0.1 Probability Basics

0.1.1 Discrete Random Variables

- X denotes a random variable and it can take on a countable number of values in $\{x_1, x_2, ..., x_n\}$
- $P(X = x_i)$ is the probability that the random variable X takes on value x_i
- $P(...)$ is called the **probability mass function**
- e.g. $P(Room) = \langle 0.7, 0.2, ..., 0.02 \rangle$

0.1.2 Continuous Random Variables

- X takes on a value in the continuum
- $P(X = x)$ is the $probability density function$
- $P(x \in (a, b)) = \int_a^b P(x)dx$

0.1.3 Axioms of Probability Theory

- $0 \leq P(a) \leq 1$
- $P(true) = 1$ and $P(false) = 0$
- $P(a \vee b) = P(a) + P(b) - P(a \wedge b)$

0.1.4 Joint and Conditional Probability

- $P(X = x \wedge Y = y) = P(x, y)$
- If X and Y are **independent** then $P(x, y) = P(x)P(y)$
- $P(x|y) = \frac{P(x,y)}{P(y)}$
- $P(x, y) = \frac{P(x|y)}{P(y)}$
- If X and Y are **independent** then $P(x|y) = P(x)$
- $P(x, y|z) = P(x|z)P(y|z)$ means that x and y are **conditionally independent**
- If I know z, I don't need to know x to compute the probability of y.

0.1.5 Law of Total Probability (Discrete)

- $\sum_x P(x) = 1$
- $P(x) = \sum_y P(x, y) = \sum_y P(x|y)P(y)$

0.2 Bayes Rule

- $P(x|y) = \frac{P(y|x)P(x)}{P(y)} = \frac{likelihood * prior}{evidence}$

- Usually $P(y)$ is difficult to compute, so use normalization trick
- $P(x|y) = \eta P(y|x)P(x)$ where $\eta = \frac{1}{\sum_{x \in X} P(y|x)P(x)}$

0.2.1 Casual and Diagnostic Reasoning

- Suppose a robot wants to determine probability of a door being open
- It obtains measurement z. What is the $P(open|z)$
- $P(open|z)$ is **diagnostic** and $P(z|open)$ is **causal**

0.2.2 Conditional Independence Example

- Consider three variables: $RobotLocation, GPSEstimate, LandmarkEstimate$
- $GPSEstimate$ and $LandmarkEstimate$ are NOT independent, $P(GPSEstimate|LandmarkEsti$ $P(GPSEstimate)$
- $GPSEstimate$ and $LandmarkEstimate$ are conditionally independent given $RobotLocation$
- If I know the robot's location, then I can compute the landmark estimate without knowing the GPS estimate

0.3 Bayes Net

- Encode conditional independence relationships in a Bayes Net. Used to describe cause-effect relationships
- Directed and acyclic graph
  - Nodes represent random variables
  - Edges represent conditional dependencies
  - Nodes that are not connected are conditionally independent of each other
  - Node is associated with a probability function $P(X_i|Parents(X_i))$, this is defined by a conditional probability table (CPT)
- Inference ....

0.3.1 Markov Random Fields

- Graph is undirected and may be cyclic

0.3.2 Conditional Random Fields

- Undirected graphical model whose nodes can be divided into exactly two disjoint sets:
  - X: the input variables
  - Y: the observed and output variables

- Used to model the conditional distribution: $P(Y|X)$
- CRFs can be used for object recognition and image segmentation

0.4 Learning a probabilistic model