Some Important Theorems on Probability:

- Conditional probability of \( A \) given \( B \): \( P(A \mid B) = \frac{P(A \cap B)}{P(B)} \)
- Bayes' Theorem:
  \[
  \frac{P(A \mid B)}{P(B)} = \frac{P(B \mid A)}{P(B)}
  \]
- Binomial coefficient: \( \binom{n}{r} = \frac{n!}{r!(n-r)!} \)

Fundamental Principle of Counting: If an operation consists of \( k \) steps, of which the first can be done in \( n_1 \) ways, for each of these the second step can be done in \( n_2 \) ways, for each of the first two the third step can be done in \( n_3 \) ways, an so forth, then the whole operation can be done in \( n_1 \cdot n_2 \cdot \cdots \cdot n_k \) ways.

Permutations and Combinations:

- Number of permutations of \( n \) distinct objects taken \( r \) at a time: \( nPr = \frac{n!}{(n-r)!} \)

Some Important Theorems on Probability:

- If \( A \subset B \) and \( B \subset C \), then \( A \subset C \)
- If \( A \subset B \), then \( A' \supset B' \) or \( B' \subset A' \)
- \( A \cup (B \cup C) = (A \cup B) \cup C = A \cup B \cup C \)
- \( A \cap B = A \cap C \)
- \( A \cap (B \cap C) = (A \cap B) \cap C = A \cap B \cap C \)
- \( A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \)
- \( A \cup B = A \cup A' = \Omega \)
- \( A \cap B = A \cap A' = \emptyset \)
- The number of combinations of \( n \) distinct objects taken \( r \) at a time is \( \binom{n}{r} = \frac{n!}{r!(n-r)!} \)

Probability distributions:

- Discrete case. A discrete random variable and the function given by \( f(x) = P(X = x) \) for each \( x \) within the range of \( X \) is called the probability distribution of \( X \).

Probability distributions (or probability functions): If \( X \) is a discrete random variable, the function given by \( f(x) = P(X = x) \) for each \( x \) within the range of \( X \) is called the probability distribution of \( X \). The function \( f(x) \) is a probability distribution if and only if \( (1) \ f(x) \geq 0 \), and \( (2) \ \sum f(x) = 1 \).

Distribution functions for discrete random variables: If \( X \) is a discrete random variable, the function given by \( F(x) = P(X \leq x) = \sum_{t \leq x} f(t) \) for \( -\infty < x < \infty \) where \( f(t) \) is the value of the probability distribution of \( X \) at \( t \), is called the distribution function of \( X \). If \( F(x) \) is a distribution function then \( (1) \ F(-\infty) = 0, F(\infty) = 1 \), and \( (2) \ f(x) = F(x) \) for any real numbers \( a \) and \( b \).

If the range of a random variable \( X \) consists of the values \( x_1 < x_2 < x_3 < \cdots < x_n \), then \( f(x_1) = F(x_1) \) and \( f(x_i) = F(x_i) - F(x_{i-1}) \) for \( i = 2, 3, \ldots, n \).

Continuous case. Continuous probability density functions: A function with values \( f(x) \), defined over the set of all real numbers, is called a probability density function of the continuous random variable \( X \) if and only if \( P(a < X < b) = \int_a^b f(x) \, dx \). Where \( f(x) \) has the properties \( (1) \ f(x) \geq 0 \), and \( (2) \ \int_{-\infty}^{\infty} f(x) \, dx = 1 \).

Distribution functions for continuous random variables: If \( X \) is a continuous random variable and the value of its probability density at \( t \) is \( f(t) \), then the function given by \( F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t) \, dt \) is called the distribution function of \( X \). If \( F(x) \) is a distribution function then \( (1) \ F(-\infty) = 0, F(\infty) = 1 \), and \( (2) \ f(a) \leq F(b) \) for any real numbers \( a \) and \( b \).
If \( f(x) \) and \( F(x) \) are the values of the probability density and the distribution of the continuous random variable \( X \) at \( x \), then \( P(a < X < b) = F(b) - F(a) \) for any real \( a \) and \( b \) with \( a \leq b \), and \( f(x) = dF(x)/dx \) where the derivative exists.

**Joint distributions:**

**Joint probability distributions:** If \( X \) and \( Y \) are discrete random variables, the function given by \( f(x,y) = P(X = x, Y = y) \) for each pair of values \((x,y)\) within the range of \( X \) and \( Y \) is called the joint probability distribution of \( X \) and \( Y \). \( f(x,y) \) is a joint probability distribution if and only if (1) \( f(x,y) \geq 0 \) and (2) \( \sum_x \sum_y f(x,y) = 1 \).

**Joint distribution functions:** If \( X \) and \( Y \) are discrete random variables, the function given by \( F(x,y) = P(X \leq x, Y \leq y) = \sum_{s \leq x} \sum_{t \leq y} f(s,t) \) for \( -\infty < x < \infty \) and \( -\infty < y < \infty \) where \( f(s,t) \) is the value of the joint probability distribution of \( X \) and \( Y \) at \((s,t)\), is called the joint distribution function of \( X \) and \( Y \). If \( F(x,y) \) is the value of the joint distribution function of two discrete random variables \( X \) and \( Y \) at \((x,y)\), then (1) \( F(-\infty, -\infty) = 0 \), (2) \( F(\infty, \infty) = 1 \), (3) if \( a < b \) and \( c < d \), then \( F(a,c) \leq F(b,d) \).

**Marginal probability functions:** If \( X \) and \( Y \) are discrete random variables and \( f(x,y) \) is the value of their joint probability distribution at \((x,y)\), the function given by \( g(x) = \sum_y f(x,y) \) for each \( x \) within the range of \( X \) is called the marginal distribution of \( X \). Correspondingly, the function given by \( h(y) = \sum_x f(x,y) \) for each \( y \) within the range of \( Y \) is called the marginal distribution of \( Y \).

**Conditional distribution:** If \( f(x,y) \) is the value of the joint probability distribution of the discrete random variables \( X \) and \( Y \) at \((x,y)\), and \( h(y) \) is the value of the marginal distribution of \( Y \) at \( y \), the function given by \( f(x|y) = \frac{f(x,y)}{h(y)} \) for each \( x \) within the range of \( X \), is called the conditional distribution of \( X \) given \( Y = y \). Correspondingly, if \( g(x) \) is the value of the marginal distribution of \( X \) at \( x \), the function \( w(y|x) = \frac{f(x,y)}{g(x)} \) for each \( y \) within the range of \( Y \), is called the conditional distribution of \( Y \) given \( X = x \).

**Independent random variables:** Suppose that \( X \) and \( Y \) are discrete random variables. If the events \( X = x \) and \( Y = y \) are independent events for all \( x \) and \( y \), then we say that \( X \) and \( Y \) are independent random variables. In such case, \( P(X = x, Y = y) = P(X = x) \cdot P(Y = y) \) or equivalently \( f(x,y) = f(x) \cdot f(y) \). Conversely, if for all \( x \) and \( y \) the joint probability function \( f(x,y) \) can be expressed as the product of a function of \( x \) alone and a function of \( y \) alone (which are then the marginal probability functions of \( X \) and \( Y \)), \( X \) and \( Y \) are independent. If, however, \( f(x,y) \) cannot be so expressed, then \( X \) and \( Y \) are dependent.

**Mathematical expectation:**

**I. Discrete case.** Let \( f(x) \) be the probability distribution of \( X \) then the expected value of \( X \) is \( E(X) = \sum_x x f(x) \). Let \( X \) be a discrete random variable with probability distribution \( f(x) \), the expected value of \( g(X) \) is given by \( E[g(X)] = \sum_x g(x) f(x) \).

**II. Continuous case.** Let \( f(x) \) be the probability density function of \( X \) then the expected value of \( X \) is \( E(X) = \int_{-\infty}^{\infty} x f(x) dx \). Let \( X \) be a continuous random variable with probability density \( f(x) \), the expected value of \( g(X) \) is given by \( E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) dx \).

**Mathematical expectation:**

**I. Discrete case.** If \( X \) is a discrete random variable and \( f(x) \) is the value of its probability distribution at \( x \), the expected value of \( X \) is \( E(X) = \sum_x x f(x) \). The expected value of \( g(X) \) is given by \( E[g(X)] = \sum_x g(x) f(x) \).

**II. Continuous case.** If \( X \) is a continuous random variable and \( f(x) \) is the value of its probability density at \( x \), the expected value of \( X \) is \( E(X) = \int_{-\infty}^{\infty} x f(x) dx \). The expected value of \( g(X) \) is given by \( E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) dx \).

**III. Joint case.** If \( X \) and \( Y \) are discrete random variables and \( f(x,y) \) is the value of their joint probability distribution at \((x,y)\), the expected value of \( g(X,Y) \) is \( E[g(X,Y)] = \sum_x \sum_y g(x,y) \cdot f(x,y) \).

**Some theorems on expectation.**

1. If \( a \) and \( b \) are any constant, then \( E(aX + b) = aE(X) + b \).
2. If \( X \) and \( Y \) are any random variables, then \( E(X + Y) = E(X) + E(Y) \).
3. If \( X \) and \( Y \) are independent random variables, then \( E(XY) = E(X)E(Y) \).

**Moments:**

The \( r \)th moment about the origin. The \( r \)th moment about the origin of a random variable \( X \), denoted by \( \mu'_r \), is the expected value of \( X^r \); symbolically, \( \mu'_r = E(X^r) = \sum_x x^r f(x) \) for \( r = 0, 1, 2, \cdots \).
when $X$ is discrete, and $\mu'_r = E(X^r) = \int_{-\infty}^{\infty} x^r f(x)dx$ when $X$ is continuous. $\mu'_r$ is called the mean of the distribution function of $X$, and it is denoted by $\mu$.

**rth moment of $X$ about the mean.** The rth moment about the mean of a random variable $X$, denoted by $\mu_r$, is the expected value of $(X - \mu)^r$; symbolically, $\mu_r = E[(X - \mu)^r] = \sum_{x}(x - \mu)^r f(x)$ for $r = 0, 1, 2, \cdots$ when $X$ is discrete, and $\mu_r = E[(X - \mu)^r] = \int_{-\infty}^{\infty}(x - \mu)^r f(x)dx$ when $X$ is continuous. $\mu_2$ is called the variance of the distribution of $X$, and it is denoted by $\sigma^2$, $\text{var}(X)$, or $V(X)$; $\sigma$, the positive square root of the variance, is called the standard deviation.

Some theorems on variance.
1. $\sigma^2 = \mu'_2 - \mu^2 = E(X^2) - [E(X)]^2$.
2. If $a$ and $b$ are any constants then $\text{var}(aX + b) = a^2\text{var}(X)$.
3. If $X$ and $Y$ are independent random variables, $\text{var}(X \pm Y) = \text{var}(X) + \text{var}(Y)$.

- **Chebyshev’s inequality.** Suppose that $X$ is a random variable (discrete or continuous) having mean $\mu$ and variance $\sigma^2$, which are finite. Then if $\varepsilon$ is any positive number $P(\{|X - \mu| \geq \varepsilon\} \leq \frac{\sigma^2}{\varepsilon^2}$ or, with $\varepsilon = k\sigma$ $P(\{|X - \mu| \geq k\sigma\}) \leq \frac{1}{k^2}$.

- **Moment generating function.** The moment-generating function of a random variable $X$, where it exists, is given by $M_X(t) = E[e^{tX}] = \sum_x e^{tx} f(x)$ when $X$ is discrete and $M_X(t) = E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f(x)dx$ when $X$ is continuous. $\frac{d^n M_X(t)}{dt^n}\bigg|_{t=0} = \mu'_n$.

Some theorems on moment generating function. If $a$ and $b$ are constant, then
1. $M_{X+a}(t) = E[e^{t(X+a)}] = e^{at}M_X(t)$;
2. $M_{bX}(t) = E(e^{bX}) = M_X(bt)$;
3. $M_{X+Y}(t) = E[e^{t(X+Y)}] = E[e^{tX}e^{tY}]$.

- **Product Moments.**

**Product moment about the origin.** The rth and sth product moment about the origin of the random variables $X$ and $Y$, denoted by $\mu_{r,s}$, is the expected value of $X^rY^s$; symbolically $\mu_{r,s} = E(X^rY^s) = \sum_x\sum_y x^r y^s f(x,y)$ for $r = 0, 1, 2, \cdots$ and $s = 0, 1, 2, \cdots$ when $X$ and $Y$ are discrete. Note that $\mu_{1,0} = E(X)$, which we denote here $\mu_X$, and that $\mu_{0,1} = E(Y)$, which we denote here by $\mu_Y$.

**Product moment about the mean.** The rth and sth product moment about the means of the random variables $X$ and $Y$, denoted by $\mu_{r,s}$, is the expected value of $(X - \mu_X)^r(Y - \mu_Y)^s$; symbolically $\mu_{r,s} = E[(X - \mu_X)^r(Y - \mu_Y)^s] = \sum_x\sum_y(x - \mu_X)^r(y - \mu_Y)^s f(x,y)$ for $r = 0, 1, 2, \cdots$ and $s = 0, 1, 2, \cdots$ when $X$ and $Y$ are discrete. $\mu_{1,1}$ is called the covariance of $X$ and $Y$, and it is denoted by $\sigma_{XY}$, $\text{cov}(X,Y)$, or $C(X,Y)$.

Some theorems on covariance.
1. $\sigma_{XY} = \text{cov}(X,Y) = E(XY) - E(X)E(Y) = \mu'_{1,1} - \mu_X\mu_Y$.
2. If $X$ and $Y$ are independent random variables, then $\sigma_{XY} = \text{cov}(X,Y) = 0$.
3. $\text{Var}(X \pm Y) = \text{Var}(X) + \text{Var}(Y) \pm 2\text{cov}(X,Y)$.

- **Conditional expectation.** If $X$ is a discrete random variable and $f(x|y)$ is the value of the conditional probability distribution of $X$ given $Y = y$ at $x$, then

  **Conditional expectation.** The conditional expectation of $u(X)$ given $Y = y$ is $E[u(X)|y] = \sum_x u(x)f(x|y)$.

  **Conditional mean.** The conditional mean of the random variable $X$ given $Y = y$, is $\mu_{X|y} = E[X|y] = \sum_x x f(x|y)$.

  **Conditional variance.** The conditional variance of the random variable $X$ given $Y = y$ is $\sigma^2_{X|y} = E[(X - \mu_{X|y})^2|y] = E[X^2|y] - \mu^2_{X|y}$.

- **Special Probability Distributions.**

  **The Discrete Uniform Distribution.** A random variable $X$ has a discrete uniform distribution if and only if its probability distribution is given by $f(x) = \frac{1}{k}$ for $x = x_1, x_2, \cdots, x_k$ where $x_i \neq x_j$ when $i \neq j$.

  **The Bernoulli Distribution.** A random variable $X$ has a Bernoulli distribution if and only if its probability distribution is given by $f(x; \theta) = \theta^x(1 - \theta)^{1-x}$ for $x = 0, 1$. $\mu = \theta$, $\sigma^2 = \theta(1 - \theta)$, $\mu'_r = \theta$ for $r = 1, 2, 3, \cdots$. 
The Binomial Distribution. A random variable $X$ has a binomial distribution if and only if its probability distribution is given by $b(x; n, \theta) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}$ for $x = 0, 1, 2, \ldots, n$. $M_X(t) = (1 + \theta(e^t - 1))^n$, $\mu = n\theta$ and $\sigma^2 = n\theta(1 - \theta)$.

The Negative Binomial and Geometric Distribution. A random variable $X$ has negative binomial distribution if and only if its probability distribution is given by $b^*(x; k, \theta) = \binom{x+k-1}{k} \theta^k (1 - \theta)^{x-k}$ for $x = k, k+1, k+2, \ldots$. $b^*(x; k, \theta) = \frac{k}{x} \cdot b(k; x, \theta)$, $\mu = \frac{k}{\theta}$, $\sigma^2 = \frac{k}{\theta^2}(1 - \theta)$. For $k = 1$ we obtain the geometric distribution denoted by $g(x; \theta)$.

The Hypergeometric Distribution. A random variable $X$ has a hypergeometric distribution if and only if its probability distribution is given by $h(x; n, N, M) = \binom{M}{x} \binom{N-M}{n-x} / \binom{N}{n}$ for $x = 0, 1, 2, \ldots, n$, $x \leq M$ and $n - x \leq N - M$. $\mu = n \frac{M}{N}$, $\sigma^2 = nM(N-M)(N-n) / \{N^2(N-1)\}$.

The Binomial Approximation to the Hypergeometric Distribution. As a rule of thumb, if $n$ is not exceed 5 percent of $N$, then we may use binomial probabilities in place of hypergeometric probabilities.

The Poisson Distribution. A random variable $X$ has a Poisson distribution if and only if its probability distribution is given by $p(x; \lambda) = \frac{e^{-\lambda} \lambda^x}{x!}$ for $x = 0, 1, 2 \ldots$ where $\lambda$ is the mean of successes. $\mu = \sigma^2 = \lambda$, $M_X(t) = e^{\lambda(e^t-1)}$.

The Poisson Approximation to the Binomial Distribution. Poisson distribution will provide a good approximation to binomial probabilities when $n \geq 20$ and $\theta \leq 0.05$. When $n \geq 100$ and $n\theta < 10$, the approximation will generally be excellent.

• Measures of Central Tendency
  
  Mean. The mean is equal to the sum of all the values in the data set divided by the number of values in the data set.
  
  Median. The median is the middle score for a set of data that has been arranged in order of magnitude.
  
  Mode. The mode is the most frequently occurring value in a set of values.

• Sampling Distributions.

  Sample Mean and Sample Variance If $X_1, X_2, \ldots, X_n$ are constitute a random sample, then the sample mean is given by $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ and the sample variance is given by $S^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n}$.

  Some theorems on sampling distributions of means.
  
  (1) If $X_1, X_2, \ldots, X_n$ are constitute a random sample from an infinite population with mean $\mu$ and the variance $\sigma^2$, then $E(\bar{X}) = \mu$ and $var(\bar{X}) = \frac{\sigma^2}{n}$.
  
  (2) For any positive constant $c$, the probability that $\bar{X}$ will take on a value between $\mu - c$ and $\mu + c$ is at least $1 - \frac{\sigma^2}{mc^2}$. When $n \to \infty$, this probability approaches 1.
  
  (3) If $X_1, X_2, \ldots, X_n$ are constitute a random sample from an infinite population with mean $\mu$, the variance $\sigma^2$, and the moment-generating function $M_X(t)$, then the limiting distribution of $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ as $n \to \infty$ is the standard normal distribution.
  
  (4) If $\bar{X}$ is the mean of a random sample of size $n$ from a normal population with mean $\mu$ and the variance $\sigma^2$, its sampling distribution is a normal distribution with mean $\mu$ and the variance $\sigma^2/n$.
  
  (5) If $\bar{X}$ is the mean of a random sample of size $n$ taken without replacement from a finite population of size $N$ with mean $\mu$ and the variance $\sigma^2$, the $E(\bar{X}) = \mu$ and $var(\bar{X}) = \frac{\sigma^2}{n} \frac{N-n}{N-1}$.

• The Estimation of Means. If $\hat{\theta}_1$ and $\hat{\theta}_2$ are values of the random variables $\hat{\theta}_1$ and $\hat{\theta}_2$ such that $P(\hat{\theta}_1 < \theta < \hat{\theta}_2) = 1 - \alpha$ for some specified probability $1 - \alpha$, we refer to the interval $\hat{\theta}_1 < \theta < \hat{\theta}_2$ as a $(1 - \alpha)100\%$ confidence interval for $\theta$. The probability $1 - \alpha$ is called the degree of confidence, and the endpoints of the interval are called the lower and upper confidence limits.

  Some theorems on estimation of means.
  
  (1) If $\bar{X}$, the mean of a random sample of size $n$ from a normal population with the known variance $\sigma^2$, is to be used as an estimator of the mean of the population, the probability is $1 - \alpha$ that the error will be less than $z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$.
  
  (2) If $\bar{x}$ is the value of the mean of a random sample of size $n$ from a normal population with the known variance $\sigma^2$, then $\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$ is a $(1 - \alpha)100\%$ confidence interval for the mean of the population.