Conference 1 Micro Theory 250D2

Jacob Hazen

material is not 100% my ideas
jacobhazen1.github.io/

5 February 2024

Random Variable

- Probability is about random variables (r.v.)
- A random variable is any “probabilistic” outcome.
  - Flip of coin
  - Height of someone randomly chosen in a population

Sample Space & Atoms

- r.v. take on values in a sample space discrete or continuous
- For example:
  - The sample space of a toss of a coin is $coin \in \{H, T\}$
  - Height of person: $person \in \{0, \infty\}$
- We call the values atoms

Discrete Distributions

- A discrete distribution assigns a probability to every atom in the sample space:
  - E.G. let $X \in \{H, T\}$ now

$$P(X = H) = 0.5$$
$$P(X = T) = 0.5$$

- Note: the entire state space **must** sum to 1

$$\sum_x P(X = x) = 1$$

- Now is a good time to introduce an Event (X) and a Outcome (x).
- An Event is simply a subset of a sample space (S), in formal terms, $X \subseteq S$.
- Sum over part of the Discrete Space Dice:
- $P(D>3) = P(D=4) + P(D=5) + P(D=6)$

Probability Theory : Bayes Rule

$$P(A \text{ and } B|B) = P(A \cap B|B) = P(A, B|B) = P(A|B) = \frac{P(A, B)}{P(B)}$$

Note: The first two formula's starting from the left is just to for your notation. Specifically $P(A, B|B) = P(A|B)$ in the literature people remove the B because it's standard/redundant, but it is important to understand the relationship.

$$P(A,B|B) = \frac{P(A,B)}{P(B)}$$
$$P(A,B|B)P(B) = P(A,B)$$

Now Note:

$$P(A,B|A) = \frac{P(A,B)}{P(A)}$$
$$P(A,B|A)P(A) = P(A,B)$$

Thus Bayes theorem is:

$$P(A, B|A) = \frac{P(A, B|B)P(B)}{P(A)}$$

Or in the standard notation where
$P(A, B|B) = P(A \cap B|B) = P(A \mid B)$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Conditional Distributions

$$P(X = x, Y = y | Y = y) = P(x|y) = \frac{P(x, y)}{P(y)}$$

Chain Rule

$$P(X, Y) = P(X, Y)\frac{P(Y)}{P(Y)} = P(X, Y|Y)P(Y)$$

Then in General:

$$\begin{aligned} P(X_1, \ldots, X_N) &= \prod_{n=1}^{N} P(X_n|X_1, \ldots, X_{n-1}) \\ &= \prod_{n=1}^{N} P(X_n| \cap_{j=1}^{n-1} X_j) \end{aligned}$$

Marginalization

- Given a collection of R.V. we are only interested in a subset of them
- Lets compute $P(X)$ from joint distribution $P(X, Y, Z)$
- Marginalization allows us to compute this:

$$P(X) = \sum_y \sum_z P(X, Y = y, Z = z)$$

Sum over all the possible values that Y and Z can take in the joint distribution of X, Y, Z gives us $P(X)$

Continuous

$$f_X(x) = \int_a^b \int_c^d f(x, y, z) dxdz$$

Where $x \in [a, b]$ and $y \in [c, d]$

Deriving above from Chain Rule:

$$\begin{aligned}\sum_y \sum_z P(x,y,z) &= P(y,z|x)P(x) \\ &= P(x)\sum_y \sum_z P(y,z|x) \\ &= P(x)\end{aligned}$$

Note: $\sum_y \sum_z P(y,z|x) = 1$

Bayes Theorem using Marginalization and Chain Rule

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}$$

$$= \frac{P(X|Y)P(Y)}{\sum_{y} P(X|Y=y)P(Y=y)}$$

Ok, this version of Bayes is a lot to take in but it is **KEY** to understanding how we can use the observable data to uncover the hidden distribution.

- $P(Y)$ is the prior distribution, in other words, the distribution we know already from the data.
- $P(Y|X)$ is the posterior distribution, in other words, we update future knowledge using our prior knowledge of the distribution (i.e $P(Y)$) given the new observed data X

Next is the $P(X)$, our data observed data comes in terms of data points, how do we estimate $P(X)$ from the data.

- Using the rules above $\sum_{y} P(X|Y=y)P(Y=y) = P(X)$, in other words, since we observe, $P(X|Y)$ using the chain rule + marginalization above we can get $P(X)$

Example

Let Y be a disease and X be a symptom. From the distribution of the symptom given the disease $P(X|Y)$ and the probability of the disease $P(Y)$, we can compute the (more useful) distribution of the disease given the symptom $P(Y|X)$.