ST2334 Cheatsheet AY21/22 Sem 1

by Richard Willie

Basic Concepts of Probability

Sample Space
The set of **all possible outcomes** of a statistical experiment. It is represented by the symbol S.

Sample Points
Every outcome in a sample space is called an element or a sample point.

Event
A subset of a sample space.

Simple Event
An event which consists of **exactly one outcome**.

Compound Event
An event which consists of **more than one outcome**.

Sure Event
The sample space itself.

Null Event
An subset of S that contains no elements, denoted with $\emptyset$.

Operations with Events

Let S denote a sample space, A and B are any two events of S.

Union of Events
The union of two events A and B, denoted by $A \cup B$, is the event containing all the elements that belong to A or B or to both. That is,

$$A \cup B = \{x : x \in A \textit{ or } x \in B\}$$

The union of n events $A_1, A_2, \ldots, A_n$ is denoted by

$$\bigcup_{i=1}^{n} A_i = A_1 \cup A_2 \cup \cdots \cup A_n$$
$$= \{x : x \in A_1 \textit{ or } \ldots \textit{ or } x \in A_n\}$$

Intersection of Events
The intersection of two events A and B, denoted by $A \cap B$, is the event containing all the elements that are common to A and B. That is,

$$A \cap B = \{x : x \in A \textit{ and } x \in B\}$$

The intersection of n events $A_1, A_2, \ldots, A_n$ is denoted by

$$\bigcap_{i=1}^{n} A_i = A_1 \cap A_2 \cap \cdots \cap A_n$$
$$= \{x : x \in A_1 \textit{ and } \ldots \textit{ and } x \in A_n\}$$

Complement of Events
The complement of event A with respect to S, denoted by A' or A^C, is he set of all elements of S that are not in A.

$$A' = \{x : x \in S \textit{ and } x \notin A\}$$

Mutually Exclusive Events
Two events A and B are said to be mutually exclusive or mutually disjoint if $A \cap B = \emptyset$. That is, if A and B have no elements in common.

Mutually Exclusive Events
Two events A and B are said to be mutually exclusive or mutually disjoint if $A \cap B = \emptyset$. That is, if A and B have no elements in common.

Basic Properties

1. Identity laws:
$$A \cup \emptyset = A$$
$$A \cap S = A$$
2. Universal bound laws:
$$A \cup S = S$$
$$A \cap \emptyset = \emptyset$$
3. Idempotent laws:
$$A \cup A = A$$
$$A \cap A = A$$
4. Commutative laws:
$$A \cup B = B \cup A$$
$$A \cap B = B \cap A$$
5. Associative laws:
$$(A \cup B) \cup C = A \cup (B \cup C)$$
$$(A \cap B) \cap C = A \cap (B \cap C)$$
6. Distributive laws:
$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$
$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$
7. De Morgan's laws:
$$(A \cup B)' = A' \cap B'$$
$$(A \cap B)' = A' \cup B'$$
8. Absorption laws:
$$A \cup (A \cap B) = A$$
$$A \cap (A \cup B) = A$$
$$A \cup (A' \cap B) = A \cup B$$
$$A \cap (A' \cup B) = A \cap B$$
9. Complement laws:
$$A \cup A' = S$$
$$A \cap A' = \emptyset$$
10. Logical adjacency:
$$(A \cup B) \cap (A \cup B') = A$$
$$(A \cap B) \cup (A \cap B') = A$$

Contained Events
$A \subset B$ if all elements in event A are also in event B. If $A \subset B$ and $B \subset A$, then $A = B$. In this module, it is assumed that contained means proper subset.

Counting Methods

Multiplication Principle
If an operation can be performed in n_1 ways, and for each of these ways a second operation can be performed in n_2 ways, then the two operations can be performed together in $n_1 n_2$ ways.
For k such operations, we have $n_1 n_2 \cdots n_k$ ways.

Addition Principle
If a first procedure can be performed in n_1 ways, and a second procedure in n_2 ways, and that it is not possible to perform both together, then the number ways we can perform either the first or second procedures is $n_1 + n_2$ ways.
For k such operations, we have $n_1 + n_2 + \cdots + n_k$ ways.

Permutation
An arrangement of r objects from a set of n objects, where $r \leqslant n$. Number of permutations of n distinct objects taken r at a time is denoted by

$${}_nP_r = \frac{n!}{(n-r)!}$$

Circular Permutation
The number of permutations of n distinct objects arranged in a circle is

$$(n-1)!$$

Permutation with repetition
Suppose we have n objects such that there are n_1 of one kind, n_2 of second kind, $\ldots$, n_k of a k-th kind, where $n_1 + n_2 + \cdots + n_k = n$. Then the number of distinct permutations is

$${}_nP_{n_1, n_2, \ldots, n_k} = \frac{n!}{n_1! n_2! \cdots n_k!}$$

Combination
Number of ways to select r objects from n objects without regard to the order. Number of combinations of n distinct objects taken r at a time is denoted by

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

Binomial Coefficient

1.
$$\binom{n}{r} = \binom{n}{n-r} \text{ for } r = 0, 1, \ldots, n$$
2.
$$\binom{n}{r} = \binom{n-1}{r} + \binom{n-1}{r-1} \text{ for } 1 \leqslant r \leqslant n$$
3.
$$\binom{n}{r} = 0 \text{ for } r < 0 \textit{ or } r > n$$

Probability

Axioms of Probability
1. $0 \leqslant \Pr(A) \leqslant 1$
2. $\Pr(S) = 1$
3. If $A_1, A_2, \ldots$ are **mutually exclusive** (disjoint) events, i.e., $A_i \cap A_j = \emptyset$ when $i \neq j$, then

$$\Pr(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \Pr(A_i)$$

Inclusion-Exclusion Principle

$$\Pr(A_1 \cup A_2 \cup \cdots \cup A_n) = \sum_{i=1}^{n} \Pr(A_i) -$$
$$\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \Pr(A_i \cap A_j) +$$
$$\sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} \sum_{k=j+1}^{n} \Pr(A_i \cap A_j \cap A_k) -$$
$$\cdots + (-1)^{n+1} \Pr(A_1 \cap A_2 \cap \cdots \cap A_n)$$

Conditional Probability
Probability of A given B

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}, \text{ if } \Pr(B) \neq 0$$
$$\Pr(A|B) = 1 - \Pr(A'|B)$$

Axioms of Conditional Probability
For fixed B, $\Pr(A|B)$ satisfies the following
1. $0 \leqslant \Pr(A|B) \leqslant 1$
2. $\Pr(S|B) = 1$
3. If $A_1, A_2, \ldots$ are **mutually exclusive** (disjoint) events, i.e., $A_i \cap A_j = \emptyset$ when $i \neq j$, then

$$\Pr(\bigcup_{i=1}^{\infty} A_i | B) = \sum_{i=1}^{\infty} \Pr(A_i|B)$$

Multiplication Rule of Probability

$$\Pr(A \cap B) = \Pr(A)\Pr(B|A) = \Pr(B)\Pr(A|B)$$
$$\Pr(A \cap B \cap C) = \Pr(A)\Pr(B|A)\Pr(C|A \cap B)$$

Law of Total Probability
Let $A_1, A_2, \ldots, A_n$ be a **partition** of the sample space S. That is, $A_1, A_2, \ldots, A_n$ are mutually exclusive and exhaustive events such that $A_i \cap A_j \neq \emptyset$ for $i \neq j$ and $\bigcup_{i=1}^{n} A_i = S$.

$$\Pr(B) = \sum_{i=1}^{n} \Pr(B \cap A_i) = \sum_{i=1}^{n} \Pr(A_i)\Pr(B|A_i)$$

Bayes' Theorem
Let $A_1, A_2, \ldots, A_n$ be a partition of the sample space S. Then

$$\Pr(A_k|B) = \frac{\Pr(A_k)\Pr(B|A_k)}{\sum_{i=1}^{n} \Pr(A_i)\Pr(B|A_i)} \text{ for } k = 1, \ldots, n$$

Proof:

$$\Pr(A_k)\Pr(B|A_k) = Pr(A_k \cap B) \tag{1}$$

$$\sum_{i=1}^{n} \Pr(A_i)\Pr(B|A_i) = \sum_{i=1}^{n} \Pr(A_i \cap B) = \Pr(B) \quad (2)$$

Independent Events

Two events A and B are said to be **independent** if and only if

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

Properties

1. Suppose $\Pr(A) > 0, \Pr(B) > 0$. If A and B are independent events, then events A and B **cannot be** mutually exclusive.
2. Suppose $\Pr(A) > 0, \Pr(B) > 0$. If A and B are mutually exclusive events, then events A and B **cannot be** independent.
3. The sample space S and the empty space $\emptyset$ are independent of any event.
4. If $A \subset B$, then A and B are dependent unless $B = S$.
5. If A and B are independent, then so are A and B', A' and B, A' and B'.

Pairwise Independent Events

A set of events $A_1, A_2, \ldots, A_n$ are said to be **pairwise independent** if and only if

$$\Pr(A_i \cap A_j) = \Pr(A_i)\Pr(A_j)$$

for $i \neq j$ and $i, j = 1, \ldots, n$.

Mutually Independent Events

A set of events $A_1, A_2, \ldots, A_n$ are said to be **mutually independent** (or simply **independent**) if and only if for any subset $\{A_{i_1}, A_{i_2}, \ldots, A_{i_k}\}$,

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

Mutually independence implies pairwise independence, but pairwise independence does not imply mutually independence.

The complements of any number of the above events will also be mutually independent with the remaining events.

Concepts of Random Variables

Random Variable

A real-valued function X which assigns a number to every element $s \in S$.

Range space of X, $R_X = \{x : x = X(s),\ s \in S\}$.

Equivalent Events

Let B be an event with respect to R_X, i.e., $B \subset R_X$. If $A = \{s \in S : X(s) \in B\}$, then A and B are equivalent events and $\Pr(A) = \Pr(B)$.

Discrete Probability Distributions

Discrete Random Variable

If the number of possible values of X is **finite** or **countable infinite**, we call X a discrete random variable.

Probability Function

Each value of X has a certain probability $f(x)$, and this function $f(x)$ is called the **probability function** (p.f.) or **probability mass function** (p.m.f.).

The collection of pairs $(x_i, f(x_i))$ is called the probability distribution of X.

It must satisfy the following two conditions

1. $0 \leqslant f(x_i) \leqslant 1$ for all x_i
2. $\sum_{i=1}^{\infty} f(x_i) = 1$

Continuous Probability Distributions

Continuous Random Variable

If R_X, the range space of a random variable X, is an interval or a collection of intervals, then X is a continuous random variable.

Probability Density Function

Let X be a continuous random variable. The **probability density function** (p.d.f.) $f(x)$ satisfy the following conditions:

1. $f(x) \geqslant 0$ for all $x \in R_X$
2. $$\int_{R_X} f(x)\ dx = 1 \text{ or } \int_{-\infty}^{\infty} f(x)\ dx = 1$$

since $f(x) = 0$ for x not in R_X

3. For any $(c, d) \subset R_X$ such that $c < d$,

$$\Pr(c \leqslant X \leqslant d) = \int_c^d f(x)\ dx$$

4. $\Pr(A) = 0$ does **not** necessarily imply $A = \emptyset$

Cumulative Distribution Function

Let X be a random variable (can be discrete or continuous). We define $F(x)$ to be the **cumulative distribution function** (c.d.f.) of the random variable X where

$$F(x) = \Pr(X \leqslant x)$$

Note: c.d.f is a **non-decreasing function**.

CDF for Discrete Random Variables

$$F(x) = \sum_{t \leqslant x} f(t) = \sum_{t \leqslant x} \Pr(X = t)$$

For any $a \leqslant b$,

$$\begin{aligned}\Pr(a \leqslant X \leqslant b) &= \Pr(X \leqslant b) - \Pr(X < a)\\ &= F(b) - F(a^-)\end{aligned}$$

where a^- is the largest possible value of X that is strictly less than a.

CDF for Continuous Random Variables

$$F(x) = \int_{-\infty}^{x} f(t)\ dt$$

The reverse is also true, **if the derivative exists**:

$$f(x) = \frac{dF(x)}{dx}$$

For any $a \leqslant b$,

$$\begin{aligned}\Pr(a \leqslant X \leqslant b) &= \Pr(X \leqslant b) - \Pr(X \leqslant a)\\ &= F(b) - F(a)\end{aligned}$$

Expressing CDF

We can also express a c.d.f. in the following form:

$$F(x) = \begin{cases} 0, & \text{for } x < 0 \\ 0.3, & \text{for } 0 \leqslant x < 1 \\ 0.9, & \text{for } 1 \leqslant x < 2 \\ 1, & \text{for } 2 \leqslant x \end{cases}$$

Expectation

The **mean** or **expected value** of X is denoted by $E(X)$ or μ_X.

Discrete Random Variable

$$\mu_X = E(X) = \sum_i x_i f_X(x_i) = \sum_x x f_X(x)$$

We can also calculate mean using

$$E(X) = \sum_{k=1}^{\infty} \Pr(X \geqslant k)$$

Continuous Random Variable

$$\mu_X = E(X) = \int_{-\infty}^{\infty} x f_X(x)\ dx$$

Function a Random Variable

For any function $g(X)$ of a random variable X with p.f. or p.d.f. $f_X(x)$,

1. if X is discrete, providing the sum exists,

$$E[g(X)] = \sum_x g(x) f_X(x)$$

2. if X is continuous, providing the integral exists,

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\ dx$$

Properties

1. $$E(aX) = aE(X)$$
2. $$E(X + b) = E(X) + b$$

Moment

The special function, $g(x) = x^k$, leads us to the definition of moment. The k-th moment of X is $E(X^k)$.

Variance

The **variance** of X is denoted by $V(X)$ or σ_X^2.

$$\begin{aligned}\sigma_X^2 &= V(X) = E[(X - \mu_X)^2]\\ &= \begin{cases} \sum_x (x - \mu_X)^2 f_X(x), & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x)\ dx, & \text{if } X \text{ is continuous} \end{cases}\end{aligned}$$

We can also calculate variance using

$$V(X) = E(X^2) - [E(X)]^2$$

Properties

1. $$V(X) \geqslant 0$$
2. $$V(aX) = a^2 E(X)$$
3. $$V(X + b) = V(X)$$

Standard Deviation

$$\sigma_X = \sqrt{V(X)}$$

Chebyshev's Inequality

We cannot reconstruct the probability distribution of X from $E(X)$ and $V(X)$. However, we can derive some bounds.

Let X be a random variable, $E(X) = \mu$ and $V(X) = \sigma^2$. Then for **any positive number** k, we have

$$\Pr(|X - \mu| \geqslant k\sigma) \leqslant \frac{1}{k^2}$$

$$\Pr(|X - \mu| < k\sigma) \geqslant 1 - \frac{1}{k^2}$$

A slightly more handy forms of the formulae:

$$\Pr(|X - \mu| \geqslant c) \leqslant \frac{V(X)}{c^2}, \text{ for } c > 0$$

$$\Pr(|X - \mu| < c) \geqslant 1 - \frac{V(X)}{c^2}, \text{ for } c > 0$$

Two-Dimensional Random Variables

(X, Y) is a two-dimensional random variable (also called a random vector), where X, Y are functions assigning a real number to each $s \in S$.

Range Space

$$R_{X,Y} = \{(x, y) : x = X(s),\ y = Y(s),\ s \in S\}$$

The above definition can be extended to more than two random variables, i.e., n-dimensional random variable/vector.

Discrete and Continuous Random Variables

1. (X, Y) is a discrete random variable if the possible values of $(X(s), Y(s))$ are **finite** or **countable infinite**.
2. (X, Y) is a continuous random variable if the possible values of $(X(s), Y(s))$ can **assume all values in some region** of the Euclidean plane $\mathbb{R}^2$.

Joint Probability Function

Discrete Random Variable

With each possible value (x_i, y_j), we associate a number $f_{X,Y}(x_i, y_j)$ representing $\Pr(X = x_i, Y = y_j)$ and satisfying the following conditions:

1. $$0 \leqslant f_{X,Y}(x_i, y_j) \leqslant 1 \text{ for all } (x_i, y_j) \in R_{X,Y}$$
2. $$\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} f_{X,Y}(x_i, y_j) = 1$$

or

$$\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \Pr(X = x_i, Y = y_j) = 1$$

$f_{X,Y}$ is the **joint probability function** of (X, Y).

Continuous Random Variable

$f_{X,Y}(x, y)$ is called a **joint probability density function** if it satisfies the following:

1.
$$f_{X,Y}(x,y) \geqslant 0 \text{ for all } (x,y) \in R_{X,Y}$$
2.
$$\iint_{(x,y)\in R_{X,Y}} f_{X,Y}(x,y)\ dx\ dy = 1$$
or
$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\ dx\ dy = 1$$

Marginal Probability Distribution

Discrete Random Variable

$$f_X(x) = \sum_y f_{X,Y}(x,y)$$
$$f_Y(y) = \sum_x f_{X,Y}(x,y)$$

Conditional Random Variable

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\ dy$$
$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\ dx$$

Basically we **fix one of the two values**, then either sum or integrate over the other. It gives the probabilities of various values of the variables in the subset without reference to the values of the other variables.

Conditional Probability Distribution

The conditional distribution of Y given that $X = x$ is given by
$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)},\ \text{if } f_X(x) > 0$$
The conditional distribution of Y given that $X = x$ is given by
$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)},\ \text{if } f_Y(y) > 0$$
The conditional p.f.'s satisfy all the requirements for a 1-dimensional p.f.. Thus,

1. For a fixed y,
$$f_{X|Y}(x|y) \geqslant 0$$
For a fixed x,
$$f_{Y|X}(y|x) \geqslant 0$$
2. For discrete r.v.'s,
$$\sum_x f_{X|Y}(x|y) = 1 \text{ and } \sum_y f_{Y|X}(y|x) = 1$$
For continuous r.v.'s,
$$\int_{-\infty}^{\infty} f_{X|Y}(x|y) = 1 \text{ and } \int_{-\infty}^{\infty} f_{Y|X}(y|x) = 1$$
3. For $f_X(x) > 0$,
$$f_{X,Y}(x,y) = f_{Y|X}(y|x)f_X(x)$$
For $f_Y(y) > 0$,
$$f_{X,Y}(x,y) = f_{X|Y}(x|y)f_Y(y)$$

Independent Random Variables

Random variables X and Y are independent if and only if $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ for all x, y. This can be extended to n random variables.

Product Space
The product of 2 positive functions $f_X(x)$ and $f_Y(y)$ results in a function which is positive on a **product space**.
If $f_X(x) > 0$ for $x \in A_1$ and $f_Y(y) > 0$ for $y \in A_2$, then $f_X(x)f_Y(y) > 0$ for $(x,y) \in A_1 \times A_2$.

Expectation

Discrete Random Variable

1. Marginal
$$E(X) = \sum_x x f_X(x)$$
$$E(Y) = \sum_y y f_Y(y)$$
2. Conditional
$$E(X|Y) = \sum_x x f_{X|Y}(X|Y)$$
$$E(Y|X) = \sum_y y f_{Y|X}(Y|X)$$
3. Function
$$E[g(X,Y)] = \sum_x\sum_y g(x,y) f_{X,Y}(x,y)$$

Continuous Random Variable

1. Marginal
$$E(X) = \int_{-\infty}^{\infty} x f_X(x)\ dx$$
$$E(Y) = \int_{-\infty}^{\infty} y f_Y(y)\ dy$$
2. Conditional
$$E(X|Y) = \int_{-\infty}^{\infty} x f_{X|Y}(X|Y)\ dx$$
$$E(Y|X) = \int_{-\infty}^{\infty} y f_{Y|X}(Y|X)\ dy$$
3. Function
$$E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)\ dx\ dy$$

Covariance

Let $g(X,Y) = (X-\mu_X)(Y-\mu_Y)$. This leads to the definition of covariance between two random variables.
$$\sigma_{X,Y} = Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$$

Discrete
$$\sum_X\sum_Y (X-\mu_X)(Y-\mu_Y) f_{X,Y}(x,y)$$

Continuous
$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (X-\mu_X)(Y-\mu_Y) f_{X,Y}(x,y)\ dx\ dy$$

Properties

1.
$$Cov(X,Y) = E(XY) - \mu_X\mu_Y$$
2.
$$X \text{ and } Y \text{ are independent } \Rightarrow Cov(X,Y) = 0$$
$$Cov(X,Y) = 0 \not\Rightarrow X \text{ and } Y \text{ are independent}$$
3.
$$Cov(aX+b, cY+d) = ac\ Cov(X,Y)$$
4.
$$V(aX+bY) = a^2V(X) + b^2V(Y) + 2ab\ Cov(X,Y)$$

Correlation Coefficient

The correlation coefficient of X and Y, denoted by $Cor(X,Y)$ or $\rho_{X,Y}$ or ρ, is defined by
$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sqrt{V(X)}\sqrt{V(Y)}}$$

Properties
1. $-1 \leqslant \rho_{X,Y} \leqslant 1$
2. $\rho_{X,Y}$ is a measure of the degree of **linear** relationship between X and Y
3.
$$X \text{ and } Y \text{ are independent } \Rightarrow \rho_{X,Y} = 0$$
$$\rho_{X,Y} = 0 \not\Rightarrow X \text{ and } Y \text{ are independent}$$

Discrete Uniform Distribution

If the random variable X assumes the values $x_1, x_2, \ldots, x_k$ with **equal probability**, then the random variable X is said to have a discrete uniform distribution and the probability function is given by
$$f_X(x) = \frac{1}{k},\ \text{for } x = x_1, x_2, \ldots, x_k$$

Mean
$$\mu = E(X) = \sum_{i=1}^{k} x_i \frac{1}{k} = \frac{1}{k}\sum_{i=1}^{k} x_i$$

Variance
$$\sigma^2 = V(X) = \sum_{i=1}^{k}(x_i-\mu)^2\frac{1}{k} = \frac{1}{k}\sum_{i=1}^{k}(x_i-\mu)^2$$
or
$$\sigma^2 = E(X^2) - \mu^2 = \frac{1}{k}\sum_{i=1}^{k} x_i^2 - \mu^2$$

Bernoulli Distribution

A Bernoulli experiment is a random experiment with only **two possible outcomes**, and we can code them as 1 and 0. A random variable X is defined to have a Bernoulli distribution if the probability function of X is given by
$$f_X(x) = p^x(1-p)^{1-x}$$
for $x = 0, 1$ and $0 < p < 1$
<u>Note:</u> $(1-p)$ is often denoted by q.

Mean
$$\mu = E(X) = p$$

Variance
$$\sigma^2 = V(X) = p\ (1-p) = pq$$

Binomial Distribution

A random variable X is defined to have a binomial distribution with two parameters n and p (i.e., $X \sim B(n,p)$), if the probability function of X is given by
$$f_X(x) = \binom{n}{x}\ p^x(1-p)^{n-x}$$
$$= \binom{n}{x}\ p^x q^{n-x}$$
<u>Note:</u> X is the number of successes that occur in n independent Bernoulli events.

Mean
$$\mu = E(X) = np$$

Variance
$$\sigma^2 = V(X) = np\ (1-p) = npq$$

Negative Binomial Distribution

A random variable X is defined to have a negative binomial distribution with two parameters k and p (i.e., $X \sim NB(k,p)$), if the probability function of X is given by
$$f_X(x) = \binom{x-1}{k-1}\ p^k q^{x-k}$$
for $x = k, k+1, k+2, \ldots$
<u>Note:</u> Basically, x-th trial and k-th sucess.

Mean
$$\mu = E(X) = \frac{k}{p}$$

Variance
$$\sigma^2 = V(X) = \frac{(1-p)\ k}{p^2}$$

Geometric Distribution

A special case of the negative binomial distribution, where $k = 1$ (i.e., $X \sim Geom(p) \Leftrightarrow X \sim NB(1,p)$).

Poisson Distribution

Poisson experiments yield the number of successes occurring during a given time interval or in a specified region.
The probability distribution of the Poisson random variable X, is called the Poisson distribution and the probability function is given by
$$f_X(x) = \frac{e^{-\lambda}\lambda^x}{x!}\ \text{for } x = 0, 1, 2, \ldots$$

where λ is the average number of successes occurring in the given time interval or specified region and $e \approx 2.1718281818\ldots$

Properties
1. The **number of successes** occurring in one time interval or specified region are **independent** of those occurring in any other disjoint time interval or region of space.
2. The **probability of a single success** occurring during a very short time interval or in a small region is **proportional to** the length of the time interval or size of the region and does not depend on the number of successes occurring outside this time interval or region.
3. The **probability of more than one success** occurring in such a short time interval or falling in such a small region is **negligible**.

Mean

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

Variance

$$\sigma^2 = V(X) = \lambda$$

Poisson Approximation to BD

Let X be a binomial random variable with parameters n and p. That is

$$\Pr(X = x) = f_X(x) = \binom{n}{x} p^x q^{n-x}$$

Suppose that $n \to \infty$ and $p \to 0$ such that $\lambda = np$ remains a constant as $n \to \infty$.
Then X will have approximately a Poisson distribution with parameter np. That is

$$\lim_{\substack{p\to 0\\ n\to\infty}} \Pr(X = x) = \frac{e^{-np}(np)^x}{x!}$$

Note: Poisson distribution is a good approximation to the binomial distribution $B(n,p)$ when n is large but np is small (so that $n(1-p)$ large).

Continuous Uniform Distribution

A random variable has a uniform distribution over the interval $[a,b], -\infty < a < b < \infty$, denoted by $U(a,b)$, if its probability density function is given by

$$f_X(x) = \frac{1}{b-a}, \text{ for } a \leqslant x \leqslant b$$

Mean

$$E(X) = \frac{a+b}{2}$$

Variance

$$V(X) = \frac{1}{12}(b-a)^2$$

CDF

$$F(x) = \begin{cases} 0, & \text{for } x < a \\ \frac{x-a}{b-a}, & \text{for } a \leqslant x \leqslant b \\ 1, & \text{for } 1 \leqslant b < x \end{cases}$$

Exponential Distribution

A continuous random variable X assuming all **nonnegative** values is said to have an exponential distribution with parameter $\alpha > 0$ if its probability density function is given by

$$f_X(x) = \alpha e^{-\alpha x}, \text{ for } x > 0, \ \alpha > 0$$

Mean

$$E(X) = \frac{1}{\alpha}$$

Variance

$$V(X) = \frac{1}{\alpha^2}$$

Note: The p.d.f. can also be written in the following form:

$$f_X(x) = \frac{1}{\mu} e^{-x/\mu}$$

No Memory Property

$$\Pr(X > s + t \mid X > s) = \Pr(X > t)$$

CDF

$$F_X(x) = \Pr(X \leqslant x) = 1 - e^{-\alpha x}$$

Hence,

$$\Pr(X > x) = e^{-\alpha x}$$

Normal Distribution

The random variable X assuming all real values, $-\infty < x < \infty$, has a normal (or Gaussian) distribution if its probability density function is given by

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \text{ for } -\infty < x < \infty$$

where $-\infty < \mu < \infty$ and $\sigma > 0$.
Note: It is denoted by $N(\mu, \sigma^2)$.

Properties
1. The graph is bell-shaped
2. Symmetrical about the vertical line $x = \mu$
3. Maximum point is at $x = \mu$ and its value is $\frac{1}{\sqrt{2\pi}\sigma}$
4. Two normal curves are identical in shape if they have the same σ^2, but centered at different positions if their μ are different
5. As σ increases, the curve flattens, and vice versa

Standardized Normal distribution
If X has distribution $N(\mu, \sigma^2)$, and if $Z = \frac{(X-\mu)}{\sigma}$, then Z has the $N(0,1)$ distribution (standardized normal distribution), and $E(Z) = 0$ and $V(Z) = 1$.
The p.d.f. of Z may be written as

$$f_Z(z) = \frac{1}{2\pi}\exp\left(-\frac{z^2}{2}\right)$$

The importance of the standardized normal distribution is the fact that it is tabulated.
Whenever X has distribution $V(\mu, \sigma^2)$, we can always simplify the process of evaluating the values of $Pr(x_1 < X < x_2)$ by using the transformation $Z = (X-\mu)/\sigma$. Hence,

$$x_1 < X < x_2 \Leftrightarrow (x_1 - \mu)/\sigma < Z < (x_2 - \mu)/\sigma$$

Let $z_1 = (x_1 - \mu)/\sigma$ and $z_2 = (x_2 - \mu)/\sigma$. Then,

$$\Pr(x_1 < X < x_2) = \Pr(z_1 < Z < z_2)$$

Statistical Tables
Statistical tables give values $\Phi(z)$ for a given z, where $\Phi(z)$ is the cumulative distribution function of a standardized normal random variable Z. Thus,
1. $\Phi(z) = \Pr(Z \leqslant z)$ 2. $1 - \Phi(z) = \Pr(Z > z)$
Some statistical tables give the 100α percentage points, z_α, of a standardized normal distribution, where

$$\alpha = \Pr(Z \geqslant z_\alpha) = \int_{z_\alpha}^{\infty} \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{z^2}{2}\right)\, dz$$

Since the p.d.f. of Z is symmetrical about 0, therefore

$$\Pr(Z \geqslant z_\alpha) = \Pr(Z \leqslant -z_\alpha) = \alpha$$

Normal Approximation to BD

When $n \to \infty$ and $p \to 1/2$, we can use normal distribution to approximate binomial distribution.
If X is a binomial random variable with mean $\mu = np$ and variance $\sigma^2 = np(1-p)$, then as $n \to \infty$,

$$Z = \frac{X - np}{\sqrt{npq}} \text{ is approximately } \sim N(0,1)$$

Note: A good rule of thumb is to use the normal approximation only when $np > 5$ and $n(1-p) > 5$.

Continuity Correction
1.
$$\Pr(X = k) \approx \Pr(k - 1/2 < X < k + 1/2)$$
2.
$$\Pr(a \leqslant X \leqslant b) \approx \Pr(a - 1/2 < X < b + 1/2)$$
3.
$$\Pr(a < X \leqslant b) \approx \Pr(a + 1/2 < X < b + 1/2)$$
4.
$$\Pr(a \leqslant X < b) \approx \Pr(a - 1/2 < X < b - 1/2)$$
5.
$$\Pr(a < X < b) \approx \Pr(a + 1/2 < X < b - 1/2)$$
6.
$$\begin{aligned}\Pr(X \leqslant c) &= \Pr(0 \leqslant X \leqslant c) \\ &\approx \Pr(-1/2 < X < c + 1/2)\end{aligned}$$
7.
$$\begin{aligned}\Pr(X > c) &= \Pr(c < X \leqslant n) \\ &\approx \Pr(c + 1/2 < X < n + 1/2)\end{aligned}$$

Population

The totality of all possible outcomes or observations of a survey or experiment. There are two kinds of populations, namely, finite and infinite populations.

Finite Population
A finite population consists of a finite number of elements.

Infinite Population
An infinite population is one that consists of an infinitely (countable or uncountable) large number of elements.
Note: some finite populations are so large that in theory we assume them to be infinite.

Sample

A sample is any subset of a population.

Simple Random Sample
A simple random sample of n members is a sample that is chosen in such a way every subset of n observations of the population has the same probability of being selected.
Let X be a random variable with certain probability distribution, $f_X(x)$. Let $X_1, X_2, \ldots, X_n$ be n independent random variables each having the same distribution as X, then $(X_1, X_2, \ldots, X_n)$ is called a random sample of size n from as population with distribution $f_X(x)$.
The joint p.f. (or p.d.f.) of $(X_1, X_2, \ldots, X_n)$ is given by

$$\begin{aligned}&f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) \\ &= f_{X_1}(x_1) f_{X_2}(x_2) \cdots f_{X_n}(x_n)\end{aligned}$$

where $f_X(x)$ is the p.f. (or p.d.f.) of the population.

Sampling from a Finite Population

Sampling without replacement
In general, there are ${}_NC_n$ samples of size n that can be drawn from a finite population of size N without replacement.
Each sample has equal chance of being selected, with a probability of $\frac{1}{{}_NC_n}$.

Sampling with replacement
In general, there are N^n samples of size n that can be drawn from a finite population of size N with replacement.
Each sample has equal chance of being selected, with a probability of $\frac{1}{N^n}$.

Sampling from an Infinite Population

Sampling with/without replacement
The concept of a random sample from an infinite population is more difficult to explain.
Refer to *Chapter 5, page 16-20* for some unclear examples.

Sampling Distribution of Sample Mean

Statistic
A function of a random sample $(X_1, X_2, \ldots, X_n)$ is called a statistic. For example, the mean is a statistic. Hence, a statistic is a random variable, and it is meaningful to consider the probability distribution of a statistic, which is also called a sampling distribution.
Note: Statistic must **not** depend on unknown parameters.

Sample Mean

For some random sample of size n represented by $X_1, X_2, \ldots, X_n$, the sample mean is defined by the statistic

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

If the values in the random sample are observed and they are $x_1, x_2, \ldots, x_n$, then the realization of the statistic $\overline{X}$ is given by

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Sampling Distribution
For random samples of size n taken from an **infinite population** or a **finite population with replacement** having population mean μ and population standard deviation σ, the sampling distribution of the sample mean $\overline{X}$ has its mean and variance given by:

$$\mu_{\overline{X}} = \mu_X \text{ ,i.e., } E(\overline{X}) = E(X)$$

$$\sigma^2_{\overline{X}} = \frac{\sigma^2_X}{n} \text{ ,i.e., } V(\overline{X}) = \frac{V(X)}{n}$$

Law of Large Number
Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a population having any distribution with mean μ and **finite** population variance σ^2.
For any $\epsilon \in \mathbb{R}$,

$$\mathrm{P}(|\overline{X} - \mu| > \epsilon) \to 0 \text{ as } n \to \infty$$

In other words, as the sample size increases, the probability that the sample mean differs from the population mean goes to zero.

Central Limit Theorem

The sampling distribution of the sample mean $\overline{X}$ is approximately normal with mean μ and variance $\frac{\sigma^2}{n}$ if n is sufficiently large.
Hence $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ follows approximately $N(0,1)$.
Sampling distribution properties of $\overline{X}$:
1. Central Tendency

$$\mu_{\overline{X}} = \mu$$

2. Variation

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

Normal Sampling Distribution
1. If $X_i, i = 1, 2, \ldots, n$ are $N(\mu, \sigma^2)$, then $\overline{X}$ is $N(\mu, \frac{\sigma^2}{n})$ regardless of the sample size n.
2. If $X_i, i = 1, 2, \ldots, n$ are approximately $N(\mu, \sigma^2)$, then $\overline{X}$ is approximately $N(\mu, \frac{\sigma^2}{n})$ regardless of the sample size n.

Sampling Distribution of Difference of Two Sample Means
If independent samples of sizes $n_1 (\geqslant 30)$ and $n_2 (\geqslant 30)$ are drawn from two populations, with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2 respectively, then the sampling distribution of the differences of the sample means $\overline{X}_1$ and $\overline{X}_2$ is approximately normally distributed with mean and standard deviation given by:
1.

$$\mu_{\overline{X}_1 - \overline{X}_2} = \mu_1 - \mu_2$$

2.

$$\sigma_{\overline{X}_1 - \overline{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

3.

$$\frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \text{ approx } \sim N(0,1)$$

Chi-square distribution

If Y is a random variable with probability density function

$$f_Y(y) = \frac{1}{2^{n/2}\Gamma(n/2)} y^{n/2-1} e^{-y/2}$$

for $y > 0$, and 0 otherwise, then Y is defined to have a chi-square distribution with n degrees of freedom, denoted by $\chi^2(n)$, where n is a positive integer, and Γ is the gamma function.
The gamma function, Γ, is defined by

$$\Gamma(n) = \int_0^\infty x^{n-1} e^{-x}\, dx = (n-1)!$$

for $n = 1, 2, 3, \ldots$

Properties
1. If $Y \sim \chi^2(n)$, then $E(Y) = n$ and $V(Y) = 2n$
2. For large n, $\chi^2(n)$ approx $\sim N(n, 2n)$
3. If $Y_1, Y_2, \ldots, Y_k$ are independent chi-square random variables with $n_1, n_2, \ldots, n_k$ degrees of freedom respectively, then $Y_1 + Y_2 + \cdots + Y_k$ has a chi-square distribution with $n_1 + n_2 + \cdots + n_k$ degrees of freedom. That is,

$$\sum_{i=1}^{k} Y_i \sim \chi^2(\sum_{i=1}^{k} n_i)$$

From Normal to Chi-square
1. If $X \sim N(0,1)$, then $\chi^2 \sim \chi^2(1)$
2. For large $X \sim N(\mu, \sigma^2)$, then $[(X-\mu)/\sigma]^2 \sim \chi^2(1)$
3. Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal population with mean μ, and variance σ^2. Define

$$Y = \sum_{i=1}^{n} \frac{(X_i - \mu)^2}{\sigma^2}$$

Then $Y \sim \chi^2(n)$

Sampling Distribution of $(n-1)S^2/\sigma^2$

The statistic

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

is the sample variance. However, it has little practical application. Instead, we shall consider the sampling distribution of the random variable

$$\frac{(n-1)S^2}{\sigma^2}$$

when $X_i \sim N(\mu, \sigma^2)$ for all i.
If S^2 is the variance of a random sample of size n taken from a normal population having the variance σ^2, then the random variable

$$\frac{(n-1)S^2}{\sigma^2}$$

has a chi-square distribution with $n-1$ degrees of freedom. That is,

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

t-distribution

Suppose $Z \sim N(0,1)$ and $U \sim \chi^2(n)$. If Z and U are independent, and let

$$T = \frac{Z}{\sqrt{U/n}}$$

then the random variable T follows the t-distribution with n degrees of freedom. That is,

$$\frac{Z}{\sqrt{U/n}} \sim t(n)$$

If T follows a t-distribution with n degrees of freedom, then its p.d.f. is given by

$$f_T(t) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\,\Gamma(\frac{n}{2})}(1 + \frac{t^2}{n})^{\frac{n+1}{2}}, \ -\infty < t < \infty$$

The gamma function, Γ, is defined by

$$\Gamma(n) = \int_0^\infty x^{n-1} e^{-x}\, dx = (n-1)! \text{ for } n = 1, 2, 3, \ldots$$

Properties
1. The graph of the t-distribution is symmetric about the vertical axis and resembles the graph of the standard normal distribution
2.

$$\lim_{n\to\infty} f_T(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$$

3. If $T \sim t(n)$, then $E(T) = 0$ and $V(T) = n/(n-2)$ for $n > 2$
4.

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

F-distribution

Let U and V be independent random variables having $\chi^2(n_1)$ and $\chi^2(n_2)$, respectively, then the distribution of random variable,

$$F = \frac{U/n_1}{V/n_2}$$

is called a F distribution with (n_1, n_2) degrees of freedom.
The p.d.f. F is given by

$$f_F(x) = \frac{n_1^{n_1/2} n_2^{n_2/2}\Gamma(\frac{n_1+n_2}{2})}{\Gamma(\frac{n_1}{2})\Gamma(\frac{n_2}{2})} \frac{x^{\frac{n_1}{2}-1}}{(n_1 x + n_2)^{(n_1+n_2)/2}}$$

for $x > 0$ or 0 otherwise.
It can be shown that

$$E(X) = \frac{n_2}{(n_2 - 2)}, \text{ with } n_2 > 2$$

$$V(X) = \frac{2n_2^2(n_1 + n_2 - 2)}{n_1(n_2-2)^2(n_2-4)}, \text{ with } n_2 > 4$$

Properties
1.

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n_1 - 1, n_2 - 1)$$

2. If $F \sim F(n, m)$ then $1/F \sim F(m, n)$

Remarks

For Chapter 6 and Chapter 7, refer to *Summary of Confidence Interval and Hypothesis Testing Version 3.pdf*

For Excel commands, refer to *Excel commands to compute prob for special distributions Version 2.xlsx*

For R commands, refer to *R commands to find prob quantiles Version 2.pdf*

For Python code, refer to *Python code to find prob quantiles Version 2.pdf*

For statistical tables, refer to *Statistical Tables.pdf*