset S$$

$$\Pr(S) = 1 \qquad \Pr(\emptyset) = 0$$

- For example, Pr(head) = 0.5 and Pr(tail) = 0.5 when flipping a fair coin. Also, $\Pr(S) = \Pr(Head \cup Tail) = 1$;
- Question: what is the probably for the temperature to be exactly 65°F at noon tomorrow?

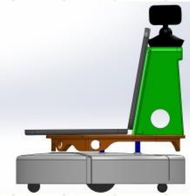
Conditional Probability

- Conditional probability Pr(A|B) is the probability of some event A, given the occurrence of some other event B; e.g., the probability of a high UV index if it's sunny outside.

$$\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

$$\Pr(A \cap B) = \Pr(A \mid B)\Pr(B)$$

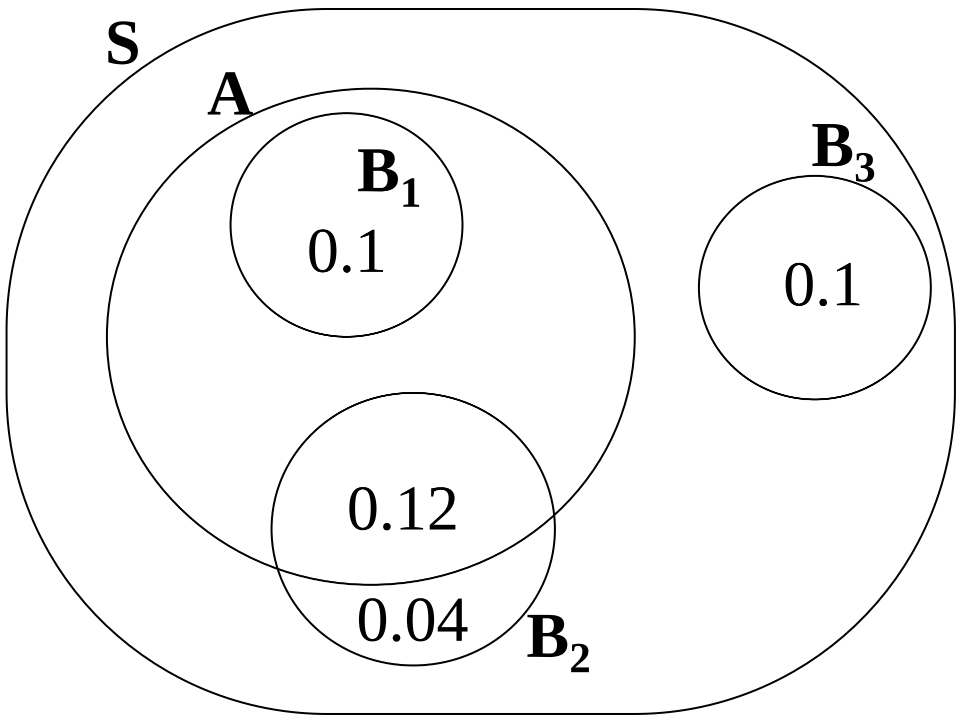
- For example, in a bag of 2 red balls and 2 blue balls (4 balls in total), the probability of taking a red ball is 1/2; however, when taking a second ball, the probability of it being either a red ball or a blue ball depends on the ball previously taken, such as, if a red ball was taken, the probability of picking a red ball again would be 1/3 since only 1 red and 2 blue balls would have been remaining.



Conditional Probability Example

- Given the Euler diagram below, the unconditional probability Pr(A) is about 0.5. What are the conditional probabilities Pr(A|B_1), Pr(A|B_2), and P(A|B_3)?

Answer: Pr(A|B_1) = 1, Pr(A|B_2) = 0.75, and P(A|B_3) = 0.



Joint Probability

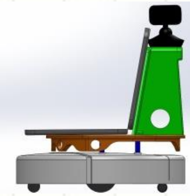
- Joint probability is a measure of two events happening at the same time, and can only be applied to situations where more than one observation can be occurred at the same time.

$$\Pr(A \cap B) = \Pr(A \mid B)\Pr(B) = \Pr(B \mid A)\Pr(A)$$

- For example, a joint probability can not be calculated when tossing a coin on the same flip. However, the joint probability can be calculated on the probability of rolling a 2 and a 5 using two different dice.
- For example, the probability that it's going to rain tomorrow is 0.3. The probability that someone would go picnic in the rain is 0.1. What is the probability that this person will be picnicking in the rain tomorrow?

$$\Pr(\text{Picnic} \cap \text{Rain}) = \Pr(\text{Picnic} \mid \text{Rain})\Pr(\text{Rain}) = 0.1 \times 0.3 = 0.03$$

- Notice that we are doing some <u>reasoning</u> here! This is how robots (and humans) make <u>decisions</u>.



Independent Events

- Two events are (statistically) independent if the occurrence of one does not affect the probability of the other:

$$\Pr(A \mid B) = \Pr(A) \qquad \Pr(B \mid A) = \Pr(B)$$

- Two events A and B are independent if and only if their joint probability equals the product of their probabilities:

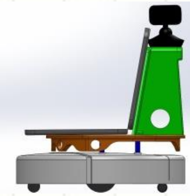
$$\Pr(A \cap B) = \Pr(A)\Pr(B)$$

- A finite set of events is mutually independent if and only if every event is independent of any intersection of the other events:

$$\Pr\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} \Pr(A_i)$$

- For example, what’s the probability of getting a total of 12 if you roll a fair dice twice?

Answer: 1/36



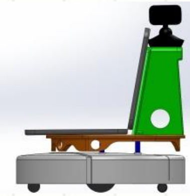
Law of Total Probability

- The law of total probability shows that if $\{B_n : n = 1, 2, 3, \ldots\}$ is a finite partition of a sample space and each event B_n is measurable, then for any event A of the same probability space: $\Pr(A) = \sum_n \Pr(A \cap B_n)$
- Or, alternatively: $\Pr(A) = \sum_n \Pr(A \mid B_n)\Pr(B_n)$
- The above mathematical statement might be interpreted as follows: given an outcome A, with known conditional probabilities given any of the B_n events, each with a known probability itself, what is the total probability that A will happen?
- For example: Suppose that two factories supply light bulbs to the market. Factory X's bulbs work for over 5000 hours in 99% of cases, whereas factory Y's bulbs work for over 5000 hours in 95% of cases. It is known that factory X supplies 60% of the total bulbs available. What is the chance that a purchased bulb will work for longer than 5000 hours?

Law of Total Probability Example

- Solution: $$\Pr(A) = \Pr(A \mid B_1)\Pr(B_1) + \Pr(A \mid B_2)\Pr(B_2)$$
  $$= \frac{99}{100}\cdot\frac{6}{10} + \frac{95}{100}\cdot\frac{4}{10} = \frac{974}{1000}$$
- Thus each purchased light bulb has a 97.4% chance to work for more than 5000 hours;
- The summation can be interpreted as a weighted average, and consequently the probability, Pr(A), is sometimes called "average probability".
- Another example: the probability that it's going to rain tomorrow is 0.3. The probability that someone will go picnic is 0.1 if it rains and 0.5 if not. What is the probability that this person will go picnic tomorrow?

  Answer: $0.1\times0.3 + 0.5\times0.7 = 0.38$

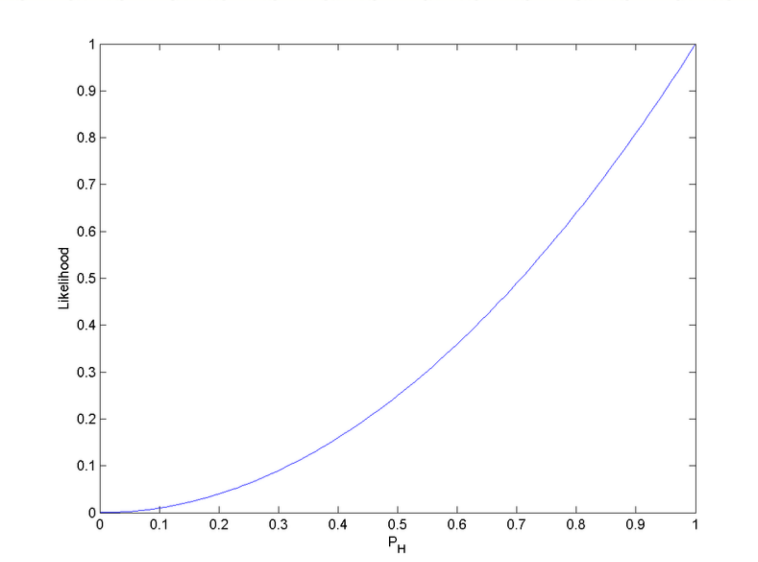


Likelihood

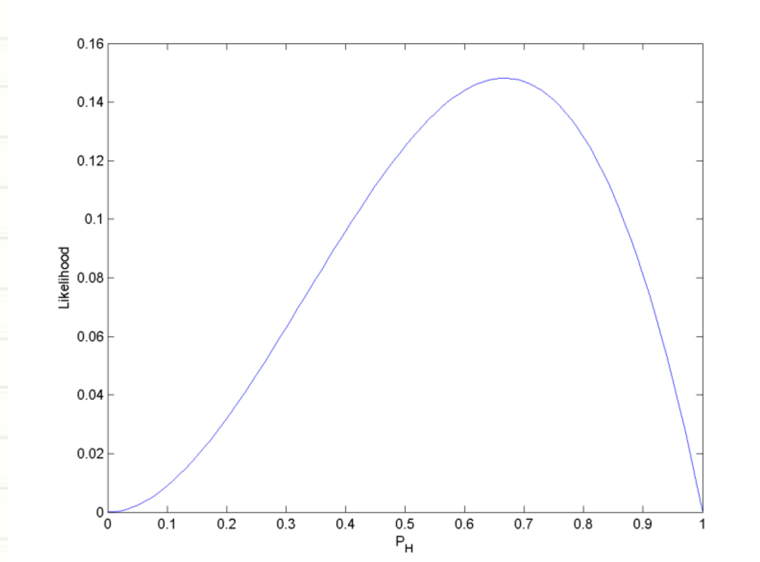
- The likelihood of a set of parameter values, θ, given outcomes x, is equal to the probability of those observed outcomes given those parameter values, that is: $L(\theta \mid x) = \Pr(x \mid \theta)$
- In statistics, *probability* is used when describing a function of the outcome given a fixed parameter value. *Likelihood* is used when describing a function of a parameter given an outcome. For example, if a coin is flipped 100 times and it has landed heads-up 100 times, what is the likelihood that the coin is fair?
- For example: Let P_H be the probability that a certain coin lands heads up (H) when tossed. So, the probability of getting two heads in two tosses (HH) is P_H^2. If $P_H = 0.5$, then the probability of seeing two heads is 0.25: $\Pr(\mathrm{HH} | P_H = 0.5) = 0.25$
- Another way of saying this is that the likelihood that $P_H = 0.5$, given the observation HH, is 0.25, that is: $L(P_H = 0.5 \mid \mathrm{HH}) = \Pr(\mathrm{HH} | P_H = 0.5) = 0.25$

Likelihood (Cont.)

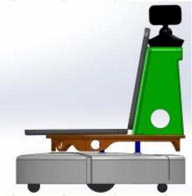
- But this is not the same as saying that the probability that $P_H = 0.5$, given the observation HH, is 0.25. The likelihood that $P_H = 1$, given the observation HH, is 1. However, two heads in a row does not prove that the coin always comes up heads, because HH is possible for any $P_H > 0$.
- A good way to estimate the parameter P_H without additional knowledge is to use the maximum likelihood value.



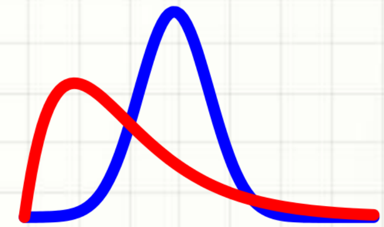
The likelihood function for estimating the probability of a coin landing heads-up without prior knowledge after observing HH



The likelihood function for estimating the probability of a coin landing heads-up without prior knowledge after observing HHT



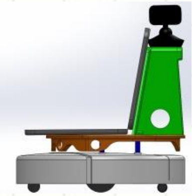
Bayers' Theorem



- Remember? $\Pr(A \cap B) = \Pr(A \mid B)\Pr(B) = \Pr(B \mid A)\Pr(A)$
- Bayers' Theorem: $\Pr(A \mid B) = \dfrac{\Pr(B \mid A)\Pr(A)}{\Pr(B)}$
- For some partition $\{A_j\}$ of the event space, It is sometime useful to compute P(B) using the law of total probability:

$$\Pr(A_i \mid B) = \frac{\Pr(B \mid A_i)\Pr(A_i)}{\sum_j \Pr(B \mid A_j)P(A_j)}$$

- This simple formulation is the foundation for a large field called Bayesian statistics. It is widely used in robot perception and decision making;
- In fact, there has been evidence shown that this is the way how brain works in processing information (sometimes);
- The Bayers' theorem expresses how a subjective degree of *belief* should rationally change to account for new *evidence*;
- I know none of these is making any *sense* to you yet, so let's take a look at a motivating example next.



Example of Bayers' Theorem

- A particular disorder has a base rate occurrence of 1/1000 people. A test to detect this disease has a false positive rate of 5% – that is, 5% of the time it mistakenly report a person to have the disease. Assume that the false negative rate is 0% – the test correctly diagnoses every person who does have the disease. What is the chance that a randomly selected person with a positive result actually has the disease?
- Mathematically, let's call Pr(A) as the probability that a person has the disease without known any test result. Pr(A)=0.001.
- Let's call Pr(B) the probability that a person is tested positive. We need to use the law of total probability here:

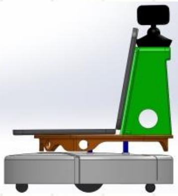
$$\Pr(B) = \Pr(B \mid A)\Pr(A) + \Pr(B \mid \neg A)\Pr(\neg A)$$

$$= 1 \times 0.001 + 0.05 \times 0.999 = 0.05095$$

- Pr(B|A) =1 is the probability of positive test is the person have disease.
- Pr(A|B) can then be calculated with $\Pr(A \mid B) = \dfrac{\Pr(B \mid A)\Pr(A)}{\Pr(B)} \approx 0.02$

Example of Bayers' Theorem (Cont.)

- Although the test is highly accurate, it in fact gives a correct positive result just 2% of the time;
- From a different perspective, considering a population of 10,000 people who are given the test. Just 1/1000th or 10 of those people will actually have the disease and therefore a true positive test result. However, 5% of the remaining 9990 people, or 500 people, will have a false positive test result. So the probability that a person has the disease given that they have a positive test result is 10/510, or around 2%;
- Looking back at the problem, several pieces of information based on past statistics were provided (base occurrence rate, false positive rate, and false negative rate);
- The Bayers' theorem allows us to make reasoning based on both past *knowledge* and new *observations*.



Breaking Up the Bayers' Equation

- Bayer's theorem rewritten: $$\Pr(H \mid E) = \frac{\Pr(E \mid H)\Pr(H)}{\Pr(E)}$$
- *H* is an hypothesis whose probability may be affected by a new observation (evidence). Often there are competing hypotheses, from which one chooses the most probable;
- The evidence *E* corresponds to new data that were not used in computing the prior probability;
- Pr(*H*), the prior probability, is the probability of *H* before *E* is observed. This indicates one's previous estimate of the probability that a hypothesis is true, before gaining the current evidence;
- Pr(*H*|*E*), the posterior probability, is the probability of *H* after *E* is observed;
- Pr(*E*|*H*), the likelihood, is the probability of observing *E* given *H*. It indicates the compatibility of the evidence with the given hypothesis.
- P(*E*) is the same for all possible hypotheses being considered. This means that P(*E*) does not enter into determining the relative probabilities of different hypotheses. Therefore, **posterior is proportional to prior times likelihood**:

$$\Pr(H \mid E) \propto \Pr(E \mid H)\Pr(H)$$

Predicting the Coin Flipping Results

- Imaging we have three types of coins in a bag: Fair Coin (FC) with Pr(Head) = Pr(Tail) = 0.5; Head-heavy Coin (HC) with Pr(Head) = 0.6; and Tail-heavy Coin (TC) with Pr(Head) = 0.40.
- Now let's randomly pick a coin out of a bag. Now, we have three hypothesis for what kind of coin is this: FC, HC, or TC. So how do we know for sure (or at least with a good confidence)?
- Without prior knowledge, we assume the prior probability for each hypothesis is the same Pr(FC) = Pr(HC) = Pr(TC) = 1/3.
- Now we flip the coin once and the result is a head. This is a new evidence. The likelihood of seeing a head when the coin is fair is Pr(Head|FC) = 0.5. Likewise, Pr(Head|HC) = 0.6 and Pr(Head|TC) = 0.4. Use Bayes' theorem, we can find out the posterior probability for the three hypothesis (Next Page).

Coin Flipping Continued

$$\Pr(FC \mid Head) \propto \Pr(Head \mid FC)\Pr(FC) = 0.50 \times \tfrac{1}{3} = \tfrac{1}{6}$$

$$\Pr(HC \mid Head) \propto \Pr(Head \mid HC)\Pr(HC) = 0.60 \times \tfrac{1}{3} = \tfrac{1}{5}$$

$$\Pr(TC \mid Head) \propto \Pr(Head \mid TC)\Pr(TC) = 0.40 \times \tfrac{1}{3} = \tfrac{2}{15}$$

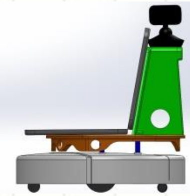
- Since we only have three possible hypothesizes here,

$$\Pr(FC \mid Head) + \Pr(HC \mid Head) + \Pr(TC \mid Head) = 1$$

- Now we have a set of new probability for the hypothesizes after observing a head:

$$\Pr(FC \mid Head) = 0.33;\ \Pr(HC \mid Head) = 0.40;\ \Pr(TC \mid Head) = 0.27$$

- Notice that the three hypothesizes are no longer evenly distributed. If we have to make a guess at this point, we would be guessing the Head-heavy coin although without too much confidence.
- So, what if we flipped the coin again and the result is now a tail? What would be out best guess after this new observation?



Flip the Coin Again

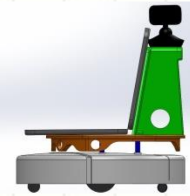
- The posterior probabilities from the previous step is now becoming the prior probabilities for the new step:

$$\Pr(FC\,|\,Tail) \propto \Pr(Tail\,|\,FC)\Pr(FC) = 0.50 \times 0.33 = 0.17$$
$$\Pr(HC\,|\,Tail) \propto \Pr(Tail\,|\,HC)\Pr(HC) = 0.40 \times 0.40 = 0.16$$
$$\Pr(TC\,|\,Tail) \propto \Pr(Tail\,|\,TC)\Pr(TC) = 0.60 \times 0.27 = 0.16$$

- Normalize the values again, we can get a new set of posterior probabilities:

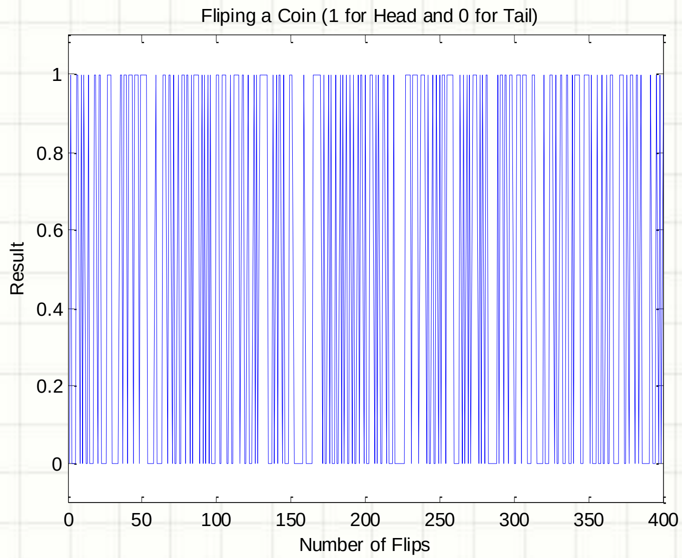
$$\Pr(FC\,|\,Head) = 0.34;\ \Pr(HC\,|\,Head) = 0.33;\ \Pr(TC\,|\,Head) = 0.33$$

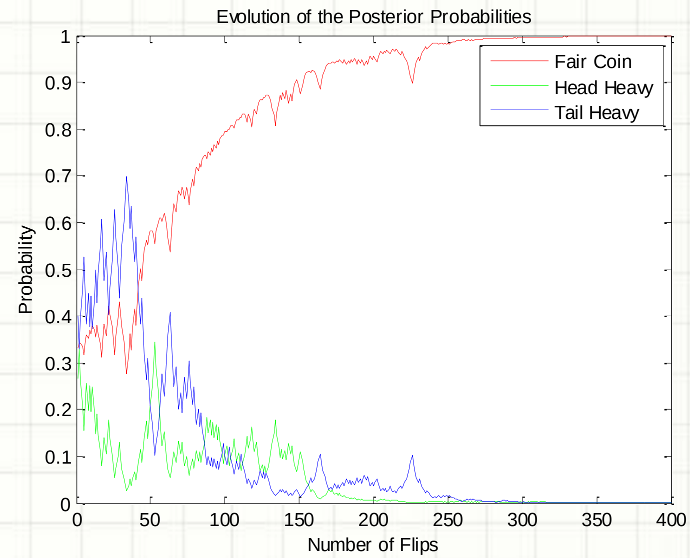
- Now the coin seems to be fair, although we are not quite sure;
- This process can continue as you keep flipping coins, and this is called a recursive process;
- Computers are really good at doing these tedious repetitive calculations;
- Some MATALB results are showing in the next slide.



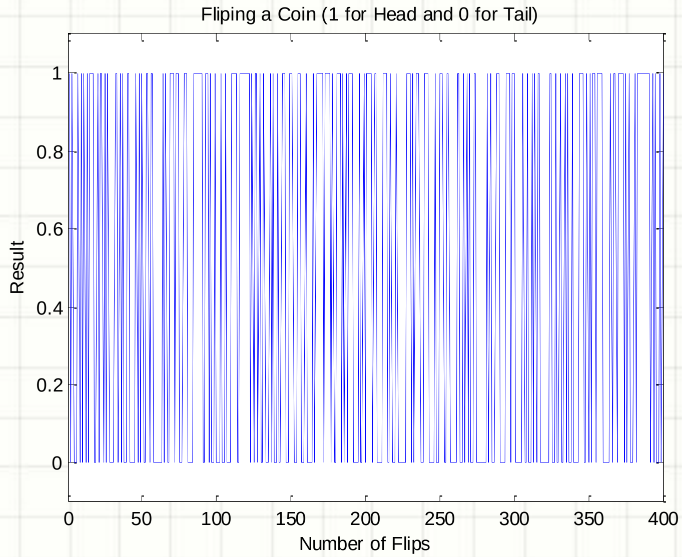
Flipping a Fair Coin for 400 Times

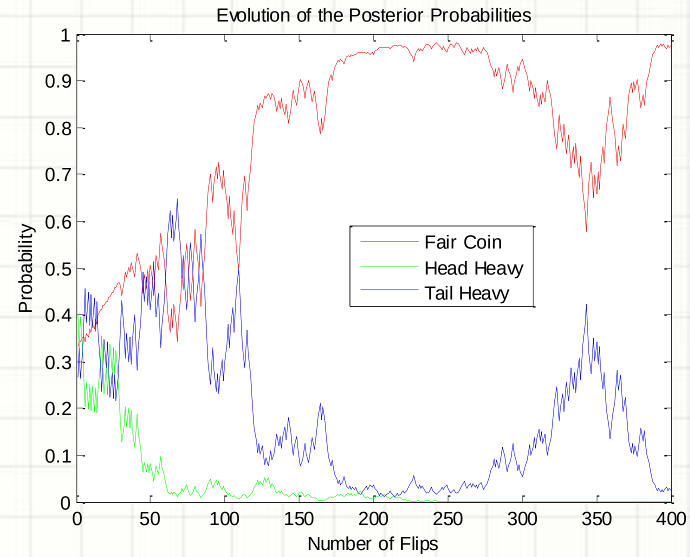
First Try:





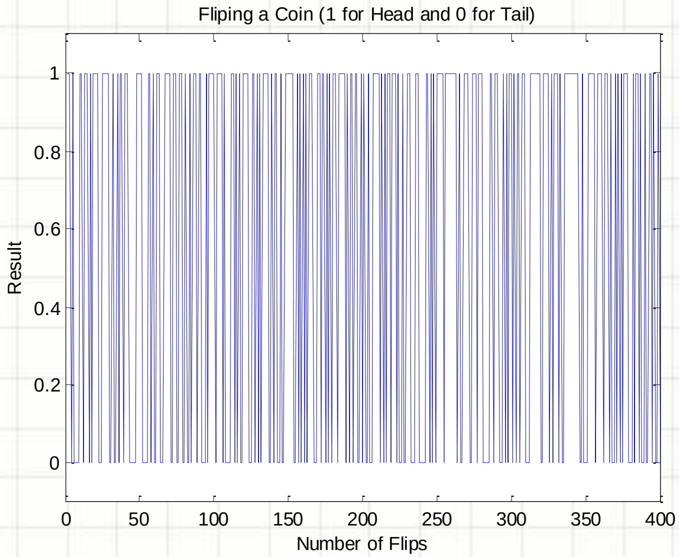
Second Try:

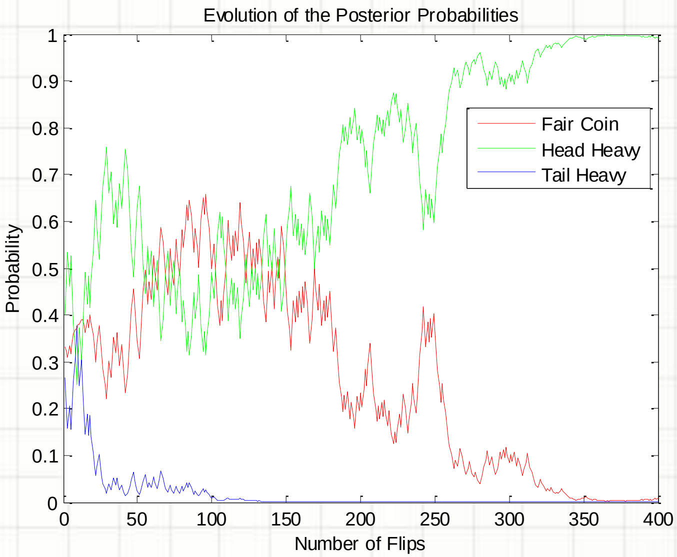




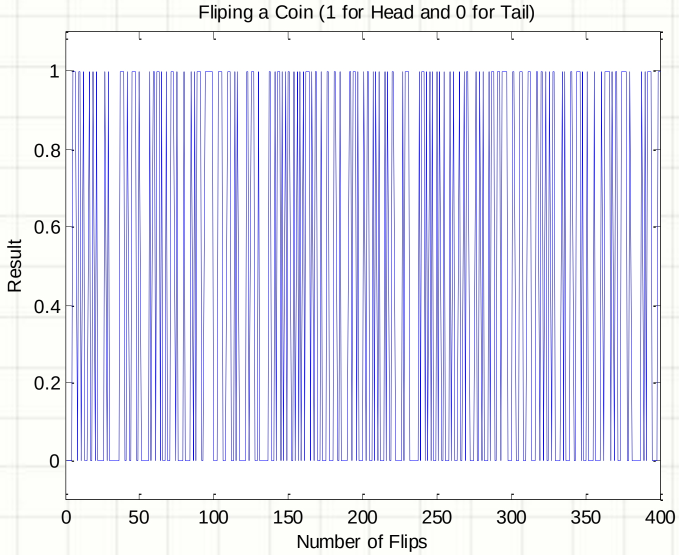
Flipping an Unfair Coin for 400 Times

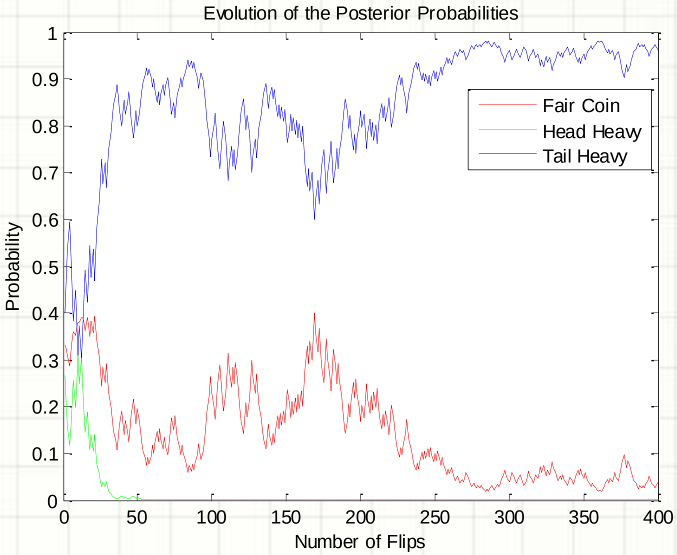
Head Heavy:





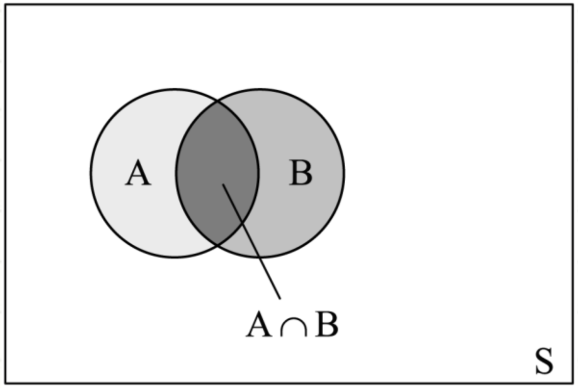
Tail Heavy:

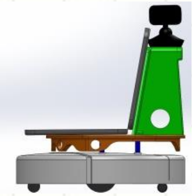




Review of Probability Properties

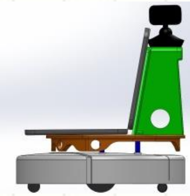
- Not A (A Complement): $\Pr(\neg A) = 1 - \Pr(A)$
- A or B: $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$
- A or B, if A and B are mutually exclusive: $\Pr(A \cup B) = \Pr(A) + \Pr(B)$
- A and B: $\Pr(A \cap B) = \Pr(A \mid B)\Pr(B) = \Pr(B \mid A)\Pr(A)$
- A and B, if A and B are independent: $\Pr(A \cap B) = \Pr(A)\Pr(B)$
- A given B: $\Pr(A \mid B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{\Pr(B \mid A)\Pr(A)}{\Pr(B)}$





Summary

- Uncertainty is everywhere in robotics;
- A robot gets smarter if it can make decisions based on knowledge of the uncertainty;
- Byers' theorem is the foundation for robot (and human) decision making;
- The posterior probability is affected by the prior probability and the likelihood (the probability of seeing the new observation giving the hypothesis).



Further Reading

- Search on Wikipedia the following key words: ‘probability’,‘likelihood’,‘conditional probability’, ‘independent events’, and ‘Bayer’s theorem’;
- Any intro level statistical book such as: “Probability, Statics, and Random Processes for Electrical Engineering” by Alberto Leon-Garcia;
- Bayes for Beginners: http://www.ualberta.ca/~chrisw/BayesForBeginners.pdf