Probability and Statistics

Alvin Lin

Probability and Statistics: January 2017 - May 2017

Independence

Two events A and B are independent if $P(A|B) = P(A)$, and are dependent otherwise. For example, with the case of rolling a die twice, the probability of rolling a 2 the second time given that a 2 occurred in the first roll is equal to the probability of rolling a 2 the second time. The events of rolling a 2 in the first roll and rolling a 2 in the second roll are independent of each other.

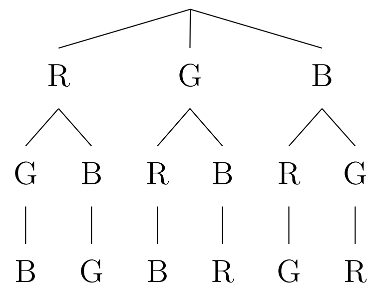
$$P(a\ 2\ in\ the\ second\ roll|a\ 2\ in\ the\ first\ roll) = \frac{1}{6}$$

$$P(a\ 2\ in\ the\ second) = \frac{6}{36} = \frac{1}{6}$$

Example

A basket has 3 balls marked R, B, and G. Select a ball 3 times without replacement.

$$E_1 : a\ G\ in\ the\ first\ and\ a\ G\ in\ the\ second$$
$$E_2 : a\ G\ in\ the\ third$$
$$E_3 : a\ G\ at\ least\ once$$
$$E_4 : a\ G\ in\ the\ first$$
$$E_5 : a\ B\ in\ the\ first$$



$$P(A \cap B) = P(A|B)P(B) \quad = P(A)P(B)(since\ A\ and\ B\ are\ independent)$$

A and B are independent if and only if $P(A \cap B) = P(A)P(B)$.

$$\begin{aligned} P(E_1) &= \frac{0}{6} \\ &= 0 \\ P(E_1 \cap E_2) &= 0 \\ P(E_1 \cap E_2) &= P(E_1)P(E_2) \end{aligned}$$

E_1 and E_2 are independent.

$$\begin{aligned} P(E_4|E_5) &= \frac{0}{2} \\ &= 0 \\ P(E_4) &= \frac{2}{6} \\ P(E_4|E_5) &\neq P(E_4) \end{aligned}$$

E_4 and E_5 are dependent.

$$\begin{aligned} P(E_4 \cap E_3) &= \frac{2}{6} \\ P(E_4) &= \frac{2}{6} \\ P(E_3) &= \frac{6}{6} \\ P(E_4 \cap E_3) = P(E_4)P(E_3) \end{aligned}$$

E_4 and E_3 are independent.

Mutual Independence of Multiple Events

Events $A_1, A_2, \ldots, A_n$ are mutually independent if for every $k(k = 2, 3, \ldots, k)$ and for every subset of indices $i_1, i_2, i_3, \ldots, i_k$ the following is true:

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = P(A_{i_1}) \times P(A_{i_2}) \times \cdots \times P(A_{i_k})$$

Example

Let A_1, A_2, A_3, A_4 be the events of an experiment such that A_1, A_2, A_3, A_4 are mutually independent.

$$\Big[P(A_1 \cap A_2) = P(A_1)P(A_2)$$
$$\wedge P(A_1 \cap A_3) = P(A_1)P(A_3)$$
$$\wedge P(A_1 \cap A_4) = P(A_1)P(A_3)$$
$$\ldots$$
$$\wedge P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3)$$
$$\ldots$$
$$\wedge P(A_1 \cap A_2 \cap A_3 \cap A_4) = P(A_1)P(A_2)P(A_3)P(A_4) \Big]$$

Proposition

If $A_1, A_2, A_3, \ldots, A_n$ are mutually independent, then $A_1', A_2', A_3', \ldots, A_n'$ are mutually indepedent where $A_1' = \textit{not } A_1$.

Consequently, if A_1, A_2, A_3, A_4 are mutually independent, then

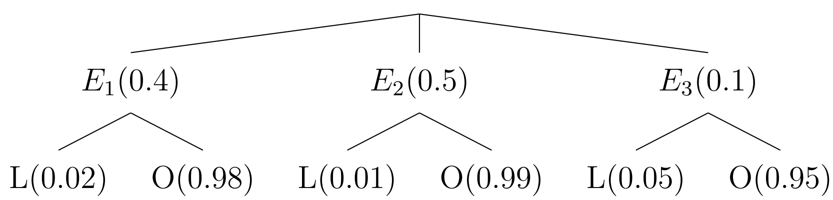
$$P(\textit{not } A_1 \cap \textit{not } A_2 \cap \textit{not } A_3) = P(\textit{not } A_1)P(\textit{not } A_2)P(\textit{not } A_3)$$

Example

1. A certain company sends 40% of its overnight mail parcels via express mail service E_1. Of these parcels, 2% arrive after the guaranteed delivery time (denote the event "late delivery" by L). If a record of an overnight mailing is randomly selected from the company's file, what is the probability that the parcel went via E_1 and was late?

2. Suppose that 50% of the overnight parcels are sent via express mail service E_2. Of those sent via E_2, only 1% arrive late, whereas 5% of the parcels handled by E_3 arrive late. What is the probability that a randomly selected parcel arrives late?

3. If a randomly selected parcel arrived on time, what is the probability that it was not sent via E_1?



1.
$$(0.4)(0.02) = 0.008$$

2.
$$(0.4)(0.02) + (0.5)(0.01) + (0.1)(0.05) = 0.018$$

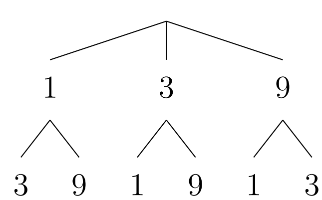
3.
$$\frac{(0.5)(0.99) + (0.1)(0.95)}{(0.4)(0.98) + (0.5)(0.99) + (0.1)(0.95)}$$

Random Variable

For a given sample space S of some experiment, a random variable (rv) is any rule that associates a number with each outcome in S. A rv is a function whose domain is S and whose range is the set of real numbers.

Example

There are 3 balls marked 1, 3, 9 in a basket. Select a ball twice without replacement. Let X be the sum of the two numbers.



$$X(1,3) = 4$$
$$X(3,9) = 12$$

Any random variable whose only possible values are 0 and 1 is called a Bernoulli random variable.

Example

Determine the number of pumps in each of the two six-pump gas stations.

- X = the total number of pumps in use at the two stations
- Y = the difference between the number of pumps in use at station 1 and the number of pumps in use at station 2
- U = the maximum number of pumps in use at the two stations

$$W(observed) = (3,4)$$
$$X(W) = 3 + 4 = 7$$
$$Y(W) = 3 - 4 = -1$$
$$U(W) = max(3,4) = 4$$

If you have any questions, comments, or concerns, please contact me at
alvin@omgimanerd.tech