Lecture 1: Overview of Probability
Monday, January 13, 2020

0.1 Probability in a Nutshell

We will begin with an object called a random variable ("RV") A. It can take many different values each time the experiment is run, and the value given follows some distribution: $A \in \{a_1, \cdots, a_n\} \mapsto \mathrm{Pr}_A(a)$, the probability that the RV A takes the value a.

Additionally, we want the probabilities to be normalized: $\sum_a \mathrm{Pr}_A(a) = 1$.

Now suppose we have two random variables, A and B.

Definition 0.1.1 (Joint Probability)**.** $\mathrm{Pr}_{A,B}(a, b)$ is the probability that A takes the value a and B takes the value b.

Definition 0.1.2 (Marginal Probability)**.** $\mathrm{Pr}_A(a) = \sum_b \mathrm{Pr}_{A,B}(a, b)$

Definition 0.1.3 (Conditional Probability)**.** $\mathrm{Pr}(a \mid b)$ is the probability of a *given* b, since it is conceivable that the random variable B is informative about the outcome of A or vice versa.

Example. Suppose we roll a die, and $A \in \{1, 2, 3, 4, 5, 6\}$ while $B \in \{\text{even}, \text{odd}\}$.

$$\mathrm{Pr}(\text{roll a 6}) = \frac{1}{6}$$

so

$$\mathrm{Pr}(\text{roll a 6} \mid \text{even}) = \frac{1}{3}$$

and

$$\mathrm{Pr}(\text{even} \mid \text{roll a 6}) = 1$$

◇

It's intuitive that

$$\Pr_{A,B}(a, b) = \mathrm{Pr}(a \mid b) \Pr_B(b) = \mathrm{Pr}(b \mid a) \Pr_A(a)$$

Therefore

$$\mathrm{Pr}(a \mid b) = \frac{\mathrm{Pr}(b \mid a)\, \mathrm{Pr}_A(a)}{\mathrm{Pr}_B(b)} \qquad \text{(Bayes' Theorem)}$$

0.1.1 Statistical Independence

We call a pair of random variables statistically independent if the joint probabilities factor into the marginals:

$$\Pr_{A,B}(a, b) = \Pr_A(a) \Pr_B(b) \iff A \text{ and } B \text{ are statistically independent}$$

In some ways, statistical independence is a good thing, since it makes calculations technically easier. Unfortunately, it's typically boring, since this means A is not informative about B and vice versa, so knowing things about one variable will tell you nothing about the other.

If two RVs are statistically independent,

$$\Pr(a \mid b) = \frac{\Pr_{A,B}(a,b)}{\Pr_B(b)} = \frac{\Pr_A(a)\cancel{\Pr_B(b)}}{\cancel{\Pr_B(b)}} = \Pr_A(a)$$

Now suppose we have more than two RVs, $\{A_1, A_2, \cdots, A_N\}$.

Definition 0.1.4 (Pairwise Independence).

$$\Pr_{A_i,A_j}(a_i, a_j) = \Pr_{A_i}(a_i)\Pr_{A_j}(a_j) \quad i \neq j$$

Definition 0.1.5 (Mutual Independence).

$$\Pr_{A_1,A_2,\ldots,A_N}(a_1, a_2, \cdots, a_N) = \Pr_{A_1}(a_1)\Pr_{A_2}(a_2)\cdots\Pr_{A_N}(a_N)$$

Note

Mutual independence obviously implies pairwise independence. However, pairwise independence **does not** imply mutual independence!

0.2 Functions of Random Variables

Suppose we have a random variable A with probability $\Pr_A(a)$. We can use a function to generate a new random variable, $F(A)$. What is $\Pr_F(f)$, the probability that our new random variable takes a value f?

$$\Pr_F(f) = \sum_a \delta_{f,F(a)} \Pr_A(a) \qquad \text{(Transformation Theorem for Probabilities)}$$

To understand this, notice that, for example, if $A \in \{-3, -2, -1, 0, 1, 2, 3\}$ and $F(x) = x^2$, there are now two ways to get $f = 4$ (namely, $a = -2$ or $a = 2$). Therefore, we have to look at both of these a values (the δ-function) and of course we must also consider the probability to get each a.

0.3 Expectation Values

Definition 0.3.1 (Expectation Value).

$$\langle A \rangle_{\Pr_A} = \sum_a \Pr_A(a)$$

is called the expectation value of A, and is simply a weighted average over all the possible values A can take.

$\langle \cdots \rangle$ acts like a linear operator:

$$\langle \alpha A + \beta B \rangle = \alpha \langle A \rangle + \beta \langle B \rangle$$