

Cheat Sheet

The Sum Rule: If an experiment can either end up being one of N outcomes, or one of M outcomes (where there is no overlap), then the total number of possible outcomes is: $N + M$.

Complementary Counting: Let \mathcal{U} be a (finite) universal set, and S a subset of interest. Then, $|S| = |\mathcal{U}| - |\mathcal{U} \setminus S|$.

k -Permutations: If we want to *pick (order matters)* only k out of n distinct objects, the number of ways to do so is:

$$P(n, k) = n \cdot (n - 1) \cdot (n - 2) \cdot \dots \cdot (n - k + 1) = \frac{n!}{(n-k)!}$$

k -Combinations/Binomial Coefficients: If we want to *choose (order doesn't matter)* only k out of n distinct objects, the number of ways to do so is:

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

The Product Rule: If an experiment has N_1 outcomes for the first stage, N_2 outcomes for the second stage, \dots , and N_m outcomes for the m^{th} stage, then the total number of outcomes of the experiment is $N_1 \times N_2 \times \dots \times N_m = \prod_{i=1}^m N_i$.

Permutation: The number of orderings of N **distinct** objects is $N! = N \cdot (N - 1) \cdot (N - 2) \cdot \dots \cdot 3 \cdot 2 \cdot 1$.

Multinomial Coefficients: If we have k distinct types of objects (n total), with n_1 of the first type, n_2 of the second, \dots , and n_k of the k -th, then the number of arrangements possible is

$$\binom{n}{n_1, n_2, \dots, n_k} = \frac{n!}{n_1! n_2! \dots n_k!}$$

Encoding/Stars and Bars Method: The number of ways to distribute n indistinguishable balls into k distinguishable bins is

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

Binomial Theorem: Let $x, y \in \mathbb{R}$ and $n \in \mathbb{N}$ a positive integer. Then: $(x+y)^n = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k}$.

Principle of Inclusion-Exclusion (PIE):

2 events: $|A \cup B| = |A| + |B| - |A \cap B|$

3 events: $|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|$

k events: singles - doubles + triples - quads + ...

Pigeonhole Principle: If there are n pigeons we want to put into k holes (where $n > k$), then at least one pigeonhole must contain at least 2 (or to be precise, $\lceil n/k \rceil$) pigeons.

Combinatorial Proofs: To prove two quantities are equal, you can come up with a combinatorial situation, and show that both in fact count the same thing, and hence must be equal.

Key Probability Definitions: The **sample space** is the set Ω of all possible outcomes of an experiment. An **event** is any subset $E \subseteq \Omega$. Events E and F are **mutually exclusive** if $E \cap F = \emptyset$.

Probability space: A *probability space* is a pair (Ω, \mathbb{P}) , where Ω is the sample space and $\mathbb{P} : \Omega \rightarrow [0, 1]$ is a *probability measure* such that $\sum_{x \in \Omega} \mathbb{P}(x) = 1$. The probability of an event $E \subseteq \Omega$ is $\mathbb{P}(E) = \sum_{x \in E} \mathbb{P}(x)$.

Equally Likely Outcomes: If Ω is a sample space such that each of the unique outcome elements in Ω are equally likely, then for any event $E \subseteq \Omega$: $\mathbb{P}(E) = |E|/|\Omega|$.

Conditional Probability: $\mathbb{P}(A | B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$

Bayes Theorem: $\mathbb{P}(A | B) = \frac{\mathbb{P}(B | A) \mathbb{P}(A)}{\mathbb{P}(B)}$

Partition: Non-empty events E_1, \dots, E_n **partition** the sample space Ω if they are both:

- **(Exhaustive)** $E_1 \cup E_2 \cup \dots \cup E_n = \bigcup_{i=1}^n E_i = \Omega$ (they cover the entire sample space).
- **(Pairwise Mutually Exclusive)** For all $i \neq j$, $E_i \cap E_j = \emptyset$ (none of them overlap)

Note that for any event E , E and E^C always form a partition of Ω .

Law of Total Probability (LTP): If events E_1, \dots, E_n partition Ω , then for any event F :

$$\mathbb{P}(F) = \sum_{i=1}^n \mathbb{P}(F \cap E_i) = \sum_{i=1}^n \mathbb{P}(F | E_i) \mathbb{P}(E_i)$$

Bayes Theorem with LTP: : Suppose A_1, \dots, A_n partition Ω and let B be any event. Then

$$\mathbb{P}(A_1 | B) = \frac{\mathbb{P}(B | A_1) \mathbb{P}(A_1)}{\sum_{i=1}^n \mathbb{P}(B | A_i) \mathbb{P}(A_i)}. \text{ In particular, } \mathbb{P}(A | B) = \frac{\mathbb{P}(B | A) \mathbb{P}(A)}{\mathbb{P}(B | A) \mathbb{P}(A) + \mathbb{P}(B | A^C) \mathbb{P}(A^C)}$$

Chain Rule: Let A_1, \dots, A_n be events with nonzero probabilities. Then:

$$\mathbb{P}(A_1 \cap \dots \cap A_n) = \mathbb{P}(A_1) \mathbb{P}(A_2 | A_1) \mathbb{P}(A_3 | A_1 \cap A_2) \dots \mathbb{P}(A_n | A_1 \cap \dots \cap A_{n-1})$$

Independence: A and B are **independent** if any of the following equivalent statements hold:
 1. $\mathbb{P}(A \cap B) = \mathbb{P}(A) \mathbb{P}(B)$ 2. $\mathbb{P}(A | B) = \mathbb{P}(A)$ 3. $\mathbb{P}(B | A) = \mathbb{P}(B)$

Mutual Independence: We say n events A_1, A_2, \dots, A_n are **(mutually) independent** if, for any subset $I \subseteq [n] = \{1, 2, \dots, n\}$, we have

$$\mathbb{P}\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} \mathbb{P}(A_i)$$

 This equation is actually representing 2^n equations since there are 2^n subsets of $[n]$.

Conditional Independence: A and B are **conditionally independent given an event C** if any of the following equivalent statements hold:
 1. $\mathbb{P}(A \cap B | C) = \mathbb{P}(A | C) \mathbb{P}(B | C)$ 2. $\mathbb{P}(A | B \cap C) = \mathbb{P}(A | C)$ 3. $\mathbb{P}(B | A \cap C) = \mathbb{P}(B | C)$

Random Variable (RV): A random variable (RV) X is a numeric function of the outcome $X : \Omega \rightarrow \mathbb{R}$.
 The set of possible values X can take on is its **range/support**, denoted Ω_X .
 If Ω_X is finite or countable infinite (typically integers or a subset), X is a **discrete RV**.

Probability Mass Function (PMF): For a discrete RV X , assigns probabilities to values in its range. That is $p_X : \Omega_X \rightarrow [0, 1]$ such that: (1) $p_X(k) \geq 0$ for all $k \in \Omega_X$; (2) $\sum_{k \in \Omega_X} p_X(k) = 1$.
 Furthermore, $p_X(k) = \mathbb{P}(X = k)$.

Cumulative Distribution Function (CDF): The **cumulative distribution function (CDF)** of ANY random variable (discrete or continuous) is defined to be the function $F_X : \mathbb{R} \rightarrow \mathbb{R}$ with $F_X(t) = \mathbb{P}(X \leq t)$.

Independence of RVs (Discrete): Discrete RVs X, Y are **independent**, written $X \perp Y$, if for all $x \in \Omega_X$ and $y \in \Omega_Y$:

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x) \cdot \mathbb{P}(Y = y)$$

Expectation (Discrete): The **expectation** of a discrete RV X is:

$$\mathbb{E}[X] = \sum_{k \in \Omega_X} k \cdot p_X(k)$$

Law of the Unconscious Statistician (LOTUS): For a RV X and function g : If X is discrete, $\mathbb{E}[g(X)] = \sum_{b \in \Omega_X} g(b) \cdot p_X(b)$.

Linearity of Expectation (LoE): For any random variables X, Y (possibly dependent):

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

Multiplicativity of expectation: For any independent random variables X, Y :

$$\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$$

Linearity of Expectation with Indicators: If asked only about the expectation of a RV X which is some sort of "count" (and not its PMF), then you may be able to write X as the sum of possibly dependent **indicator** RVs X_1, \dots, X_n , and apply LoE, where for an indicator RV X_i , $\mathbb{E}[X_i] = 1 \cdot \mathbb{P}(X_i = 1) + 0 \cdot \mathbb{P}(X_i = 0) = \mathbb{P}(X_i = 1)$.

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

Standard Deviation (SD):
$$\sigma_X = \sqrt{\text{Var}(X)}$$

Property of Variance:
$$\text{Var}(aX + b) = a^2 \text{Var}(X)$$

Variance Adds for Independent RVs: If X, Y are independent, then
$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$

Bernoulli/Indicator Random Variable: $X \sim \text{Bernoulli}(p)$
(Ber(p) for short) iff X has PMF:

$$p_X(k) = \begin{cases} p, & k = 1 \\ 1 - p, & k = 0 \end{cases}$$

$\mathbb{E}[X] = p$ and $\text{Var}(X) = p(1-p)$. A Bernoulli r.v. is intended to capture whether an event occurred or not. An example of a Bernoulli/indicator RV is one flip of a coin with $\mathbb{P}(\text{head}) = p$.

Binomial Random Variable: $X \sim \text{Binomial}(n, p)$ (Bin(n, p) for short) iff X has PMF

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k \in \Omega_X = \{0, 1, \dots, n\}$$

$\mathbb{E}[X] = np$ and $\text{Var}(X) = np(1-p)$. X is the sum of n i.i.d. Ber(p) random variables and models the number of successes/events that occur in the n independent Bernoulli trials.

Poisson Random Variable: $X \sim \text{Poisson}(\lambda)$ (Poi(λ) for short) iff X has PMF:

$$p_X(k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k \in \Omega_X = \{0, 1, 2, \dots\}$$

$\mathbb{E}[X] = \lambda$ and $\text{Var}(X) = \lambda$. A Poisson r.v. counts the number of events that occur in one unit of time when events happen at a rate of λ per unit of time.

If X_1, \dots, X_n are independent Poisson RV's, where $X_i \sim \text{Poi}(\lambda_i)$, then $X = X_1 + \dots + X_n \sim \text{Poi}(\lambda_1 + \dots + \lambda_n)$.

A Poisson random variable can be used to approximate a Binomial random variable if n is large and p is small, in which case we set $\lambda = np$. (The approximation improves as $n \rightarrow \infty$ and $p \rightarrow 0$, with np a constant).

Uniform Random Variable (Discrete): $X \sim \text{Uniform}(a, b)$
(Unif(a, b) for short), for integers $a \leq b$, iff X has PMF:

$$p_X(k) = \frac{1}{b-a+1}, \quad k \in \Omega_X = \{a, a+1, \dots, b\}$$

$\mathbb{E}[X] = \frac{a+b}{2}$ and $\text{Var}(X) = \frac{(b-a)(b-a+1)}{12}$. A Uniform r.v. captures situations in which each of the outcomes $\{a, a+1, \dots, b\}$ is equally likely.

Geometric Random Variable: $X \sim \text{Geometric}(p)$ (Geo(p) for short) iff X has PMF:

$$p_X(k) = (1-p)^{k-1} p, \quad k \in \Omega_X = \{1, 2, 3, \dots\}$$

$\mathbb{E}[X] = \frac{1}{p}$ and $\text{Var}(X) = \frac{1-p}{p^2}$. A Geometric r.v. counts the number of independent Ber(p) trials until an event/success happens (including the final event/success).