I. Set Theory and Generalized Inclusion-Exclusion

I.A. Fundamental Set Identities and Cardinality

Definition 0.1 (Key Set Operations)**.** Let A, B be sets in a universal set U.

(i) **Difference:** $A \setminus B = A \cap B^c$.

(ii) **Symmetric Difference:** $A \Delta B = (A \setminus B) \cup (B \setminus A) = (A \cup B) \setminus (A \cap B)$.

(iii) **Cardinality of Symmetric Difference:** $|A \Delta B| = |A| + |B| - 2\,|A \cap B|$.

(iv) **De Morgan's Laws (Generalized):** $\left(\bigcup_{i=1}^{n} A_i\right)^c = \bigcap_{i=1}^{n} A_i{}^c$.

Theorem 0.1 (PIE for Two and Three Sets)**.** (i) $|A \cup B| = |A| + |B| - |A \cap B|$.

(ii) $|A \cup B \cup C| = |A| + |B| + |C| - (|A \cap B| + |A \cap C| + |B \cap C|) + |A \cap B \cap C|$.

Proof (Detailed).[Proof for Two Sets $|A \cup B|$] $A \cup B$ can be written as the disjoint union $A \cup (B \setminus A)$. By the Sum Rule for disjoint sets:

$$|A \cup B| = |A| + |B \setminus A|$$

Since $B = (A \cap B) \cup (B \setminus A)$ (disjoint union), we have:

$$|B| = |A \cap B| + |B \setminus A| \implies |B \setminus A| = |B| - |A \cap B|$$

Substituting this back into the first equation yields $|A \cup B| = |A| + |B| - |A \cap B|$. ▲

I.B. Generalized Inclusion-Exclusion

Theorem 0.2 (Principle of Inclusion-Exclusion (PIE) - General Form)**.** Let $A_1, A_2, \ldots, A_n$ be finite sets. The cardinality of their union is:

$$\left|\bigcup_{i=1}^{n} A_i\right| = \sum_{k=1}^{n} (-1)^{k-1} S_k$$

where $S_k = \sum_{1 \le i_1 < \cdots < i_k \le n} |A_{i_1} \cap \cdots \cap A_{i_k}|$ (the sum of the cardinalities of all $\binom{n}{k}$ intersections of k distinct sets).

Proof (Detailed).[Detailed Proof of General PIE] Consider an element x. If x belongs to exactly $m \ge 1$ sets, we must show that x is counted exactly once in the RHS sum. x is counted in S_k if the intersection $A_{i_1} \cap \cdots \cap A_{i_k}$ contains x. Since x is in m sets, there are $\binom{m}{k}$ such k-intersections. The net contribution of x to the RHS is:

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

Since $m \geq 1$, $(1-1)^m = 0$, so the contribution is 1. If $m = 0$, the sum is 0. Thus, every element in the union is counted exactly once. ▲

Theorem 0.3 (Generalized PIE: Exactly k Elements (E_k))**.** Let E_k be the number of elements belonging to **exactly** k of the n sets $A_1, \ldots, A_n$.

$$E_k = \sum_{j=k}^{n}(-1)^{j-k}\binom{j}{k}S_j$$

Proof (Detailed).[Proof of E_k Formula] Consider an element x that belongs to exactly m sets.

(a) If $m < k$, x is not in E_k, and since $\binom{j}{k} = 0$ for $j < k$, the sum is 0.

(b) If $m \geq k$, x is counted 1 time in E_k. x contributes to S_j exactly $\binom{m}{j}$ times. The total count for x on the RHS is:

$$\sum_{j=k}^{m}(-1)^{j-k}\binom{j}{k}\binom{m}{j}$$

Using the identity $\binom{j}{k}\binom{m}{j} = \binom{m}{k}\binom{m-k}{j-k}$ and letting $i = j-k$:

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

By the Binomial Theorem, $\sum_{i=0}^{N}(-1)^i\binom{N}{i} = (1-1)^N = 0$ for $N > 0$. Thus, the total count is $\binom{m}{k}(1-1)^{m-k}$, which is 1 if $m = k$ and 0 if $m > k$. This correctly equals E_k's contribution.

▲

II. Advanced Combinatorics and Counting

II.A. Permutations and Combinations

Formula 0.1 (k-Permutations (Ordered, No Repetition))**.**

$$P(n,k) = \frac{n!}{(n-k)!}$$

Formula 0.2 (k-Combinations (Unordered, No Repetition))**.**

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

Formula 0.3 (Combinations with Repetition (Multisets))**.** The number of ways to choose k items from n categories with repetition allowed:

$$\text{Multiset}\binom{n}{k} = \binom{n+k-1}{k}$$

Principle 0.1 (Stars and Bars)**.** The number of non-negative integer solutions to $x_1 + x_2 + \cdots + x_n = k$ is $\binom{n+k-1}{k}$. This is equivalent to arranging k "stars" and $n-1$ "bars."

II.B. Advanced Counting Identities

Formula 0.4 (Derangements)**.** The number of permutations π of n objects such that $\pi(i) \neq i$ for all i is the number of derangements, $!n$ (or D_n):

$$!n = n! \sum_{k=0}^{n} \frac{(-1)^k}{k!} = \left\lfloor \frac{n!}{e} + \frac{1}{2} \right\rfloor$$

Formula 0.5 (Vandermonde's Identity)**.**

$$\binom{r+s}{n} = \sum_{k=0}^{n} \binom{r}{k}\binom{s}{n-k}$$

Formula 0.6 (Hockey-Stick Identity)**.**

$$\sum_{i=r}^{n} \binom{i}{r} = \binom{n+1}{r+1}$$

II.C. Stirling Numbers

Definition 0.2 (Stirling Numbers of the Second Kind, $\left\{ {n \atop k} \right\}$)**.** The number of ways to partition a set of n distinct objects into k non-empty, indistinguishable subsets.

Formula 0.7 (Stirling Numbers of the Second Kind (Explicit Formula))**.**

$$\left\{ {n \atop k} \right\} = \frac{1}{k!} \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} j^n$$

Formula 0.8 (Stirling Numbers of the Second Kind (Recurrence))**.**

$$\left\{ {n \atop k} \right\} = \left\{ {n-1 \atop k-1} \right\} + k \left\{ {n-1 \atop k} \right\}$$

Definition 0.3 (Stirling Numbers of the First Kind (Absolute), ${n \atop k}$)**.** The number of permutations of n distinct objects that have exactly k disjoint cycles.

Formula 0.9 (Stirling Numbers of the First Kind (Recurrence))**.**

$${n \atop k} = {n-1 \atop k-1} + (n-1) {n-1 \atop k}$$

III. Axiomatic Probability and Total Probability

III.A. Axiomatic Foundations

Axiom 0.1 (Kolmogorov's Axioms). A probability measure $\mathbb{P}$ on a sample space Ω with σ-algebra $\mathcal{F}$ satisfies:

(i) **Non-negativity:** $\mathbb{P}(A) \geq 0$ for all $A \in \mathcal{F}$.

(ii) **Unit Measure:** $\mathbb{P}(\Omega) = 1$.

(iii) **Countable Additivity:** If $A_1, A_2, \ldots$ are pairwise disjoint events, then $\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

Property 0.1 (Derived Properties). (i) $\mathbb{P}(\emptyset) = 0$.

(ii) $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.

(iii) **Monotonicity:** If $A \subseteq B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$.

III.B. Conditional Probability and Independence

Definition 0.4 (Conditional Probability). For events A and B where $\mathbb{P}(B) > 0$:

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

Definition 0.5 (Independence). Events A and B are independent if and only if:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B)$$

Formula 0.10 (Multiplication Rule (Chain Rule)).

$$\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) = \mathbb{P}(A_1)\,\mathbb{P}(A_2 \mid A_1)\,\mathbb{P}(A_3 \mid A_1 \cap A_2) \cdots \mathbb{P}\left(A_n \mid \bigcap_{i=1}^{n-1} A_i\right)$$

Theorem 0.4 (Law of Total Probability (LTP) - Generalized). If $B_1, B_2, \ldots, B_n$ form a partition of the sample space Ω (disjoint, union is Ω, and $\mathbb{P}(B_i) > 0$), then for any event A:

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A \mid B_i)\,\mathbb{P}(B_i)$$

IV. Bayes' Theorem and Generalized Formulae

Theorem 0.5 (Bayes' Theorem). For events A and B (with $\mathbb{P}(A) > 0, \mathbb{P}(B) > 0$):

$$\mathbb{P}(B \mid A) = \frac{\mathbb{P}(A \mid B)\,\mathbb{P}(B)}{\mathbb{P}(A)}$$

Theorem 0.6 (Generalized Bayes' Theorem for N Events (Partition))**.** If $B_1, B_2, \ldots, B_n$ form a partition of the sample space Ω, then for any event A (with $\mathbb{P}(A) > 0$) and any $j \in \{1, \ldots, n\}$:

$$\mathbb{P}(B_j \mid A) = \frac{\mathbb{P}(A \mid B_j)\,\mathbb{P}(B_j)}{\sum_{i=1}^{n} \mathbb{P}(A \mid B_i)\,\mathbb{P}(B_i)}$$

Theorem 0.7 (Bayes' Theorem for 3 Events)**.** Specifically, if B_1, B_2, B_3 partition Ω:

$$\mathbb{P}(B_1 \mid A) = \frac{\mathbb{P}(A \mid B_1)\,\mathbb{P}(B_1)}{\mathbb{P}(A \mid B_1)\,\mathbb{P}(B_1) + \mathbb{P}(A \mid B_2)\,\mathbb{P}(B_2) + \mathbb{P}(A \mid B_3)\,\mathbb{P}(B_3)}$$

Proof (Detailed).[Proof of Generalized Bayes' Theorem] We start with the definition of conditional probability:

$$\mathbb{P}(B_j \mid A) = \frac{\mathbb{P}(A \cap B_j)}{\mathbb{P}(A)}$$

The numerator, $\mathbb{P}(A \cap B_j)$, can be rewritten using the product rule:

$$\mathbb{P}(A \cap B_j) = \mathbb{P}(A \mid B_j)\,\mathbb{P}(B_j) \quad \text{(Numerator)}$$

The denominator, $\mathbb{P}(A)$, can be rewritten using the Law of Total Probability over the partition $\{B_i\}_{i=1}^{n}$:

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A \cap B_i) = \sum_{i=1}^{n} \mathbb{P}(A \mid B_i)\,\mathbb{P}(B_i) \quad \text{(Denominator)}$$

Substituting the expressions for the numerator and denominator yields the generalized formula.
▲

V. Probability Inequalities

V.A. Bounds on Probability

Inequality 0.1 (Boole's Inequality (Union Bound))**.** For any sequence of events $A_1, A_2, \ldots$:

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

Proof (Detailed).[Proof of Boole's Inequality] Define a sequence of disjoint events B_i such that $\bigcup A_i = \bigcup B_i$: Let $B_1 = A_1$, and for $i \geq 2$, $B_i = A_i \setminus (\bigcup_{j=1}^{i-1} A_j) = A_i \cap \left(\cap_{j=1}^{i-1} A_j{}^c\right)$. Since the B_i are disjoint and $B_i \subseteq A_i$, we apply Countable Additivity (Axiom iii) and Monotonicity:

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

Since $B_i \subseteq A_i$, $\mathbb{P}(B_i) \leq \mathbb{P}(A_i)$.

$$\sum_{i=1}^{\infty} \mathbb{P}(B_i) \leq \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

▲

Inequality 0.2 (Bonferroni Inequalities)**.** Let $S_k = \sum_{1 \le i_1 < \cdots < i_k \le n} \mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_k})$ be the k-th term sum from the probabilistic PIE formula. The probability of the union is bounded by the partial sums:

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{k=1}^{n} (-1)^{k-1} S_k$$

(i) **Lower Bound (m odd):** $\mathbb{P}(\bigcup_{i=1}^{n} A_i) \ge \sum_{k=1}^{m} (-1)^{k-1} S_k$

(ii) **Upper Bound (m even):** $\mathbb{P}(\bigcup_{i=1}^{n} A_i) \le \sum_{k=1}^{m} (-1)^{k-1} S_k$

Proof (Detailed).[Proof of Bonferroni Inequalities] The inequalities arise from truncating the alternating series for the PIE. The error term $R_m = \mathbb{P}(\bigcup A_i) - \sum_{k=1}^{m} (-1)^{k-1} S_k$ is equal to:

$$R_m = \sum_{k=m+1}^{n} (-1)^{k-1} S_k = (-1)^m \sum_{k=m+1}^{n} (-1)^{k-(m+1)} S_k$$

By the generalized PIE for E_{m+1} (the probability that *exactly* $m+1$ events occur), one can show that the remaining sum $\sum_{k=m+1}^{n} (-1)^{k-(m+1)} S_k$ has the same sign as the first term S_{m+1}, and in fact, is equivalent to the probability that at least $m+1$ events occur (which is non-negative).

$$R_m = (-1)^m \mathbb{P}(\text{at least } m+1 \text{ events occur})$$

(a) If m is odd, $R_m \le 0$, hence $\mathbb{P}(\bigcup A_i) \le \sum_{k=1}^{m} (-1)^{k-1} S_k$. (Wait, this is $R_m = \mathbb{P}(\bigcup A_i) -$ Lower Bound, so $\mathbb{P}(\bigcup A_i) \le$ Lower Bound is wrong. Let's redefine the signs.) The error term should be defined in terms of $E_{\ge m+1}$, the probability that at least $m+1$ events occur. Let $P(k)$ be the probability that exactly k events occur. $P(k) = \sum_{j=k}^{n} (-1)^{j-k} \binom{j}{k} S_j$. $\mathbb{P}(\bigcup A_i) = \sum_{k=1}^{n} P(k)$. The remainder $R_m = \mathbb{P}(\bigcup A_i) - \sum_{k=1}^{m} (-1)^{k-1} S_k$. It can be shown that $(-1)^{m+1} R_m \ge 0$.

(b) If m is odd ($m+1$ is even), $R_m \ge 0$, so $\mathbb{P}(\bigcup A_i) \ge \sum_{k=1}^{m} (-1)^{k-1} S_k$.

(c) If m is even ($m+1$ is odd), $R_m \le 0$, so $\mathbb{P}(\bigcup A_i) \le \sum_{k=1}^{m} (-1)^{k-1} S_k$.

▲

V.B. Moment Inequalities

Inequality 0.3 (Markov's Inequality)**.** If X is a non-negative random variable ($X \ge 0$) and $a > 0$, then:

$$\mathbb{P}(X \ge a) \le \frac{\mathbb{E}[X]}{a}$$

Proof (Detailed).[Proof of Markov's Inequality (Continuous Case)] Assume X has a probability density function $f(x)$. Since $X \ge 0$:

$$\mathbb{E}[X] = \int_{0}^{\infty} x f(x) dx$$

We split the integral at a:

$$\mathbb{E}[X] = \int_0^a xf(x)dx + \int_a^\infty xf(x)dx$$

Since $x \geq 0$, $\int_0^a xf(x)dx \geq 0$. Also, for $x \in [a, \infty)$, $x \geq a$.

$$\mathbb{E}[X] \geq \int_a^\infty xf(x)dx \geq \int_a^\infty af(x)dx$$

Factoring out a:

$$\mathbb{E}[X] \geq a\int_a^\infty f(x)dx = a\mathbb{P}(X \geq a)$$

Dividing by a yields the result. ▲

Inequality 0.4 (Chebyshev's Inequality)**.** For any random variable X with finite mean $\mu = \mathbb{E}[X]$ and finite variance $\sigma^2 = \mathrm{Var}(X)$, and for any $\epsilon > 0$:

$$\mathbb{P}(|X - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}$$

Proof (Detailed).[Proof of Chebyshev's Inequality] Define the non-negative random variable $Y = (X - \mu)^2$. Then $\mathbb{E}[Y] = \mathbb{E}[(X - \mu)^2] = \mathrm{Var}(X) = \sigma^2$. The event $|X - \mu| \geq \epsilon$ is equivalent to $(X - \mu)^2 \geq \epsilon^2$. We apply Markov's inequality to Y with $a = \epsilon^2$:

$$\mathbb{P}(Y \geq a) \leq \frac{\mathbb{E}[Y]}{a}$$

Substituting the terms:

$$\mathbb{P}\left((X - \mu)^2 \geq \epsilon^2\right) \leq \frac{\mathbb{E}[(X - \mu)^2]}{\epsilon^2} = \frac{\sigma^2}{\epsilon^2}$$

Since $\mathbb{P}((X - \mu)^2 \geq \epsilon^2) = \mathbb{P}(|X - \mu| \geq \epsilon)$, the inequality holds. ▲