Probability and Statistics Review

random experiment: a process leading to at least two possible outcomes with uncertainty as to which will occur.

basic outcomes: possible outcomes of a random experiment

sample space: the set of **all** possible outcomes

event: a set of basic outcomes from the sample space; it **occurs** if the random experiment gives rise to one of its basic outcomes.

intersection: for two events *A* and *B* in the sample space *S*, their **intersection**, denoted $A \cap B$ is the set of all basic outcomes in *S* that belong to both *A* and *B.*

more generally, for *K* events $E_1, E_2, \ldots E_K$, their intersection $E_1 \cap E_2 \ldots \cap E_K$ is the set of all basic outcomes that belong to every $E_i \, (i = 1, 2, \ldots K)$.

if the events *A* and *B* have no basic outcomes, they are called **mutually exclusive** and their intersection is said to be the **empty(null) set**.
for *K* events, they are said to be mutually exclusive if every pair of them is mutually exclusive.

union: for two events *A*,*B* in the sample space *S*, their **union**, denoted $A \cup B$ is the set of all basic outcomes in *S* that belong to at least one of these events.

more generally, for the *K* events $E_1, E_2, \ldots E_K$, their union, $E_1 \cup E_2 \ldots \cup E_K$ is the set of all basic outcomes belonging to at least one of these events.

if these *K* events in the sample space *S* satisfy $E_1 \cup E_2 \ldots E_K = S$, the *K* events are said to be **collectively exhaustive.**

complement: for an event *A* in the sample space *S*, the **complement** of *A*, denoted by $\overline{A}$ is the set of basic outcomes of a random experiment belonging to *S* but not in *A*.

Probability Postulates:

1. for any event *A* in the sample space *S* $0 \le P(A) \le 1$

2. let *A* denote an event, and let O_i denote the basic outcomes. Then,

$$P(A) = \sum_{A} P(O_i)$$

3. $P(S) = 1$

Probability Rules

let *A* be an event in the sample space *S*. then

$$P(\overline{A}) = 1 - P(A)$$

decomposition rule:

$$P(B) = P(A \cap B) + P(\overline{A} \cap B)$$

since *A*, $(\overline{A} \cap B)$ are mutually exclusive, nd their union is $A \cup B$, we get the **addition rule:**

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

Conditioning on Information

the chance a particular event **will** occur is often likely to depend on whether or not some other event(s) **have** occured.

we'll approach this issue through the notion of **conditional probability.**

Specifically, let *A* and *B* be two events in the sample space *S*. The **conditional probability** of event *A* given event *B*, denoted by $P(A \mid B)$, is defined as :

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

provided that $P(B) > 0$. Similarly,

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

multiplication rules:

for two events *A*,*B* in the sample space *S*, the probability of their intersection is :

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

and

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

Statistical Independence

Two events *A*,*B* in the sample space *S* said to be **statistically independent** if

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

or, by the multiplication rules

$$P(A|B) = P(A) \quad \text{if} \quad P(B) > 0$$

$$P(B|A) = P(B) \quad \text{if} \quad P(A) > 0$$

in general, the events $E_1, E_2, \ldots E_K$ are statistically independent if

$$P(E_1 \cap E_2 \cap \ldots \cap E_K) = P(E_1)P(E_2)\cdots P(E_K)$$

Bayes theorem

fundamental law for revising probabilities given new information.

recall the multiplication rules of probability:

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

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

equating these two gives:

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

can interpret this in terms of subjective definition of probability. Suppose individual is interested in event *B* and assigns a subjective probability *P(B)*. This is called a **prior** probability.

With additional information, say the occurence of the event *A*, the probability assigned to event *B* may be modified. This new probability is called the **posterior** probability.

Bayes' Theorem is thus a mechanism for updating a prior probability to a posterior probability when additional information becomes available.

Bayes theorem can be generalized to *K* events in the following way. Suppose $E_1, E_2, \ldots E_K$ are muttually exclusive and collectively exhaustive events, and let *A* be some other event.

Suppose for some *i* we want $P(E_i \mid A)$. This can be evaluated with the previous formula, but the denominator can be expressed differently.

Recall that:

$$P(A) = P(E_1 \cap A) + P(E_2 \cap A) + \ldots + P(E_K \cap A)$$

and the multiplication rule:

$$P(E_j \cap A) = P(A \mid E_j)P(E_j) \qquad (j = 1, 2, \ldots K)$$

we thus get :

$$P(A) = P(A \mid E_1)P(E_1) + P(A \mid E_2)P(E_2) + \ldots P(A \mid E_K)P(E_K)$$

Thus we can restate Bayes Theorem as :

$$P(E_i \mid A) = \frac{P(A \mid E_i)P(E_i)}{P(A \mid E_1)P(E_1) + \ldots P(A \mid E_K)P(E_K)}$$

e.g. Test for type of cancer.

$$P(+ \mid C) = 0.95 \qquad P(- \mid C) = 0.05$$

$$P(+ \mid C^c) = 0.05 \qquad P(- \mid C^c) = 0.95$$

$$P(C) = 0.00001$$

Is this a good test?

$$P(C \mid +) = \frac{0.95 * 0.00001}{0.95 * 0.00001 + 0.05 * 0.99999} = 1.9 * 10^{-4}$$

Random Variables

a **random variable** is a variable that takes on numerical values determined by the outcome of a random experiment.

notationally, we will let capital letters denote (*X)* denote the random variable and the corresponding lowercase (x) denote the possible values it can take.

we classify random variables with respect to how many values they can take:

a random variable is **discrete** if it can take on no more than a countable number of values.

a random variable is **continuous** if it can take any value in an interval.

Probability Distributions for Discrete Random Variables

the **probability distribution** of a random variable is any representation of the probabilities of all the possible outcomes.

the **probability function,** or **probability mass function**, denoted by $P_X(x)$, of a discrete random variable X, expresses the probability that X takes the value x as a function of x:

$$P_X(x) = P(X = x)$$

Properties of Probability Functions

1. $P_X(x) \geq 0$ for all values of x.

2. $$\sum_x P_X(x) = 1$$