ST2131 Probability Notes

Contents

Chapter 1

Combinatorial Analysis

1.1 Multiplication Rule

1.1.1 Basic Principle of Counting

Suppose that two experiments are to be performed. If

- experiment 1 can result in any one of m possible outcomes; and
- experiment 2 can result in any one of n possible outcomes;

then together there are mn possible outcomes of the two experiments.

1.1.2 Generalised Basic Principle of Counting

Suppose that r experiments are to be performed. If

- experiment 1 results in n_1 possible outcomes;
- experiment 2 results in n_2 possible outcomes;
- $\dots$
- experiment r results in n_r possible outcomes;

then together there are $n_1 n_2 \dots n_r$ possible outcomes of the r experiments.

1.2 Addition Rule

Suppose that two experiments, experiment 1 and experiment 2, are mutually exclusive, i.e. either experiment 1 or experiment 2 occurs, but not both. If

- experiment 1 can result in any one of m possible outcomes; and
- experiment 2 can result in any one of n possible outcomes;

then together there are $m + n$ possible outcomes.

1.3 Factorial

Definition. $n! = n(n-1)(n-2)\dots 3(2)(1)$ with the convention that $0! = 1$.

1.4 Permutations

Number of Objects	**Arrangment**	**Distinction Between Objects**	**Number of Permutations**
n	Linear	All Distinct	$n!$
	Linear	n_1 are alike, n_2 are alike, …, n_r are alike	$\dfrac{n!}{n_1!\,n_2!\,\dots n_r!}$
	Circular	All Distinct	$(n-1)! = \dfrac{n!}{n}$
	Necklace	All Distinct	$\dfrac{(n-1)!}{2}$

1.5 Binomial Coefficient

Definition. The **binominal coefficient** is

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

If n is a nonnegative integer, and $r < 0$ or $r > n$, define

$$\binom{n}{r} = 0$$

1.6 Combinations

If there are n distinct objects, of which we choose a group of r items, then the number of possible groups is given by $\binom{n}{r}$.

1.7 Properties of Binomial Coefficient

1.7.1 Symmetric Property

For $r = 0, 1, 2, \dots, n$,

$$\binom{n}{r} = \binom{n}{n-r}$$

1.7.2 Only one way to choose nothing or everything

$$\binom{n}{0} = \binom{n}{n} = 1$$

1.7.3 Recurrence Formula

For $1 \leq r \leq n$,

$$\binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r}$$

1.8 Binomial Theorem

Theorem. Let n be a nonnegative integer, then

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

1.8.1 Special Case: $x = y = 1$

$$2^n = \sum_{k=0}^{n} \binom{n}{k}$$

1.8.2 Special Case: $x = -1, y = 1$

$$0 = \sum_{k=0}^{n} \binom{n}{k} (-1)^k$$

1.8.3 Sum of Alternate Binomial Coefficients are Equal

It follows from section 1.8.2 that

$$\binom{n}{0} + \binom{n}{2} + \binom{n}{4} + \ldots = \binom{n}{1} + \binom{n}{3} + \binom{n}{5} + \ldots$$

1.9 Multinomial Coefficient

Definition. The **multinominal coefficient** is

$$\binom{n}{n_1, n_2, \ldots, n_r} = \frac{n!}{n_1!\, n_2!\, \ldots n_r!}$$

1.10 Divide n Objects into r Groups

The number of ways to divide n objects into r distinct groups of size $n_1, n_2, \ldots, n_r$ such that $\sum_{i=1}^{r} n_i = n$ is given by

$$\binom{n}{n_1}\binom{n-n_1}{n_2}\binom{n-n_1-n_2}{n_3}\cdots\binom{n-n_1-n_2-\ldots-n_{r-1}}{n_r} = \binom{n}{n_1, n_2, \ldots, n_r}$$

1.11 Multinomial Theorem

Let n be a nonnegative integer, then

$$(x_1 + x_2 + \ldots + x_r)^n = \sum_{n_1+n_2+\ldots+n_r=n} \binom{n}{n_1, n_2, \ldots, n_r} x_1^{n_1} x_2^{n_2} \ldots x_r^{n_r}$$

1.12 The Number of Positive Integer Solutions of Equations

Given the following equation in $x_1, x_2, \ldots, x_r$,

$$x_1 + x_2 + \ldots + x_r = n$$

where $x_1, x_2, \ldots, x_r$, and n are positive integers. The number of solutions to this equation is

$$\binom{n-1}{r-1}$$

1.13 The Number of Nonnegative Integer Solutions of Equations

Given the following equation in $x_1, x_2, \ldots, x_r$,

$$x_1 + x_2 + \ldots + x_r = n$$

where $x_1, x_2, \ldots, x_r$, and n are nonnegative integers. The number of solutions to this equation is

$$\binom{n+r-1}{r-1}$$

Chapter 2

Axioms of Probability

2.1 Sample Space

Definition. The **sample space** is the set of all possible outcomes of an experiment, usually denoted by S.

2.1.1 Remarks

The **sample space** depends on the outcomes of interest.

2.2 Event

Definition. Any subset A of the **sample space** is an **event**.

2.3 Basic Set Identities

For any sets E, F, G, E_i for $i \in \mathbb{N}, 1 \leq i \leq n$, the following identities are true:

- Commutative Laws
  - $EF = FE$
  - $E \cup F = F \cup E$
- Associative Laws
  - $(EF)G = E(FG)$
  - $(E \cup F) \cup G = E \cup (F \cup G)$
- Distributive Laws
  - $(E \cup F)G = (EG) \cup (FG)$
  - $(EF) \cup G = (E \cup G)(F \cup G)$

- De Morgan's Laws
  - $\left(\bigcup_{i=1}^{n} E_i\right)^c = \bigcap_{i=1}^{n} E_i^c$
  - $\left(\bigcap_{i=1}^{n} E_i\right)^c = \bigcup_{i=1}^{n} E_i^c$

2.4 Subset Notation

Definition. For this course, for any two sets A and B, $A \subset B$ if every element of A is also an element of B.

2.5 Definitions of Probability

2.5.1 Classical Approach

Assume all the sample points are **equally likely** to occur. Then

$$P(E) = \frac{|E|}{|S|}$$

where $|E|$ is the number of sample points in event E and $|S|$ is the number of sample points in S.

2.5.2 Relative Frequency Approach

$$P(E) = \lim_{n\to\infty} \frac{n(E)}{n}$$

where $n(E)$ is the number of times in n repetitions of the experiment that E occurs.

2.5.3 Subjective Approach

Probability is considered as a measure of belief.

2.6 Mutually Exclusive Events

A sequence of events $E_1, E_2, \ldots$ is **mutually exclusive** if $E_i \cap E_j = \varnothing$ when $i \neq j$.

2.7 Axioms of Probability

Consider an **experiment** whose **sample space** is S. For each event E of the **sample space**, we assume that a number $P(E)$ is defined and satisfies the following three axioms:

1. For any event E,

$$0 \leq P(E) \leq 1$$

2. $P(S) = 1$

3. For any sequence of **mutually exclusive** events $E_1, E_2, \ldots$,

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$$

2.7.1 Remarks

- We call $P(E)$, the probability of the event E.
- Axiom 3 states that, for any sequence of mutually exclusive events, the probability of at least one of these events occurring is just the sum of their respective probabilities.

2.8 Probability of the Empty Set is Zero

Proposition. $P(\varnothing) = 0$.

2.9 Probability is Finitely Additive

Proposition. For any finite sequence of **mutually exclusive** events $E_1, E_2, \ldots, E_n$,

$$P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i)$$

2.10 The Complement Rule

Proposition. Let E be an event, then

$$P(E^c) = 1 - P(E)$$

2.11 Probability is Monotonic

Proposition. If A and B are events such that $A \subset B$, then

$$P(A) \leq P(B)$$

2.12 The Sum Rule

Proposition. Let A and B be any two events, then

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

2.13 Inclusion/Exclusion Principle

Proposition. Let $E_1, E_2, \ldots, E_n$ be any events, then

$$\begin{aligned} P(E_1 \cup E_2 \cup \ldots \cup E_n) = &\sum_{i=1}^{n} P(E_i) - \sum_{1 \leq i_1 < i_2 \leq n} P(E_{i_1} \cap E_{i_2}) + \ldots \\ &(-1)^{r+1} \sum_{1 \leq i_1 < \ldots < i_r \leq n} P(E_{i_1} \cap \ldots \cap E_{i_r}) + \ldots \\ &+ (-1)^{n+1} P(E_1 \cap \ldots \cap E_n) \end{aligned}$$

2.14 Sample Spaces Having Equally Likely Outcomes

Given a **sample space** $S = \{s_1, s_2, \ldots, s_N\}$, where $N = |S|$ denotes the number of outcomes of S and the outcomes are assumed to be equally likely to occur. Then for $1 \leq i \leq n$,

$$P(\{s_i\}) = \frac{1}{|S|}$$

Furthermore, if event A has $|A|$ outcomes, then

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

2.15 Probability as a Continuous Set Function

2.15.1 Increasing Sequence of Events

Definition. A sequence of events $\{E_n\}, n \geq 1$ is an **increasing** sequence if

$$E_1 \subset E_2 \subset \ldots \subset E_n \subset E_{n+1} \subset \ldots$$

2.15.2 Decreasing Sequence of Events

Definition. A sequence of events $\{E_n\}, n \geq 1$ is a **decreasing** sequence if

$$E_1 \supset E_2 \supset \ldots \supset E_n \supset E_{n+1} \supset \ldots$$

2.15.3 Limit of Increasing Sequence of Events

Definition. If $\{E_n\}, n \geq 1$ is an **increasing** sequence of events, then we define a new event, denoted by $\lim_{n \to \infty} E_n$ as

$$\lim_{n \to \infty} E_n = \bigcup_{i=1}^{\infty} E_i$$

2.15.4 Limit of Decreasing Sequence of Events

Definition. If $\{E_n\}, n \geq 1$ is an **decreasing** sequence of events, then we define a new event, denoted by $\lim_{n\to\infty} E_n$ as

$$\lim_{n\to\infty} E_n = \bigcap_{i=1}^{\infty} E_i$$

2.15.5 Probability of the Limit of a Monotonic Sequence of Events

Proposition. If $\{E_n\}, n \geq 1$ is either an **increasing** or **decreasing** sequence of events, then

$$P\left(\lim_{n\to\infty} E_n\right) = P\left(\bigcap_{i=1}^{\infty} E_i\right) = \lim_{n\to\infty} P(E_n)$$

Chapter 3

Conditional Probability and Independence

3.1 Conditional Probability

3.1.1 Definition of Conditional Probability

Definition. Let E and F be two events. Suppose that $P(F) > 0$, the conditional probability of E given F, denoted by $P(E|F)$, is defined as

$$P(E|F) = \frac{P(EF)}{P(F)}$$

3.1.1.1 Remark

$P(E|F)$ can also be read as: the conditional probability that E occurs given that F has occurred.

3.1.2 Multiplication Rule

Suppose that $P(A) > 0$, then

$$P(AB) = P(A)P(B|A)$$

3.1.3 General Multiplication Rule

Let $A_1, A_2, \ldots, A_n$ be n events, then

$$P(A_1A_2 \ldots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2) \ldots P(A_n|A_1A_2 \ldots A_{n-1})$$

3.2 Bayes' Formulas

3.2.1 Partition of a Sample Space

$A_1, A_2, \ldots, A_n$ **partition** the **sample space** S if

- They are "mutually exclusive", i.e. $A_i \cap A_j = \varnothing$, for all $i \neq j$; and
- They are "collectively exhaustive", i.e. $\bigcup_{i=1}^{n} A_i = S$.

3.2.2 Bayes' First Formula / Law of Total Probability

Proposition. Suppose the events $A_1, A_2, \ldots, A_n$ **partition** the **sample space**. Assume further that $P(A_i) > 0$ for $1 \leq i \leq n$. Let B be any event, then

$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \ldots + P(B|A_n)P(A_n)$$

3.2.2.1 Special Case: Partition into A and A^c

Let A and B be any two events. Then the events A, A^c partition the **sample space**. Hence,

$$P(B) = P(B|A)P(A) + P(B|A^c)P(A^c)$$

3.2.2.2 Remark

- This formula is useful in computing the probability of a composite event, i.e. an event B which depends on a series of causes $A_1, A_2, \ldots, A_n$.
- The summation can be interpreted as a weighted average of n cases: case 1 being A_1 occurs, case 2 being A_2 occurs, ..., case n being A_n occurs.
- The weights are placed according to how likely $A_1, A_2, \ldots, A_n$ occur.

3.2.3 Bayes' Second Formula

Suppose the events $A_1, A_2, \ldots, A_n$ **partition** the **sample space**. Assume further that $P(A_i) > 0$ for $1 \leq i \leq n$. Let B be any event. Then for $1 \leq i \leq n$,

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \ldots + P(B|A_n)P(A_n)}$$

3.2.3.1 Remarks

- This formula has been interpreted as a formula for "inverse probabilities".
- If $A_1, A_2, \ldots, A_n$ is a series of causes and B is a possible effect, then
  - $P(B|A_i)$ is the probability of B when it is known that A_i is the cause, whereas
  - $P(A_i|B)$ is the probability that A_i is the cause when it is known that B is the effect.

3.2.4 Odds of an Event

Definition. The **odds** of an event A is defined by

$$\frac{P(A)}{P(A^c)} = \frac{P(A)}{1 - P(A)}$$

3.3 Independent Events

3.3.1 Definition of Two Independent / Dependent Events

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

$$P(AB) = P(A)P(B)$$

and they are said to be **dependent** if

$$P(AB) \neq P(A)P(B)$$

3.3.1.1 Remarks

The following phrases mean the same:

- A and B are independent;
- A is independent of B;
- B is independent of A.

3.3.2 Independent Events and their Conditional Probabilities

If two events A and B are **independent**. Suppose $P(B) > 0$, then

$$\begin{aligned} P(A|B) &= \frac{P(AB)}{P(B)} \\ &= \frac{P(A)P(B)}{P(B)} \\ &= P(A) \end{aligned}$$

Similarly, if $P(A) > 0$,

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

3.3.3 Independent Events and their Complement Events

Proposition. If A and B are **independent**, then so are

- A and B^c;
- A^c and B;
- A^c and B^c.

3.3.4 Event is Not Always Independent of the Product of Events it is Independent to

If A is independent of B, and A is also independent of C, it is NOT necessarily true that A is independent of BC.

3.3.5 Definition of Three Independent Events

Definition. Three events A, B, and C are said to be independent if all of the following 4 conditions hold:

1. $P(ABC) = P(A)P(B)P(C)$
2. $P(AB) = P(A)P(B)$
3. $P(AC) = P(A)P(C)$
4. $P(BC) = P(B)P(C)$

3.3.5.1 Remarks

The second to fourth conditions mean that A, B, and C are pairwise independent, i.e.:

- A and B are independent (from second condition)
- A and C are independent (from third condition)
- B and C are independent (from fourth condition)

3.3.6 Definition of n Independent Events

Definition. Events $A_1, A_2, \ldots, A_n$ are **independent** if, for every sub-collection of events $A_{i_1}, A_{i_2}, \ldots, A_{i_r}$,

$$P(A_{i_1}A_{i_2}\cdots A_{i_r}) = P(A_{i_1})P(A_{i_2})\cdots P(A_{i_r})$$

3.4 Algebra of Conditional Probability

Proposition. Let A be an event with $P(A) > 0$. Then the following three conditions hold:

1. For any event B,
$$0 \leq P(B|A) \leq 1$$
2. $P(S|A) = 1$
3. Let $B_1, B_2, \ldots$ be a sequence of mutually exclusive events, then
$$P\left(\bigcup_{k=1}^{\infty} B_k \middle| A\right) = \sum_{k=1}^{\infty} P(B_k|A)$$

3.4.1 Remarks

Therefore, $P(\cdot|A)$ as a function of events satisfy the three axioms of probability.

Chapter 4

Random Variables

4.1 Definition of Random Variable

Definition. A **random variable** X is a mapping from the **sample space** to real numbers.

$$X : S \to \mathbb{R}$$

4.1.1 Commonly Used Notation for Random Variable and Their Values

- Use U, V, X, Y, Z upper case letters to denote random variables (for they are functions); and
- Use $u, v, \ldots$ lower case letters to denote values of random variables (for they are real numbers).

4.2 Discrete Random Variables

4.2.1 Definition of Discrete Random Variable

Definition. A random variable X is said to be **discrete** if the range of X is either finite or countably infinite.

4.2.2 Probability Mass Function

Definition. Suppose that a **random variable** X is **discrete**, taking values $x_1, x_2, \ldots$, then the **probability mass function** of X, denoted by p_X (or simply p if the context is clear), is defined as

$$p_X(x) = \begin{cases} P(X = x) & \text{if } x = x_1, x_2, \ldots, \\ 0 & \text{otherwise.} \end{cases}$$

4.2.3 Properties of Probability Mass Function

- $p_X(x_i) \geq 0$ for $i = 1, 2, \ldots$;
- $p_X(x) = 0$ for other values of x;
- Since X must take on one of the values of x_i,

$$\sum_{i=1}^{\infty} p_X(x_i) = 1$$

4.2.4 Cumulative Distribution Function

Definition. The **cumulative distribution function** of X, abbreviated to **distribution function** (d.f.) of X, denoted by F_X or F if the context is clear, is defined as

$$F_X : \mathbb{R} \to \mathbb{R}$$

where

$$F_X(x) = P(X \leq x) \quad \text{for } x \in \mathbb{R}$$

4.2.4.1 Cumulative Distributive Function of a Discrete Random Variable is a Step Function

Suppose that X is discrete and takes values $x_1, x_2, x_3, \ldots$ where $x_1 < x_2 < x_3 < \cdots$. Then F is a step function, where

- F is constant in the interval $[x_{i-1}, x_i)$, taking the value $p(x_1) + p(x_2) + \cdots + p(x_{i-1})$, and then
- take a jump of size $= p(x_i)$.

4.3 Expected Value / Expectation / Mean

Definition. If X is a **discrete random variable** having a **probability mass function** p_X, the **expectation** or the **expected value** of X, denoted by $E(X)$ or μ_X, is defined by

$$E(X) = \sum_x x p_X(x)$$

4.3.1 Remarks

$E(X)$ is also referred to as the **first moment** or the **mean** of X.

4.3.2 Interpretations of Expectation

$E(X)$, the **expectation** of a random variable X, may be interpreted as

- Weighted average of possible values that X can take on, where the weights are the probability that X assumes it;
- Frequency point of views (relative frequency); or
- Center of gravity.

4.4 Tail Sum Formula for Expectation

If **random variable** X is **only nonnegative integer-valued**, i.e. only takes values $0, 1, 2, \ldots$, then

$$E(X) = \sum_{k=1}^{\infty} P(X \geq k) = \sum_{k=0}^{\infty} P(X > k)$$

4.5 Expectation of a Function of a Random Variable

4.5.1 Function of a Random Variable is a Random Variable

Given a **discrete random variable** X and a function $g : \mathbb{R} \to \mathbb{R}$. Then the function $g(X)$ is also a **random variable**.

4.5.2 Expectation of a Function of a Random Variable

Proposition. If X is a **discrete random variable** that takes values x_i, $i \geq 1$, with respective probabilities $p_X(x_i)$, then for any real value function $g : \mathbb{R} \to \mathbb{R}$,

$$E[g(X)] = \sum_{i=1}^{\infty} g(x_i) p_X(x_i)$$

4.5.3 Expectation of a Linear Function of a Random Variable

Corollary. Let a and b be constants, and X be a **random variable**, then

$$E[aX + b] = aE(X) + b$$

4.5.4 Moment of Random Variables

4.5.4.1 k^{th} Moment of Random Variables

Let X be a **random variable**. Then for $k \geq 1, k \in \mathbb{N}$,

$$E(X^k)$$

is called the $\boldsymbol{k^{th}}$ **moment** of X.

4.5.4.2 k^{th} Central Moment of Random Variables

Let X be a **random variable** and $\mu = E(X)$. Then for $k \geq 1, k \in \mathbb{N}$,

$$E[(X-\mu)^k]$$

is called the $\boldsymbol{k^{th}}$ **central moment** of X.

4.5.4.3 Remarks

- The expected value of a random variable X, $E(X)$, is also referred to as the **first moment** or the **mean** of X.
- The first central moment is 0.
- The second central moment, namely, $E[(X-\mu)^2]$ is called the **variance** of X.

4.6 Variance and Standard Deviation

4.6.1 Definition of Variance

Definition. If X is a **random variable** with mean μ, then the **variance** of X, denoted by $\mathrm{Var}(X)$, is defined as

$$\mathrm{Var}(X) = E[(X-\mu)^2]$$

4.6.1.1 Interpretation of Variance

$\mathrm{Var}(X)$ is a measure of scattering (or spread) of the value of X around its **expected value**, μ.

4.6.2 Alternative Formula for Variance

Let X be a **random variable**, then

$$\mathrm{Var}(X) = E(X^2) - [E(X)]^2$$

4.6.3 Properties of Variance

- $\mathrm{Var}(X) \geq 0$.
- $\mathrm{Var}(X) = 0 \iff X$ is a **degenerate** random variable (i.e., the random variable taking only one value, its expected value, μ).
- It follows from the formula that $E(X^2) \geq [E(X)]^2 \geq 0$.

4.6.4 Definition of Standard Deviation

Definition. The **standard deviation** of a **random variable** X, denoted by σ_X or $\mathrm{SD}(X)$, is defined as

$$\sigma_X = \sqrt{\mathrm{Var}(X)}$$

4.6.5 Scaling and Shifting Property of Variance and Standard Deviation

Let X be a **random variable** and $a, b \in \mathbb{R}$. Then

- $\mathrm{Var}(aX + b) = a^2\,\mathrm{Var}(X)$
- $\mathrm{SD}(aX + b) = |a|\,\mathrm{SD}(X)$

4.7 Discrete Random Variables Arising from Repeated Trials

4.7.1 Bernoulli Trials

A **Bernoulli trial**, denoted by Bernoulli(p) trials, is an experiment with only two outcomes:

1. success with probability p; and
2. failure with probability $q = 1 - p$.

4.7.2 Bernoulli Random Variable

4.7.2.1 Definition of Bernoulli Random Variable

Definition. If X is a Bernoulli random variable, denoted by $X \sim Be(p)$. Only one Bernoulli(p) trial is performed, and

$$X = \begin{cases} 1, & \textbf{if it is a success} \\ 0, & \textbf{if it is a failure} \end{cases}$$

with

- $P(X = 0) = 1 - p$
- $P(X = 1) = p$

4.7.2.2 Properties of Bernoulli Random Variable

Let $X \sim Be(p)$. Then

$$E(X) = 0(1 - p) + 1(p) = p$$
$$\mathrm{Var}(X) = p(1 - p)$$

4.7.3 Binomial Random Variable

4.7.3.1 Definition of Binomial Random Variable

Definition. If X is a binomial random variable, denoted by $X \sim Bin(n,p)$, n Bernoulli(p) trials are performed (under identical conditions and independently), and

$$X = \textbf{number of successes} \text{ in } n \text{ Bernoulli}(p) \text{ trials}$$

Thus, X takes values $0, 1, 2, \ldots, n$ and for $0 \leq k \leq n$,

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

4.7.3.2 Properties of Binomial Random Variable

Let $X \sim Bin(n,p)$. Then

$$E(X) = np$$
$$\text{Var}(X) = np(1-p)$$

4.7.3.3 Binomial Random Variable is a Sum of Bernoulli Random Variables

Let $X_i \sim Be(p)$ for $i = 1, \ldots, n$, where $X_1, \ldots, X_n$ are independent. Let

$$X = X_1 + X_2 + \cdots + X_n$$

Then $X \sim Bin(n,p)$.

4.7.4 Geometric Random Variable

4.7.4.1 Version 1: Include Success

4.7.4.1.1 Definition of Geometric Random Variable

Definition. If X is a geometric random variable, denoted by $X \sim Geom(p)$, Bernoulli(p) trials are performed (under identical conditions and independently) until the first success is obtained.

- $X =$ **number of Bernoulli(p) trials** required to obtain the first success
  - **including** the trial leading to the first success.

Thus, X takes values $1, 2, 3, \ldots$ and for $k = 1, 2, 3, \ldots$,

$$P(X = k) = pq^{k-1}$$

4.7.4.1.2 Properties of Geometric Random Variable

Let $X \sim Geom(p)$. Then

$$E(X) = \frac{1}{p}$$
$$\text{Var}(X) = \frac{1-p}{p^2}$$

4.7.4.2 Version 2: Exclude Success

4.7.4.2.1 Definition of Geometric Random Variable

Definition. If X' is a geometric random variable, denoted by $X' \sim Geom(p)$, Bernoulli(p) trials are performed (under identical conditions and independently) until the first success is obtained.

- X' = **number of failures** of Bernoulli(p) trials **before obtaining the first success**.

Thus, X' takes values $0, 1, 2, \ldots$ and for $k = 0, 1, 2, \ldots$,

$$P(X' = k) = pq^k$$

4.7.4.2.2 Properties of Geometric Random Variable

Let $X' \sim Geom(p)$. Then

$$E(X') = \frac{1-p}{p}$$
$$\text{Var}(X') = \frac{1-p}{p^2}$$

4.7.4.3 Relationship between Version 1 and 2

Let $X \sim Geom(p)$ from Version 1 and $X' \sim Geom(p)$ from Version 2. Then $X = X' + 1$.

4.7.5 Negative Binomial Random Variable

4.7.5.1 Definition of Negative Binomial Random Variable

Definition. If X is a negative binomial random variable, denoted by $X \sim NB(r, p)$, Bernoulli(p) trials are performed until the first r successes are obtained, and

- X = **number of Bernoulli (p) trials** required to obtain r **successes**.

Thus, X takes values $r, r+1, \ldots$ and for $k = r, r+1, \ldots$,

$$P(X = k) = \binom{k-1}{r-1} p^r q^{k-r}$$

4.7.5.2 Properties of Negative Binomial Random Variable

Let $X \sim NB(r, p)$. Then

$$E(X) = \frac{r}{p}$$
$$\text{Var}(X) = \frac{r(1-p)}{p^2}$$

4.7.5.3 Geometric Random Variable is a Special Case of Negative Binomial Random Variable

Note that $Geom(p) = NB(1, p)$.

4.7.5.4 Negative Binomial Random Variable is the Sum of Geometric Random Variables

Let $X_i \sim Geo(p)$, for $i = 1, 2, \ldots, r$. Suppose $X_1, \ldots, X_r$ are independent. Let

$$X = X_1 + \cdots + X_r$$

Then $X \sim NB(r, p)$.

4.8 Poisson Random Variable

4.8.1 Definition of Poisson Random Variable

Definition. If X is a Poisson random variable, denoted by $X \sim Po(\lambda)$, then

- X takes values $0, 1, 2, \ldots$

For $k = 0, 1, 2, \ldots$,

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

4.8.2 Properties of Poisson Random Variable

Let $X \sim Po(\lambda)$. Then

$$\begin{aligned} E(X) &= \lambda \\ \mathrm{Var}(X) &= \lambda \end{aligned}$$

4.8.3 Poisson Random Variable Approximates Binomial Random Variable for large n and moderate np

Let $X \sim Bin(n, p)$ and $Y \sim Po(\lambda)$ where $\lambda = np$. Then for large values of n and moderate values of $\lambda = np$,

$$P(X = k) \approx P(Y = k), \quad \forall k = 0, 1, 2, \ldots$$

4.8.4 Many Random Variables Approximate Poisson Random Variable

Each of the beforementioned random variables, and numerous other random variables, are approximately Poisson, because of Poisson approximation to the binomial.