STSCI 5080
Probability Models and Inference

Lecture 3: Independence and Random Variables

August 30, 2018

Review of Lecture 2

Property D

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Law of total probability

If $0 < P(B) < 1$, then

$$P(A) = P(A \mid B)P(B) + P(A \mid B^c)P(B^c).$$

Bayes rule

If $P(A) > 0$ and $P(B) > 0$, then

$$P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A)}.$$

Beyes rule as a backward calculation

- Suppose that the logical flow is that:

  $$B \text{ or } B^c \text{ occurs} \Rightarrow A \text{ occurs.}$$

  In such cases, you can compute $P(A \mid B), P(A \mid B^c)$, and $P(B)$ directly.
- The law of total probability tells you how to compute $P(A)$.
- The Bayes rule is a sort of backward calculation:

  given that A occurs, what is the probability that B occured before A?

Generalized law of total probability

Theorem

Let $B_1, \ldots, B_n$ form a partition of Ω in the sense that (1) $B_i \cap B_j = \emptyset$ for each $i \neq j$ and (2) $\bigcup_{i=1}^{n} B_i = \Omega$. Suppose that $P(B_i) > 0$ for all $i = 1, \ldots, n$. Then for any event A,

$$P(A) = \sum_{i=1}^{n} P(A \mid B_i) P(B_i).$$

Independence

Definition

Events A and B are said to be independent if $P(A \cap B) = P(A)P(B)$.

Independence via conditional probability

Theorem

If $P(B) > 0$, then A and B are independent if and only if $P(A \mid B) = P(A)$.

$P(A \mid B) = P(A) \leftrightarrow$ event B does not affect how likely A occurs.

Caveat

If A and B are disjoint, $P(A) > 0$, and $P(B) > 0$, then A and B are NOT independent.

Since $A \cap B = \emptyset$, $P(A \cap B) = P(\emptyset) = 0$. On the other hand, $P(A)P(B) > 0$, and so $P(A \cap B) \neq P(A)P(B)$.

Example 3.1

Example

Consider tossing a coin three times and suppose that all the outcomes occur equally likely. In addition, let

$$A = \text{the first toss is a head} = \{hhh, hht, hth, htt\},$$
$$B = \text{the second toss is a head} = \{hhh, hht, thh, tht\}.$$

In this case, $P(A) = P(B) = 1/2$. Further, since $A \cap B = \{hhh, hht\}$, we have $P(A \cap B) = 1/4$, so that

$$P(A \cap B) = P(A)P(B).$$

Namely, A and B are independent.

Generalization

Definition

Events $A_1, \ldots, A_n$ are independent if for any subset S of $\{1, \ldots, n\}$,

$$P\left(\bigcap_{i \in S} A_i\right) = \prod_{i \in S} P(A_i).$$

For example, three events A_1, A_2, A_3 are independent if

$$\begin{cases} P(A_1 \cap A_2) = P(A_1)P(A_2) \\ P(A_1 \cap A_3) = P(A_1)P(A_3) \\ P(A_2 \cap A_3) = P(A_2)P(A_3) \\ P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3) \end{cases}.$$

Example 3.2

Example

Consider tossing a coin three times and suppose that all the outcomes occur equally likely. In addition, let

$$A_i = \text{the } i\text{-th toss is a head}, \; i = 1, 2, 3.$$

Then A_1, A_2, A_3 are independent.

Pairwise independence

Definition

Events $A_1, \ldots, A_n$ are pairwise independent if for any $i \neq j$, A_i and A_j are independent, i.e., $P(A_i \cap A_j) = P(A_i)P(A_j)$.

Important!

pairwise independence $\not\Rightarrow$ independence.

Example 3.3

Example

Suppose we toss a coin two times and all the outcomes occur equally likely. Consider

$$\begin{aligned} A_1 &= \text{the first toss is a head} = \{hh, ht\}, \\ A_2 &= \text{the second toss is a head} = \{hh, th\}, \\ A_3 &= \text{the first and second tosses are the same} = \{hh, tt\}. \end{aligned}$$

Since $A_1 \cap A_2 = A_1 \cap A_3 = A_2 \cap A_3 = \{hh\}$, we see that A_1, A_2, A_3 are pairwise independent. However,

$$P(A_1 \cap A_2 \cap A_3) = P(\{hh\}) = \frac{1}{4} \neq \frac{1}{8} = P(A_1)P(A_2)P(A_3).$$

So, A_1, A_2, A_3 are not independent.

Chapter 2 Random Variables

Definition

Definition (Random variable)

A random variable X is a function from the sample space Ω to the real numbers, i.e., for every $\omega \in \Omega$, $X(\omega)$ is a real number.

Caution!: A probability measure is a function of subsets of Ω; a random variable is a function of elements of Ω.

If X and Y are random variables, then $X + Y$ is also a random variable:

$$(X + Y)(\omega) = X(\omega) + Y(\omega).$$

Kiyosi Ito (former faculty of Cornell math dept.)

Before Kolmogorov, it was not clear whether $X + Y$ is a random variable if X and Y are random variables.

Kiyosi Ito (1915-2008)

Taken from Wikipedia

Discrete random variable

Definition (Discrete random variable)

A random variable X is discrete if X takes values in a finite or countably infinite set.

Example 3.4

Example

Suppose that we toss a coin three times:
$\Omega = \{hhh, hht, hth, thh, htt, tht, tth, ttt\}$.

$$X(\omega) = \text{total number of heads in } \omega$$

is a discrete random variable taking values in $\{0, 1, 2, 3\}$

Example 3.5

Example

Consider the lifetime (in hours) of a laptop: $\Omega = \{t \mid t \geq 0\}$.

$$X(t) = \text{integer part of } t$$

is a discrete random variable taking values in $\{0, 1, 2, \dots\}$.

Probability mass function

- The probability mass function (pmf) and cumulative distribution function (cdf) describe how a discrete random variable behaves randomly.
- For a discrete random variable X that takes values in $\{x_1, x_2, \dots\}$ (called support of X), define

$$p(x_i) = P(X = x_i), \ i = 1, 2, \dots,$$

where $\{X = x\} = \{\omega \mid X(\omega) = x\}$. In general, we write

$$\{X \in S\} = \{\omega \mid X(\omega) \in S\}$$

for a given subset S of the real numbers.
- In addition, define $p(x) = P(X = x) = 0$ for $x \notin \{x_1, x_2, \dots\}$.

Probability mass function (cont.)

Definition

For a discrete random variable X, the probability mass function (pmf) $p(x)$ is a function defined by

$$p(x) = P(X = x)$$

for any real number x.

The pmf $p(x)$ satisfies that $p(x) \geq 0$ for any x and $\sum_x p(x) = 1$. In addition,

$$P(X \in S) = \sum_{x \in S} p(x).$$