**Definition 1** The sample space is the collection of all possible different outcomes of an experiment.

**Definition 2** A point in the sample space is a possible outcome of an experiment.

**Definition 3** An event is any set of points in the sample space.

**Definition 4** If \( A \) is an event associated with an experiment, and if \( n_A \) represents the number of times \( A \) occurs in \( n \) independent repetitions of the experiment, the probability of the event \( A \), denoted by \( P(A) \), is given by

\[
P(A) = \lim_{n \to \infty} \frac{n_A}{n}
\]  

**Definition 5** A probability function is a function that assigns probabilities to the various events in the sample space.

**Definition 6** If \( A \) and \( B \) are two events in a sample space \( S \), the event "both \( A \) and \( B \) occur" representing those points in the sample space that are in both \( A \) and \( B \) at the same time, is called the joint event \( A \) and \( B \) and is represented by \( AB \). The probability of the joint event is represented by \( P(AB) \).

**Definition 7** The conditional probability of \( A \) given \( B \) is the probability that \( A \) occurred given that \( B \) occurred and is given by

\[
P(A|B) = \frac{P(AB)}{P(B)}
\]  

where \( P(B) > 0 \).

**Definition 8** Two events \( A \) and \( B \) are independent if

\[
P(AB) = P(A)P(B)
\]  

**Definition 9** Two experiments are independent if for every event \( A \) associated with one experiment and every event \( B \) associated with the second experiment, \( P(AB) = P(A)P(B) \).

**Definition 10** \( n \) experiments are mutually independent if for every set of \( n \) events, formed by considering one event from each of the \( n \) experiments, the following equation is true,

\[
P(A_1A_2 \cdots A_n) = P(A_1)P(A_2) \cdots P(A_n)
\]  

where \( A_i \) represents an outcome of the \( i \)th experiment, for \( i = 1, 2, \ldots, n \).
Definition 11 A random variable is a function that assigns real numbers to the points in a sample space.

Definition 12 The conditional probability of $X$ given $Y$, written $P(X = x | Y = y)$, is the probability that the random variable $X$ assumes the value $x$, given that the random variable $Y$ has assumed the value $y$.

Definition 13 The probability function of the random variable $X$, usually denoted by $f(x)$, is the function that gives the probability of $X$ taking the value $x$, for any real number $x$. In other words,

$$f(x) = P(X = x)$$

(5)

The probability function always equals 0 at values of $x$ that $X$ cannot assume.

Definition 14 The distribution function of a random variable $X$, usually denoted by $F(x)$, is the function that gives the probability of $X$ being less than or equal to any real number $x$. In other words,

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t)$$

(6)

where the summation extends over all values of $t$ that do not exceed $x$. Distribution functions are often called cumulative distribution functions (c.d.f. for short) to emphasize their property of representing cumulative probabilities.

Definition 15 Let $X$ be a random variable. The binomial distribution is the probability distribution represented by the probability function

$$f(x) = P(X = x) = \binom{n}{x} p^x q^{n-x}$$

$$x = 0, 1, \ldots, n$$

(7)

where $n$ is a positive integer, $0 \leq p \leq 1$, and $q = 1 - p$. Note that we are using the usual convention that $0! = 1$.

Definition 16 Let $X$ be a random variable. The discrete uniform distribution is the probability distribution represented by the probability distribution represented by the probability function

$$f(x) = \frac{1}{N}$$

$$x = 1, 2, \ldots, N$$

(8)

Definition 17 The joint probability function $f(x_1, x_2, \ldots, x_n)$ of the random variables $X_1, X_2, \ldots, X_n$, is the probability of the joint occurrence of $X_1 = x_1$, $X_2 = x_2$, \ldots, and $X_n = x_n$. Stated differently,

$$f(x_1, x_2, \ldots, x_n) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$$

(9)
Definition 18 The joint distribution function \( F(x_1, x_2, \ldots, x_n) \) of the random variables \( X_1, X_2, \ldots, X_n \), is the probability of the joint occurrence of \( X_1 \leq x_1, X_2 \leq x_2, \ldots, \) and \( X_n \leq x_n \). Stated differently,

\[
F(x_1, x_2, \ldots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n) \tag{10}
\]

Definition 19 The conditional probability function of \( X \) given \( Y \), \( f(x | y) \), is

\[
f(x | y) = P(X = x | Y = y) \tag{11}
\]

Definition 20 Let \( X \) be a random variable. The hypergeometric distribution is the probability distribution represented by the probability function

\[
f(x) = P(X = x) = \frac{\binom{A}{x} \binom{B}{k-x}}{\binom{A+B}{k}} \quad 0 \leq x \leq A, \quad 0 \leq k - x \leq B \tag{12}
\]

where \( A, B, \) and \( k \) are nonnegative integers and \( k \leq A + B \).

Definition 21 Let \( X_1, X_2, \ldots, X_n \) be random variables with the respective probability functions \( f_1(x_1), f_2(x_2), \ldots, f_n(x_n) \) and with the joint probability function \( f(x_1, x_2, \ldots, x_n) \). Then \( X_1, X_2, \ldots, X_n \) are mutually independent if

\[
f(x_1, x_2, \ldots, x_n) = f_1(x_1)f_2(x_2) \cdots f_n(x_n) \tag{13}
\]

for all combinations of values of \( x_1, x_2, \ldots, x_n \).

Definition 22 The number \( x_p \), for a given value of \( p \) between 0 and 1, is called the \( p \)th quantile of the random variable \( X \), if \( P(X < x_p) \leq p \) and \( P(X > x_p) \leq 1 - p \).

If more than one number satisfies the definition of the \( p \)th quantile, we will avoid confusion by adopting the convention that \( x_p \) equals the average of the largest and the smallest numbers that satisfy Definition 22.

Definition 23 Let \( X \) be a random variable with the probability function \( f(x) \) and let \( u(X) \) be a real valued function of \( X \). The expected value of \( u(X) \), written \( E[u(X)] \), is

\[
E[u(X)] = \sum u(x)f(x) \tag{14}
\]

where the summation extends over all possible values of \( X \). If the sum on the right side of Equation 14 is infinite, or does not exist, we say that the expected value of \( u(X) \) does not exist.
Definition 24 Let $X$ be a random variable with the probability function $f(x)$. The mean of $X$, usually denoted by $\mu$, is

$$\mu = E(X)$$ (15)

Definition 25 Let $X$ be a random variable with mean $\mu$ and the probability function $f(x)$. The variance of $X$, usually denoted by $\sigma^2$ or by $\text{Var}(X)$, is

$$\sigma^2 = E[(X - \mu)^2]$$ (16)

Definition 26 Let $X_1, X_2, \ldots, X_n$ be random variables with the joint probability function $f(x_1, x_2, \ldots, x_n)$, and let $u(X_1, X_2, \ldots, X_n)$ be a real valued function of $X_1, X_2, \ldots, X_n$. Then the expected value of $u(X_1, X_2, \ldots, X_n)$ is

$$E[u(X_1, X_2, \ldots, X_n)] = \sum u(x_1, x_2, \ldots, x_n)f(x_1, x_2, \ldots, x_n)$$ (17)

where the summation extends over all possible combinations of values of $x_1, x_2, \ldots, x_n$.

Definition 27 Let $X_1$ and $X_2$ be two random variables with mean $\mu_1$ and $\mu_2$, probability functions $f_1(x_1)$ and $f_2(x_2)$, respectively, and joint probability function $f(x_1, x_2)$. The covariance of $X_1$ and $X_2$ is

$$\text{Cov}(X_1, X_2) = E[(X_1 - \mu_1)(X_2 - \mu_2)]$$ (18)

Definition 28 The correlation coefficient between two random variables is their covariance divided by the product of their standard deviations. That is, the correlation coefficient, usually denoted by $\rho$, between two random variables $X_1$ and $X_2$ is given by

$$\rho = \frac{\text{Cov}(X_1, X_2)}{\sqrt{\text{Var}(X_1)\text{Var}(X_2)}}$$ (19)

Definition 29 A random variable $X$ is discrete if there exists a one-to-one correspondence between the possible values of $X$ and some or all of the positive integers.

Definition 30 A random variable $X$ is continuous if no two quantiles $x_{p_1}$ and $x_{p_2}$ of $X$ are equal to each other, where $p_1$ is not equal to $p_2$. Equivalently, a random variable $X$ is continuous if $P(X \leq x)$ equals $P(X < x)$ for all numbers $x$.

Definition 31 Let $X$ be a random variable. Then $X$ is said to have the normal distribution if the distribution function of $X$ is given by

$$F(x) = P(X \leq x) = \frac{1}{\sqrt{2\pi}\sigma}e^{-1/2[(y-\mu)/\sigma]^2}dy$$ (20)
**Theorem 1** (Central Limit Theorem) Let $Y_n$ be the sum of the $n$ random variables $X_1, X_2, \ldots, X_n$, let $\mu_n$ be the mean of $Y_n$, and let $\sigma_n^2$ be the variance of $Y_n$. Under some general, easily met conditions, as $n$, the number of random variables, goes to infinity, the distribution function of the random variable

$$\frac{Y_n - \mu_n}{\sigma_n}$$

approaches the standard normal distribution function.

**Definition 32** A random variable $X$ has the chi-squared distribution with $k$ degrees of freedom if the distribution function of $X$ is given by

$$F(x) = P(X \leq x) = \begin{cases} \int_0^x \frac{y^{(k/2)-1}e^{-y/2}}{2^{k/2}1(2)^{k/2}}dy & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases} \quad (21)$$

**Theorem 2** Let $X_1, X_2, \ldots, X_k$ be $k$ independent and identically distributed standard normal random variables. Let $Y$ be the sum of the squares of the $X_i$’s.

$$Y = X_1^2 + X_2^2 + \cdots + X_k^2 \quad (22)$$

Then $Y$ has the chi-squared distribution with $k$ degrees of freedom.