Probability Theory

Gennaro De Luca, PhD
Arizona State University

The lecture is based on the slides developed by Prof. Yu Zhang from ASU School of Computing and Augmented Intelligence

Ira A. Fulton Schools of Engineering
Arizona State University

Objectives

Objective

Define Probability Space and Conditional Probability

Objective

Discuss Bayes Rule

Uncertainty

General situation:

- **Observed variables (evidence):** Agent knows certain things about the state of the world (e.g., sensor readings or symptoms)
- **Unobserved variables**: Agent needs to reason about other aspects (e.g., where an object is or what disease is present)
- **Model**: Agent knows something about how the known variables relate to the unknown variables

Probabilistic reasoning gives us a framework for managing our beliefs and knowledge.

Probability Space (1/2)

- **A probability space is a triplet ($\mathit{\Omega}$, $\mathscr{B}$, P) that is used to model a process or an experiment with random outcomes.**

- **$\mathit{\Omega}$: The sample space Ω is the set of all possible outcomes of an experiment.**
  - Consider two different experiments:
    - (1) Tossing a coin
    - (2) Tossing a die

Probability Space (2/2)

- $\mathscr{B}$: a σ-algebra (or Borel field), or, informally, a collection of events to consider.
  - **Event:** A subset of Ω, subject to some constraints (e.g., containing the empty set, being closed under complements and countable union)
- P : A measure called "probability" defined on $\mathscr{B}$ that satisfies these conditions:
  - $P(A) \geq 0$ for all $A \in \mathscr{B}$
  - $P(\Omega) = 1$
  - If $A_1, A_2, \ldots \in \mathscr{B}$ are pairwise disjoint, then $P(\cup A_i) = \Sigma\, P(A_i)$

Conditional Probability

Let (Ω, $\mathcal{B}$, P) be a probability space, and let $H \in \mathcal{B}$, with $P(H)>0$.

- For any $B \in \mathcal{B}$, we define:

$$P(B \mid H) = P(B \cap H) \,/\, P(H)$$

and call $P(B \mid H)$ the **conditional probability** of B, given H.

The Total Probability Rule

- **Let (Ω, $\mathcal{B}$, P) be a probability space, and let $\{H_j\}$ be pairwise disjoint events in $\mathcal{B}$ (i.e., let $H_j \cap H_k = \emptyset$, $\forall j \neq k$), and $\cup_{j=1,\ldots,n} H_j = \Omega$. Let $A \in \mathcal{B}$.**
  - Such $\{H_j\}$ is called a **partition of Ω** and is finite or countably infinite.
  - Suppose $P(H_j)>0$, $\forall j$, then the total probability rule states:

$$P(A) = \Sigma_{j=1,\ldots,n} P(A \cap H_i)$$

$$P(A) = \Sigma_{j=1,\ldots,n} P(A \mid H_i)\mathrm{P}(H_i)$$

The Product Rule

- **Sometimes we have conditional distributions but we might want the joint distribution.**
  - Note: $P(x, y) = P(x \cap y)$

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

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

The Chain Rule

- More generally, we can always write any joint distribution as an incremental product of conditional distributions.

$$P(x_1,x_2,x_3) = P(x_1)P(x_2 \mid x_1)P(x_3 \mid x_1,x_2)$$

$$P(x_1,x_2,\ldots,x_n) = \prod_i P(x_i \mid x_1,\ldots,x_{i-1})$$