Review of Probability

ECE57000: Artificial Intelligence

David I. Inouye

Sep. 16, 2019

Why probability?
Probability is useful for handling *uncertainty*

- Inherent stochasticity
  - Quantum mechanics
  - Card games

- Incomplete observability
  - “Let’s Make a Deal” game show of three doors (called “Monty Hall” problem)

- Incomplete modeling
  - Discretization of space for object locations

Why probability?
Sometimes more practical than deterministic

- “Most birds fly”

- “Birds fly, except for very young birds that have not yet learned to fly, sick or injured birds that have lost the ability to fly, flightless species of birds including the cassowary, ostrich and kiwi…”
  - (Example from Deep Learning, Goodfellow et al., 2016, Ch. 3)

Why probability?
Can be seen as an extension of formal logic rules

- Original AI systems based on formal logic and reasoning
  - Chess
  - TurboTax

- Many AI applications based on deterministic logic were too brittle and failed often
  - Traditional linguistic approaches to natural language processing

- Modern AI systems almost always rooted in probability
  - Computer vision
  - Speech recognition
  - Natural language processing

How are these statements similar or different?

- A boardgame player: “The probability of getting a heads when flipping a fair coin is 50%.”
- The weather forecaster: “The probability of rain tomorrow is 50%.”
- Your doctor after examining your symptoms: “The probability of you having the flu is 50%.”

Frequentist and Bayesian interpretations lead to the same set of axioms

- Frequentist
  - Related to rates that events occur under repeated experimentation
- Bayesian interpretation
  - "Degree of belief"
- Pragmatic interpretation
  - They lead to the same math and are useful in similar circumstances
  - Use whichever interpretation is most useful

A <u>random variable</u> maps outcomes/events of a random/uncertain process to numbers

- Flipping a coin
  - Outcomes: {“Heads”, “Tails”}
  - Possible random variables: (show on board)

- Flipping two coins
  - Outcomes: {(H,H), (H,T), (T, H), (T, T)}
  - Possible random variables: # heads, # tails, same, different

- Flipping coins until you get one tails
  - Outcomes: ?
  - Random variables: ?

A **random variable** maps outcomes to numbers: Defining a random variable is the first step

- Random Tweet
  - Outcomes: ?
  - Random variables: ?

- Random Instagram image
  - Outcomes: ?
  - Random variables: ?

Random variables can be <u>discrete</u> or <u>continuous</u>

- Discrete
  - Values are in some finite set or countably infinite set
  - $\{-1,\ \ 1\}, \{5, 10, -20, 3\}, \{0, 1, 2, \dots\}, \mathbb{Z},$
- Continuous
  - Values associated with intervals of $\mathbb{R}$
  - $[0,1], [-1, 1], [0.5, 1] \cup [-1, 0.5], \mathbb{R}_+ \equiv [0, \infty)$
- These can be unified under the study of *measure theory* but that is out of scope
- *Note: Random variables by themselves do not provide any probability information.*

<u>Probability distributions</u> attach probabilities to all possible values of a random variable

- **<u>Probability mass functio</u>**n (PMF) is used for *discrete* random variables
- A PMF P for random variable X that satisfies the following:
  1. Domain of P must include all possible states of X
  2. Unit domain: $\forall\, x \in X, 0 \leq P(x) \leq 1$
  3. Sum to 1: $\sum_{x \in X} P(x) = 1$
  - (Write on board)

Let the random variable X be 0 if heads and 1 if tails (could be unfair coin)

- Are the following functions valid PMFs? Why?
- $P(x = 0) = 0.2, P(x = 1) = 0.9$
- $P(x) = 3x^2$
- $P(x) = x^2$
- $P(x) = \log x$

Probability distributions attach probabilities to all possible values of a random variable

- *Probability density function* (PDF) is used for *continuous* random variables
- A PDF p for random variable X that satisfies the following:
  1. Domain of p must include all possible states of X
  2. Non-negative: $\forall\, x \in X, p(x) \geq 0$ ** $p(x)$ **could be greater than 1**
  3. Integrate to 1: $\int_{x \in X} p(x) = 1$
- $p(x)$ is NOT a probability, rather ***integrating*** the PDF gives probabilities over **sets**

Suppose $X \in (0,1)$ (note: 0 is not included)

- Are the following functions valid PDFs? Why?
- $\forall x \in (0, 0.5), p(x) = 2;\ \ \forall x \notin (0, 0.5), p(x) = 0$
- $p(x) = 3x^2$
- $p(x) = -\log x$

Next time: Joint, marginal and conditional probability, chain rule, independence, expectations