TAs' Notes - STAT-4300 Spring'26

February 2, 2026

Contents

1 Class 1 - Set Theory, Probability, and Indicator Functions

1.1 Set Theory

1.1.1 Review of basic definitions

We begin with definitions that should be familiar.

Definition 1.1 (Set). : A **set** is a collection of elements.

Definition 1.2 (Subset and superset). : A set A is a **subset** of B if every element of A is also an element of B, denoted

$$A \subset B$$

Equivalently, B is a **superset** of A.

Definition 1.3 (Null set and empty set). The set with no elements is called the **null set** or the **empty set**, denoted $\emptyset$.

Remark. The null set is a subset of any set. I.e. for any A,

$$\emptyset \subset A$$

Definition 1.4. (Universal set): The **universal set** is the set of all things that we could possibly consider in the context we are studying.

Remark. In probability, the universal set is typically the sample space denoted Ω

1.1.2 Review of set operations

Except for **symmetric difference**, most of these set operations should be familiar.

Definition 1.5 (Union). : The **union** of two sets, A and B, is a set containing all the elements that are in A or in B (possibly both).

The union of two sets, A and B is denoted

$$A \cup B$$

The union of three or more sets, say $A_1, A_2, \ldots A_n$ is denoted

$$A_1 \cup A_2 \cup A_3 \ldots \cup A_n = \bigcup_{i=1}^{n} A_i$$

Definition 1.6 (Intersection). : The **intersection** of two sets, A and B, is a set containing all the elements that are both in A and B.

The intersection of two sets, A and B is denoted

$$A \cap B$$

The intersection of three or more sets, say $A_1, A_2, \ldots A_n$ is denoted

$$A_1 \cap A_2 \cap A_3 \ldots \cap A_n = \bigcap_{i=1}^{n} A_i$$

Definition 1.7 (Complement). : The **complement** of a set A denoted by A^C is the set of all elements that are in the universal set S but are not in A.

Definition 1.8 (Difference, subtraction of sets). : The **difference** of two sets is defined where $A - B$ consists of the elements that are in A but not in B. This is denoted

$$A - B = A \setminus B = \{x \in \Omega : x \in A, x \notin B\}$$

Remark. From the above definition, it should be clear that

$$A - B = A \cap B^C$$

Definition 1.9 (Mutually exclusive, disjoint). : Two sets, A and B, are **mutually exclusive** or **disjoint** if they do not have any shared elements, i.e.

$$A \cap B = \emptyset$$

For three or more sets, the sets having a trivial intersection does not mean they are disjoint. Instead, we require the stronger condition below.

Definition 1.10 (Pairwise disjoint). : Several sets are **pairwise disjoint** if no two sets share a common element.

1.1.3 Review of common set properties

Theorem 1.11 (De Morgan's law). : For sets $A_1, A_2, \dots A_n$, we have

- $(A_1 \cup A_2 \cup \dots A_n)^C = A_1^C \cap A_2^C \cap \dots A_n^C$
- $(A_1 \cap A_2 \cap \dots A_n)^C = A_1^C \cup A_2^C \cup \dots A_n^C$

Theorem 1.12 (Distributive law). : For any sets A, B, C, we have

- $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
- $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

1.1.4 The symmetric difference

Pay especially close attention to the symmetric difference, which may not have been emphasized upon in previous courses.

Definition 1.13 (Symmetric difference). : The **symmetric difference** of two sets, A and B, is defined as the set of elements that are only in A or in B, but not both. This is denoted

$$A \bigtriangleup B$$

The symmetric difference of two sets is their union, minus their intersection

$$A \bigtriangleup B = (A \cup B) \setminus (A \cap B)$$

Result 1.14. Properties of the symmetric difference

1. Commutativity
$$A \bigtriangleup B = B \bigtriangleup A$$
2. Associativity
$$(A \bigtriangleup B) \bigtriangleup C = A \bigtriangleup (B \bigtriangleup C)$$
3. Distributivity of the intersection
$$A \cap (B \bigtriangleup C) = (A \cap B) \bigtriangleup (A \cap C)$$
4. The symmetric difference is trivial if and only if the two sets are equal
$$A \bigtriangleup B = \emptyset \Leftrightarrow A = B$$
5. Taking complement with respect to the same universal set,
$$A \bigtriangleup B = A^C \bigtriangleup B^C$$
6. The symmetric difference is a subset of the union
$$A \bigtriangleup B \subseteq A \cup B$$
7. The symmetric difference is equal to the union if and only if the sets are disjoint
$$A \bigtriangleup B = A \cup B \Leftrightarrow A \cap B = \emptyset$$
8. The symmetric difference and the intersection partition the union, since
$$(A \bigtriangleup B) \cap (A \cap B) = \emptyset$$
$$(A \bigtriangleup B) \cup (A \cap B) = A \cup B$$
The concept of partition will be introduced in Class 2

9. We can define the union using
$$A \cup B = (A \triangle B) \triangle (A \cap B)$$

The following result is especially important. Hence we discuss it separately.

Lemma 1.15. Given arbitrary sets $A_1, A_2, \dots A_n$ and $B_1, B_2, \dots B_n$,
$$\left(\bigcup_{i=1}^{n} A_i\right) \triangle \left(\bigcup_{i=1}^{n} B_i\right) \subseteq \bigcup_{i=1}^{n} (A_i \triangle B_i)$$

The proof of this lemma is revisited later in the lecture after the introduction of indicator functions. We will see that the proof is simple once we introduce indicators.

1.2 Probabilities

1.2.1 Review

Definition 1.16 (Random experiment, outcome, sample space)**.** :

- A **random experiment** is a process by which we observe something uncertain.
- The result of a random experiment is an **outcome**.
- The set of possible outcomes is the **sample space**.

Definition 1.17 (Event)**.** : An **event** E is a subset of the sample space, i.e. a collection of outcomes.

Remark. If A and B are events, then $A \cup B$ and $A \cap B$ are also events.

$A \cup B$ occurs if A **or** B occurs.

$A \cap B$ occurs if A **and** B occurs.

Definition 1.18 (Probability)**.** : The **probability** measure of event A is denoted $P(A)$.

1.2.2 Axioms of probability

Definition 1.19 (Axioms of probability)**.** : The axioms of probability state that

1. For any event A, $P(A) \geq 0$
2. Probability of the sample space Ω is $P(\Omega) = 1$
3. If $A_1, A_2, A_3 \dots$ are disjoint, then
$$P(A_1 \cup A_2 \cup A_3 \dots) = P(A_1) + P(A_2) + P(A_3) + \dots$$

1.2.3 Inclusion-Exclusion Principle

Result 1.20 (Inclusion-exclusion)**.** : By the **inclusion-exclusion principle**, we have
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

In general for n events $A_1, \dots A_n$,

$$\begin{aligned}
P\left(\cup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) \\
&- \sum_{i<j} P(A_i \cap A_j) \\
&+ \sum_{i<j<k} P(A_i \cap A_j \cap A_k) \\
&\vdots \\
&+ (-1)^{n-1} P\left(\bigcap_{i=1}^{n} A_i\right)
\end{aligned}$$

1.3 Indicator Functions

1.3.1 Definition

Definition 1.21 (Indicator function)**.** : Given an arbitrary set X, and a subset $A \subseteq X$, the **indidcator function** of A is

$$\mathbf{1}_A(x) = \begin{cases} 1 \text{ if } x \in A \\ 0 \text{ if } x \notin A \end{cases}$$

1.3.2 Properties of the indicator function

Result 1.22. Properties of the indicator function

1. Indicator of the intersection is the product of indicators

$$\mathbf{1}_{A\cap B}(x) = \min\{\mathbf{1}_A(x), \mathbf{1}_B(x)\} = \mathbf{1}_A(x) \cdot \mathbf{1}_B(x)$$

2. Ihe indicator of the union is sum of indicators minus their product

$$\begin{aligned}
\mathbf{1}_{A\cup B}(x) &= \max\{\mathbf{1}_A(x), \mathbf{1}_B(x)\} \\
&= \mathbf{1}_A(x) + \mathbf{1}_B(x) - \mathbf{1}_A(x) \cdot \mathbf{1}_B(x) \\
&= \mathbf{1}_A(x) + \mathbf{1}_B(x) - \mathbf{1}_{A\cap B}(x)
\end{aligned}$$

3. Indicator of the complement

$$\mathbf{1}_{A^C} = 1 - \mathbf{1}_A$$

4. If A, B disjoint

$$\begin{aligned}
\mathbf{1}_{A\cup B} &= \mathbf{1}_A + \mathbf{1}_B \\
\mathbf{1}_{A\cap B} &= 0
\end{aligned}$$

5. Indicators of subsets

$$A \subseteq B \Leftrightarrow \mathbf{1}_A \leq \mathbf{1}_B$$

6. Indicators of difference of subsets

$$\begin{aligned}
\mathbf{1}_{A-B} &= \mathbf{1}_{A\cap B^C} \quad \text{by definition of set subtraction} \\
&= \mathbf{1}_A \cdot \mathbf{1}_{B^C} \quad \text{by indicator of intersections} \\
&= \mathbf{1}_A\big(1 - \mathbf{1}_B\big) \quad \text{by indicator of complement} \\
&= \mathbf{1}_A - \mathbf{1}_{A\cap B} \quad \text{by indicator of intersections}
\end{aligned}$$

7. Indicators of symmetric difference

$$
\begin{aligned}
\mathbf{1}_{A \,\triangle\, B} &= \mathbf{1}_{(A\cup B)\setminus(A\cap B)} \quad \text{by definition of symmetric difference} \\
&= \mathbf{1}_{A\cup B} \cdot \mathbf{1}_{(A\cap B)^C} \quad \text{by definition of set subtraction} \\
&= \mathbf{1}_{A\cup B}\big(1 - \mathbf{1}_{A\cap B}\big) \quad \text{by indicator of complement} \\
&= \big(\mathbf{1}_A + \mathbf{1}_B - \mathbf{1}_{A\cap B}\big)\big(1 - \mathbf{1}_{A\cap B}\big) \quad \text{by indicator of union} \\
&= \mathbf{1}_A + \mathbf{1}_B - 2\,\mathbf{1}_{A\cap B} \quad \text{since } \mathbf{1}_{A\cap B}^2 = \mathbf{1}_{A\cap B} \\
&= \big|\mathbf{1}_A - \mathbf{1}_B\big|
\end{aligned}
$$

1.3.3 Demonstrating the usefulness of indicators

Recall Lemma 1.15. Given arbitrary sets $A_1, \dots, A_n$ and $B_1, \dots, B_n$,

$$\left(\bigcup_{i=1}^{n} A_i\right) \triangle \left(\bigcup_{i=1}^{n} B_i\right) \subseteq \bigcup_{i=1}^{n} (A_i \triangle B_i).$$

We can prove this lemma by reducing the set inclusion to an inequality involving indicator functions.

Proof. By property 1.23.5, it suffices to show

$$\mathbf{1}_{\left(\cup_{i=1}^{n} A_i\right) \,\triangle\, \left(\cup_{i=1}^{n} B_i\right)} \le \mathbf{1}_{\cup_{i=1}^{n}(A_i \,\triangle\, B_i)}$$

By property 1.23.7 (indicator of symmetric differences), the LHS can be written as

$$\left|\mathbf{1}_{\cup_{i=1}^{n} A_i}(x) - \mathbf{1}_{\cup_{i=1}^{n} B_i}(x)\right|$$

By property 1.23.2 (indicator of unions)

$$\mathbf{1}_{\cup_{i=1}^{n} A_i}(x) = \max_{1\le i\le n} \mathbf{1}_{A_i}(x), \qquad \mathbf{1}_{\cup_{i=1}^{n} B_i}(x) = \max_{1\le i\le n} \mathbf{1}_{B_i}(x),$$

Hence we get

$$\left|\max_{1\le i\le n} \mathbf{1}_{A_i}(x) - \max_{1\le i\le n} \mathbf{1}_{B_i}(x)\right| \le \mathbf{1}_{\cup_{i=1}^{n}(A_i \,\triangle\, B_i)}(x)$$

We now prove this inequality by enumerating the possible values of the right-hand side.

Case 1: RHS $= 0$

$$
\begin{aligned}
&\mathbf{1}_{\cup_{i=1}^{n}(A_i \,\triangle\, B_i)}(x) = 0 \\
\implies\ & x \notin A_i \triangle B_i \text{ for all } i \\
\implies\ & \mathbf{1}_{A_i}(x) = \mathbf{1}_{B_i}(x) \text{ for all } i \\
\implies\ & \max_i \mathbf{1}_{A_i}(x) = \max_i \mathbf{1}_{B_i}(x) \\
\implies\ & \left|\max_{1\le i\le n} \mathbf{1}_{A_i}(x) - \max_{1\le i\le n} \mathbf{1}_{B_i}(x)\right| = 0
\end{aligned}
$$

Hence the inequality holds.

Case 2: RHS $= 1$
Since the LHS is the absolute value of the difference of indicators, it takes values 0 or 1, and this is a simple upper bound.

In both cases, the inequality holds. Since this inequality is equivalent to the desired set inclusion, the lemma follows. ✓

2 Class 2 - Conditional Probability, Independence, and Law of Total Probability

2.1 Conditional Probability

Definition 2.1 (Conditional Probability)**.** Let A, B be events in a sample space S, with $P(B) > 0$, then the **conditional probability** of A given B is defined as

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

Remark. Conditional probability is a probability. Fix B, define $P_B(\cdot) = P(\cdot|B)$. Then, $P(\cdot|B)$ satisfies the three axioms of probability on the reduced sample space B.

1. Nonnegativity. For any A, $P(A|B) \geq 0$
2. Normalized. $P(B|B) = 1$
3. Countable additivity. If $A_1, A_2, A_3, \ldots$ disjoint, then

$$P(A_1 \cup A_2 \cup A_3 \ldots |B) = P(A_1|B) + P(A_2|B) + P(A_3|B) + \ldots$$

Proof. We verify the three axioms:

1. For any A, since $P(A \cap B) \geq 0$ and $P(B) > 0$, we have

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \geq 0$$

2.

$$P(B|B) = \frac{P(B \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$$

3. If $A_1, A_2, A_3, \ldots$ are disjoint, then

$$\begin{aligned}
P(A_1 \cup A_2 \cup A_3 \ldots |B) &= \frac{P((A_1 \cup A_2 \cup A_3 \ldots) \cap B)}{P(B)} \\
&= \frac{P((A_1 \cap B) \cup (A_2 \cap B) \cup (A_3 \cap B) \ldots)}{P(B)} && \text{(by distributing the intersection)} \\
&= \frac{P(A_1 \cap B) + P(A_2 \cap B) + P(A_3 \cap B) + \ldots}{P(B)} && \text{(by axiom 3 of probability)} \\
&= P(A_1|B) + P(A_2|B) + P(A_3|B) + \ldots
\end{aligned}$$

✓

Remark. In class, prof used a slightly different notation to prove countable additivity.

$$\begin{aligned}
P_B\left(\bigcup_{i=1}^{n} A_i\right) &= \frac{P\left(B \cap \bigcup_{i=1}^{n} A_i\right)}{P(B)} \\
&= \frac{P\left(\bigcup_{i=1}^{n} (B \cap A_i)\right)}{P(B)} \\
&= \frac{\sum\limits_{i=1}^{n} P\left(B \cap A_i\right)}{P(B)} \\
&= \sum_{i=1}^{n} P(A_i|B)
\end{aligned}$$

2.2 Independence

Definition 2.2 (Independence)**.** Events A, B are **independent** if and only if

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

Equivalent

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

Independence is sometimes denoted

$$A \cap B$$

Remark (Independence vs Disjointness)**.** Two events are **disjoint** if $A \cap B = \emptyset$.

Two events are independent if $P(A \cap B) = P(A)P(B)$ or $P(A|B) = P(A)$

If two events are disjoint (and each event has non-zero probability), knowing one event provides full information about the other. Therefore, disjoint events are **not** independent.

Result 2.3 (Independence and complements)**.** Suppose A, B are independent events. Then the following pairs of events are also independent:

- A^c and B
- A and B^c
- A^c and B^c

2.3 Law of Total Probability

Definition 2.4 (Partition)**.** We say that a collection of nonempty sets $A_1, A_2, \ldots$ form a **partition** of A if they are disjoint and their union is A.

That is,

- pairwise disjoint: $A_i \cap A_j = \emptyset$ for all $i \neq j$
- collectively exhaustive: $\bigcup_i A_i = A$
- nonempty: $A_i \neq \emptyset$ for all i

Remark. If $A_1, A_2, \ldots$ parition Ω, then any $\omega \in \Omega$ lives in exactly one of the A_i's.

Theorem 2.5 (Law of Total Probability)**.** If $B_1, B_2, \ldots$ is a partition of the sample space S, then for any event A, we have

$$P(A) = \sum_i P(A \cap B_i) = \sum_i P(A|B_i)P(B_i)$$

Proof. Decompose A into disjoint unions and apply axiom 3.

$$\begin{aligned} A &= \bigcup_{i=1}^{n} (A \cap B_i) \text{ and union is disjoint} \\ P(A) &= P\left(\bigcup_{i=1}^{n} (A \cap B_i)\right) \\ &= \sum_{i=1}^{n} P(A \cap B_i) \quad \text{(by axiom 3)} \\ &= \sum_{i=1}^{n} P(A|B_i)P(B_i) \end{aligned}$$

✓

3 Class 3 - Bayes Rule and Counting

3.1 Bayes Rule

Theorem 3.1 (Bayes Rule)**.** For any two events A, B, where $P(A) \neq 0, P(B) \neq 0$, we have

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

Remark.

$$\underbrace{P(A \mid B)}_{\text{posterior}} = \underbrace{\frac{P(B \mid A)}{P(B)}}_{\text{update function}} \underbrace{P(A)}_{\text{prior}}.$$

Proof.

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

✓

Example (Biased coins)**.** Three coins with head probabilities $(p_1, p_2, p_3) = (.9, .6, .3)$. Suppose we select a coin at random and observe (H, H, H, T).

What is the probability that we chose coin i given the data?

$$P(i \text{ chosen}|D) = \frac{P(D|i \text{ chosen })}{P(D)} \cdot P(i \text{ chosen})$$

More concretely, say $i = 1$

$$\begin{aligned} P(D|1 \text{ chosen}) &= P(\{H, H, H, T\}|1 \text{ chosen}) \\ &= P(H|1 \text{ chosen})^3 \cdot P(T|1 \text{ chosen}) \text{ by independence} \\ &= 0.9^3 \cdot 0.1 \end{aligned}$$

From last week, we apply law of total probability to find $P(D)$

$$P(D) = \sum_{i \in \{1,2,3\}} P(D|i \text{ chosen}) \cdot P(i \text{ chosen})$$

More generally, the probability that we chose coin i given the data is

$$P(i \text{ chosen}|D) = \frac{p_i^3(1-p_i)}{\sum_{j=1}^{3} p_j^3(1-p_j)} = \begin{cases} 0.56 & i = 1 \\ 0.33 & i = 2 \\ 0.11 & i = 3 \end{cases}$$

3.2 Counting

3.2.1 Counting Principles

Definition 3.2 (Counting Principle)**.** Let $A_1, \ldots A_n$ be finite sets with $|A_i| = k_i$, and $(a_1, \ldots a_n) \in A_1 \times \ldots A_n$, then

$$|A_1 \times A_2 \times \ldots A_n| = |A_1| \cdot |A_2| \cdot \ldots \cdot |A_n| = \prod_{i=1}^{n} k_i$$

Example (Phone PIN code)**.** Consider 4 digit PIN codes, (a_1, a_2, a_3, a_4), then

$$|A_1| = |A_2| = |A_3| = |A_4| = 10$$

Thus, by the counting principle, the total number of possible PIN codes is

$$|A_1 \times A_2 \times A_3 \times A_4| = 10^4 = 10,000$$

Remark. Every n element set is equivalent to $\{1, 2, \dots n\}$.

3.2.2 Sampling Taxonomy - 4 kinds of sampling regimes

We want to count the number of ways to draw k things from n-elements set, i.e. $S = \{1, 2, \dots n\}$. There are 4 scenarios.

	Ordered	Unordered
With Replacement	Sequences	Multisets
Without Replacement	Permutations	Combinations

3.2.3 Ordered, without replacement - Permutations

Result 3.3 (Counting Permutations)**.** We want to count the number of ordered samples of size k drawn without replacement from an n element set. Formally, we want to count the size of the sample space

$$\Omega = \{(a_1, \dots a_k) : a_i \neq a_j \text{ if } i \neq j\}$$

We call the number of permutations "n permute k" and denote it by ${}_nP_k$.
By counting principle,

$${}_nP_k = |\Omega| = n \cdot (n-1) \cdot (n-2) \cdots (n-k+1) = \frac{n!}{(n-k)!}$$

Example (Collision problem)**.** In a group of n people, what is the probability p_n that at least two have the same birthday? Assume birthdays are uniform over 365 days and independent.

Instead of counting the number of ways with at least 2, which would require us to consider the number of cases with exactly 2, exactly 3, etc., we use the **complement rule**.

$$\begin{aligned} p_n &= 1 - P(\text{no collision}) \\ &= 1 - \frac{\text{number of ways to assign birthdays with no collision}}{\text{number ways to assign birthdays}} \end{aligned}$$

We can define

$$\begin{aligned} \Omega &= \{(a_1, \dots a_n) \in \{1, \dots 365\}^n\} \\ A &= \{(a_1, \dots a_n) \in \{1, \dots 365\}^n, a_i \neq a_j \text{ if } i \neq j\} = \{\text{permutations of } \{1, \dots 365\}\} \end{aligned}$$

Then

$$\begin{aligned} |\Omega| &= 365^n \\ |A| &= {}_{365}P_n = \frac{365!}{(365-n)!} \end{aligned}$$

$$\begin{aligned} p_n &= 1 - P(\text{no collision}) \\ &= 1 - \frac{\text{number of ways to assign birthdays with no collision}}{\text{number ways to assign birthdays}} \\ &= 1 - \frac{{}_{365}P_n}{365^n} \\ &= 1 - \frac{365!}{(365-n)!365^n} \end{aligned}$$

We can simplify this equation

$$\begin{aligned} p_n &= 1 - \frac{365 \cdot 364 \dots (365-n+1)}{365^n} \\ &= 1 - 1 \cdot \frac{365}{365} \dots (365-n+1)/365 \\ &= 1 - \prod_{i}^{n-1} \left(1 - \frac{i}{365}\right) \end{aligned}$$

Since $1 - x \leq e^{-x}$, we have

$$1 - \frac{i}{365} \leq e^{-\frac{i}{365}}$$

Hence

$$\begin{aligned}\prod_{i=1}^{n-1}\left(1 - \frac{i}{365}\right) &\leq \prod_{i=1}^{n-1} e^{-\frac{i}{365}} \\ &= e^{-\sum\limits_{i=0}^{n-1}\frac{i}{365}} \\ &= e^{-\frac{n(n-1)}{2\cdot 365}}\end{aligned}$$

Therefore

$$p_n = 1 - \prod_{i=0}^{n-1}\left(1 - \frac{i}{365}\right) \geq 1 - e^{\frac{-n(n-1)}{2\cdot 365}}$$

Hence, p_n becomes significant when n is on the order of $\sqrt{365} \approx 19$.

To see a collision among m categories, we need about $\sqrt{m}$ samples.

3.2.4 Unordered, without replacement - Combinations

Remark. Combinations are permutations modulo order.

Result 3.4 (Counting Combinations)**.** We want to count the number of unordered samples of size k drawn without replacement from an n element set. This is equivalent to counting the number of unordered k-subsets from n elements. Formally, we want to count the size of the sample space

$$\Omega = \{\{a_1, a_2 \ldots a_n\} \subseteq \{1, 2, \ldots n\}\}$$

We call the number of combinations "n choose k" and denote it by ${}_nC_k$ or $\binom{n}{k}$

$${}_nC_k = |\Omega| = \binom{n}{k}$$

For each unordered subset, we can generate $k!$ ordered sequences. Hence

$$k!\,|\Omega| = k!\;{}_nC_k = {}_nP_k$$

Therefore

$${}_nC_k = |\Omega| = \binom{n}{k} = \frac{{}_nP_k}{k!} = \frac{n!}{k!(n-k)!}$$

Example (Flipping coins)**.** Flip a coin 20 times. Each ordered sequence of H/T is equally likely. Which outcome has more ways to happen?

- exactly 10 heads
- exactly 2 heads

The number of outcomes are

- exactly 10 heads: ${}_{20}C_{10} = \binom{20}{10} = 184,756$ ways
- exactly 2 heads: ${}_{20}C_2 = \binom{20}{2} = 190$ ways

3.2.5 Unordered, with replacement - Multisets

Definition 3.5 (Multisets)**.** Fix $S = \{1, 2, \ldots n\}$. A **multiset** on S is a function $m : S \to \mathbb{N}$ where $m(\sigma)$ is the multiplicity of $\sigma \in S$. A **multiset of size** k is one such that

$$\sum_{\sigma \in S} m(\sigma) = k$$

Result 3.6 (Counting Multisets)**.** We want to count the number of unordered samples of size k drawn with replacement from an n element set. This is equivalent to counting the number of multisets of size k from n elements.

The sample space is

$$\Omega = \{m : S \to \mathbb{N}, \sum_{\sigma \in S} m(\sigma) = k\}$$

We call the number of multisets "n multichoose k" and denote it by $\left(\!\binom{n}{k}\!\right)$.

We count this using the **bijective argument**: Let $x_i = m(i)$ for $i \in \{1, 2, \ldots n\}$, then

$$x_1 + x_2 \ldots x_n = k \text{ where } (x_1, \ldots x_n) \in \{0, 1, 2, \ldots\}^n = \mathbb{N}^n$$

We build a bijection

$$\Phi\left(x_1, \ldots x_n\right) = \underbrace{* * \ldots *}_{x_1} | \underbrace{* * \ldots *}_{x_2} | \ldots | \underbrace{* * \ldots *}_{x_n}$$

This is equivalent to arranging k stars and $n-1$ bars, which is a total of $k+n-1$ symbols. We need to choose $n-1$ positions for the bars, hence

$$|\Omega| = \left(\!\!\binom{n}{k}\!\!\right) = \binom{n+k-1}{n-1} = \binom{n+k-1}{k}$$

3.2.6 Summary

As a summary, we have the following counting formulas:

	Ordered	**Unordered**
With Replacement	Sequences: $\prod_{i=1}^{n} k_i$	Multisets: $\left(\!\!\binom{n}{k}\!\!\right) = \binom{n+k-1}{k}$
Without Replacement	Permutations: ${}_nP_k = \frac{n!}{(n-k)!}$	Combinations: ${}_nC_k = \binom{n}{k} = \frac{n!}{k!(n-k)!}$

4 Class 4 - Multinomial Coefficients and Discrete Random Variables

4.1 Multinomial Coefficients

Definition 4.1 (Multinomial coefficient)**.** Let S be a finite set with $|S| = n$. For nonnegative integers $m_1, \dots, m_n$ satisfying

$$\sum_{i=1}^{n} m_i = k,$$

the **multinomial coefficient**

$$\binom{k}{m_1, \dots, m_n} = \frac{k!}{m_1! \cdots m_n!}$$

counts the number of ordered sequences in S^k in which element $i \in S$ appears exactly m_i times.

Example. Let $\Omega = \{1, 2, \dots 6\}^{60}$, and each ordered outcome is equally likely. The most likely outcome is the one where each element appears exactly 10 times. The number of such outcomes is given by the multinomial coefficient

$$\binom{60}{10, 10, 10, 10, 10, 10} = \frac{60!}{(10!)^6}$$

Proof: Suppose two of the multiplicities differ by more than 1, i.e. $m_a \geq m_b + 2$ for some a, b. Then define

$$m'_a = m_a - 1, \quad m'_b = m_b + 1$$

We then compare the ratio of the two multinomial coefficients:

$$\frac{\binom{60}{\dots m'_a, \dots, m'_b, \dots}}{\binom{60}{\dots m_a, \dots, m_b, \dots}} = \frac{m_a}{m_b + 1} > 1$$

This means we can always increase the multinomial coefficient by balancing the multiplicities. Hence the maximum occurs when all multiplicities are within 1 apart.

4.2 Discrete Random Variables

4.2.1 Discrete Random Variables, PMF, CDF

Definition 4.2 (Discrete random variables)**.** A **discrete random variable** is a function $X : \Omega \to A \subseteq \mathbb{R}$ where A is countable (i.e. there is a one-to-one mapping between A and $\mathbb{N}$, $|A| \leq |\mathbb{N}|$).

Remark. For discrete X, Ω can be arbitrarily complicated, but the range of X, A needs to be countable.

Definition 4.3 (Probability mass function)**.** The **probability mass function** (pmf) of a discrete random variable X is the function $p_X : \mathbb{R} \to [0, 1]$ defined by

$$p_X(x) = P(X = x) = P(\{\omega \in \Omega : X(\omega) = x\})$$

Definition 4.4 (Support)**.** The **support** of a discrete random variable X is the set of values in $\mathbb{R}$ where the pmf is positive:

$$\mathrm{supp}(X) = \{x \in \mathbb{R} : p_X(x) > 0\}$$

Remark. The pmf satisfies

- $p_X(x) \geq 0$ for all $x \in \mathbb{R}$
- $\sum\limits_{x \in A} p_X(x) = 1$, where A is the range of X

Definition 4.5 (Cumulative distribution function)**.** The **cumulative distribution function** (cdf) of a discrete random variable X is the function $F_X : \mathbb{R} \to [0, 1]$ defined by

$$F_X(t) = P(X \leq t) = \sum_{a \leq t} p_X(a)$$

Example. Let $D_1, D_2 \in \{1, 2, \ldots 6\}$ be outcomes of 2 dice rolls. Define

$$S := D_1 + D_2 \in A = \{2, 3, \ldots 12\}$$

The sample space is

$$\Omega = \{1, 2, \ldots 6\}^2$$

The number of outcomes corresponding to sum s is

$$N(s) = \begin{cases} s - 1 & \text{if } 2 \leq s \leq 7 \\ 13 - s & \text{if } 8 \leq s \leq 12 \end{cases}$$

The probability mass function of S is

$$p_S(s) = \frac{N(s)}{|\Omega|} = \frac{N(s)}{36}$$

4.2.2 Expectations and Variance

Definition 4.6 (Expectation of discrete random variable)**.** If X is a discrete random variable with pmf p_X, and $\sum\limits_{x \in A} |x|\, p_X(x) < \infty$, then the **expectation** of X is defined as

$$\mathbb{E}[X] = \sum_{x \in \text{supp}(X)} x \cdot p_X(x)$$

If $g : \mathbb{R} \to \mathbb{R}$, then the **expectation** of $g(X)$ is defined as

$$\mathbb{E}[g(X)] = \sum_{x \in \text{supp}(X)} g(x) \cdot p_X(x)$$

Definition 4.7 (Variance)**.** Given a random variable X with finite expectation, the **variance** of X is defined as

$$\text{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2$$

4.2.3 Indicator random variables

Definition 4.8 (Indicator random variables)**.** For an event $A \subseteq \Omega$, the **indicator function** is a discrete random variable $\mathbf{1}_A : \Omega \to \{0, 1\}$ defined by

$$\mathbf{1}_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A \end{cases}$$

Result 4.9 (Expectation of indicator r.v.)**.** The expectation of an indicator random variable is the probability of the corresponding event:

$$\mathbb{E}[\mathbf{1}_A] = P(A)$$

Proof. By definition of expectation

$$\begin{aligned} \mathbb{E}[\mathbf{1}_A] &= 1 \cdot P(\mathbf{1}_A = 1) + 0 \cdot P(\mathbf{1}_A = 0) \\ &= P(A) \end{aligned}$$

✓

Theorem 4.10 (Linearity of Expectation)**.** Let X, Y be discrete random variables, and a, b be constants, then

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$$

Remark. In general, $\mathbb{E}[g(X)] \neq g(\mathbb{E}[X])$.

Theorem 4.11 (Law of total probability)**.** Let $A_1, \ldots A_n$ be a partition of Ω, define a discrete random variable $I : \Omega \to \{1, 2, \ldots n\}$ by

$$I(w) = i \text{ if } w \in A_i$$

Then

$$P(I = i) = P(A_i)$$

Define $g(I) : P(B|I = i)$. Then the law of total probability can be stated as

$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i) = \mathbb{E}\left[P(B|I)\right]$$