1. Definition of a random variable?

Let's use a scenario to introduce the idea of a random variable.
Suppose we flip a fair coin three times and record if it shows a head or a tail. The outcome or sample space is $= \{HHH, HHT, HTH, THH, TTT, TTH, THT, HTT\}$. There are eight possible outcomes and each of the outcomes is equally likely. Now, suppose we flipped a fair coin four times. How many possible outcomes are there? There are . How about ten times? possible outcomes! Instead of considering all the possible outcomes, we can consider assigning the variable , say, to be the number of heads in flips of a fair coin. If we flipped the coin times (as above), then X can take on possible values of $0, 1, 2$ or 3. By defining the variable, X, as we have, we created a random variable.

Random Variable

A random variable is a variable that takes on different values determined by chance. In other words, it is a numerical quantity that varies at random.

Types of Random Variables

There are mainly two types of random variables:

A. Discrete Random Variable

A random variable that takes on a finite or countably infinite number of values (see page 4) is called a discrete random variable while one which takes on a non-countably infinite number of values is called a non-discrete random variable. When the random variable can assume only a countable, sometimes infinite, number of values.

B. Continuous Random Variable

When the random variable can assume an uncountable number of values in a line interval. A non-discrete random variable X is said to be absolutely continuous, or simply continuous, if its distribution function may be represented as:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)du \; \{-\infty < x < \infty\} \tag{1}$$

where the function f (x) has the properties

1. $f(x) \geq 0$
2. $\int_{-\infty}^{\infty} f(x)dx = 1$

It follows from the above that if X is a continuous random variable, then the probability that X takes on any one particular value is zero, whereas the interval probability that X lies between two different values, say, a and b, is given by:

$$P(a < X < b) = \int_{a}^{b} f(x)dx \tag{2}$$

Binomial Distribution, Variance, Standard Deviation

Binomial Distribution: It is defined as when there is only one outcome for each trail, that each trail has the same probability of success, and that each trial is mutually exclusive or independent of the other.

Its probability distribution function is given by:

$$f(k, n, p) = Pr(k; n, p) = Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

where

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

2. We have $n = 12$ and $p = 0.5$. Thus, $Var(X) = np(1-p) = 6(1-0.5) = 3$.
The standard deviation is $\sigma_X = \sqrt{npq} = \sqrt{3}$

3. To Calculate the variance for these final exam scores.

24, 58, 61, 67, 71, 73, 76, 79, 82, 83, 85, 87, 88, 88, 92, 93, 94, 97

First find the mean as follows:

$$\bar{x} = \frac{24+58+61+67+71+73+76+79+82+83+85+87+88+88+92+93+94+97}{18} = \frac{233}{18} = 77.67$$

Now use the following formula to compute the variance:

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{18-1} = 306.588$$

Definition (Conditional Probability):

If $P(F) > 0$,we define the probability of E given F as:

$$P(E|F) := \frac{P(E \cap F)}{P(F)} \tag{3}$$

Multiplication Rule

The following formula is called the multiplication rule.

$$P(E \cap F) = P(E|F)P(F) \tag{4}$$

This is simply a rewriting of the definition in Equation (1) of conditional probability.

4. Let A denote the event that the first die shows 4 and B denote the event that the sum of the dice is 7. Notice that for any number the first die shows, there is only one number the second die can show to make the sum 7 (e.g. if the first die shows 5 then the second die must show 2 to make the sum 7). There are six ways in which we can get a 7. They are (1, 6), (6, 1), (2, 5), (5, 2), (4, 3), (3, 4), so the probability of getting a 7 is $\frac{6}{36}$.

So $P(B) = \frac{6}{36}$

Now, if the first die shows 4 there is only one way to make the sum of both dice equal 7 which means $P(A \cap B) = \frac{1}{36}$

Therefore, the probability that the first die shows 4 given that the sum is 7 is:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/36}{1/6} = \frac{1}{6}$$

Independence

Two events are independent if knowledge that one occurred does not change the probability that the other occurred. Informally, events are independent if they do not influence one another.

Translating the verbal description of independence into symbols gives:

A is independent of B if $P(A|B) = P(A)$

That is, knowing that B occurred does not change the probability that A occurred. In terms of events as subsets, knowing that the realized outcome is in B does not change the probability that it is in A.

If A and B are independent in the above sense, then the multiplication rule gives

$$P(A \cap B) = P(A|B)P(B) = P(A) \cdot P(B) \tag{5}$$

This justifies the following technical definition of independence.

Formal definition of independence:

Two events A and B are independent if:

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

This is a nice symmetric definition which makes clear that A is independent of B if and only if B is independent of A. Unlike the equation with conditional probabilities, this definition makes sense even when $P(B) = 0$. In terms of conditional probabilities, we have:

1. If $P(B) \neq 0$ then A and B are independent if and only if $P(A|B) = P(A)$.

2. If $P(A) \neq 0$ then A and B are independent if and only if $P(B|A) = P(B)$.

Bayes' Theorem

Bayes' theorem is a pillar of both probability and statistics and it is central to probability. For two events A and B Bayes' theorem (also called Bayes' rule and Bayes' formula) says:

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)} \tag{6}$$

Comments:

1. Bayes' rule tells us how to 'invert' conditional probabilities, i.e. to find $P(B|A)$ from $P(A|B)$.

2. In practice, $P(A)$ is often computed using the law of total probability.

5. We solve this problem using the Bayes' Theorem as follows:

Let M = Student is Male, F = Student is Female.

Note that M and F partition the sample space of students.

Let T = Student is over 6 feet tall.

We know that $P(M) = 2/5$, $P(F) = 3/5$, $P(T|M) = 4/100$ and $P(T|F) = 1/100$.

We require $P(F|T)$. Using Bayes' theorem we have:

$$P(F|T) = \frac{P(T|F)P(F)}{P(T|F)P(F) + P(T|M)P(M)}$$

$$= \frac{\frac{1}{100} \times \frac{3}{5}}{\frac{1}{100} \times \frac{3}{5} + \frac{4}{100} \times \frac{2}{5}}$$

$$= \frac{3}{11}$$

6. Let D represent the event that the person actually has the disease, and let $+$ represent the event that the test gives a positive signal. We wish to find $P(D|+)$. We are given the following probabilities:

$P(D) = 0.005$

$P(+|D) = 0.99$

$P(+|D^c) = 0.01$

Using Bayes' Rule, we have:

$$P(D|+) = \frac{P(+|D)P(D)}{P(+|D)P(D) + P(+|D^c)P(D^c)} = \frac{(0.99)(0.005)}{(0.99)(0.005) + (0.01)(0.995)} = 0.332$$

7. For $i = 1, \cdots, 5$, let A_i denote the event that the i^{th} microprocessor works. Then $P(all\ 5\ work) = P(A_1 \cap A_2 \cap A_3 \cap A_4 \cap A_5) = P(A_1)P(A_2)P(A_3)P(A_4)P(A_5) = (1 - 0.20)^5 = 0.328$

8(a). Let A = Late Delivery and B = Poor quality of the product. Let $n(A)$ and $n(B)$ be the number of cases in favor of A and B. So $n(A) = 78$ and $n(B) = 40$. Since the total number of cases is 112 (here the complaints is treated as the sample space), hence we have:

$$n(A \cap B) = 118 - 112 = 6$$

Probability of a complaint about both, late delivery an poor quality is:

$$P(A \cap B) = \frac{n(A \cap B)}{Total\ number\ of\ complaints} = \frac{6}{112} = 0.0535$$

8(b). Probability that the complaint is only about poor quality $= \frac{34}{112} = 0.3035$