Final Reference Sheet

Counting

Sum Rule
If you are choosing one thing between n options in one group and m in another group with no overlap, the total number of options is: $n + m$.

Product Rule
If you have a sequential process, where step 1 has n_1 options, step 2 has n_2 options,...,step k has n_k options, and you choose one from each step, the total number of possibilities is $n_1 \cdot n_2 \cdots n_k$

n Factorial
$n! = n \cdot (n-1) \cdot (n-2) \cdots 1$
We only define $n!$ for natural numbers n. As a convention, we define: $0! = 1$

k-permutation
The number of k-element sequences of distinct symbols from a universe of n symbols is:

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

Edge cases: $P(n,n) = n!$, $P(n,0) = 1$, $P(n,k)$ for $k < 0$ or $k > n$ is undefined.

k-combination
The number of k-element subsets from a set of n symbols is:

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

Edge cases: $\binom{n}{0} = \binom{n}{n} = 1$; $P(n,k)$ for $k < 0$ or $k > n$ is undefined.

Combination Facts
Symmetry of combinations:

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

Pascal's Rule:

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

Binomial Theorem
Let $x, y \in \mathbb{R}$ and $n \in \mathbb{N}$ positive integer. Then:

$$(x+y)^n = \sum_{i=0}^{n} \binom{n}{i} x^i y^{n-1}$$

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 $\lceil \frac{n}{k} \rceil$ pigeons.

Stars and Bars
To pick n objects from k groups (where order doesn't matter and every element of each group is indistinguishable), use the formula:

$$\binom{n+(k-1)}{k-1}$$

Probability

Sample Space
The sample space is the set Ω of all possible outcomes of an experiment.

Event
An event $E \subseteq \Omega$ is a subset of possible outcomes (i.e. a subset of Ω).

Mutually Exclusive Events
Two events E, F are mutually exclusive if they cannot happen simultaneously (they are disjoin subsets of the sample space). In notation,

$$E \cap F = \varnothing$$

Probability Space
A (discrete) probability space is a pair $(\Omega, \mathbb{P})$, where Ω is the sample space and $\mathbb{P} : \Omega \to [0,1]$ is the probability measure such that:

- $\mathbb{P}(x) \geqslant 0$ for all x
- $\sum_{\omega \in \Omega} \mathbb{P}(\omega) = 1$
- If $E, F \subseteq \Omega$ and $E \cap F = \varnothing$ then $\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F)$

Probability Consequences (Corollaries)
Complementation: $\mathbb{P}(\bar{E}) = 1 - \mathbb{P}(E)$
Monotonicity: If $E \subseteq F$, then $\mathbb{P}(E) \leqslant \mathbb{P}(F)$
Inclusion-exclusion: $\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(E \cap F)$

Uniform Probability Space
If every outcome in your sample space is equally likely then,

$$\mathbb{P}(E) = \frac{|E|}{|\Omega|}$$

Conditional Probability
For an event B, with $\mathbb{P}(B) > 0$, the "Probability of A conditioned on B" is

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

$\mathbb{P}(A|B)$ is undefined when $\mathbb{P}(B) = 0$.

Bayes' Rule
$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}$

Partition
Non-empty events $A_1, \ldots, A_n$ partition the sample space Ω if:

- **Exhaustive:** $A_1 \cup A_2 \cup \cdots \cup A_n = \Omega$ (they cover the entire sample space).
- **Pairwise Mutually Exclusive:** For all $i \neq j$, $A_i \cap A_j = \varnothing$ (none of them overlap)

Law of Total Probability (LTP)
Let $A_1, A_2 \ldots, A_n$ partition Ω.
For any event E:

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

Independence
Two events A, B are independent if

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

If A, B both have non-zero probabilities, then they are independent if any of the following equivalent statements hold:

1. $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$
2. $\mathbb{P}(A|B) = \mathbb{P}(A)$
3. $\mathbb{P}(B|A) = \mathbb{P}(B)$

Chain Rule
$\mathbb{P}(A_1 \cap A_2 \cap \ldots \cap A_n) =$
$\mathbb{P}(A_1) \cdot \mathbb{P}(A_2|A_1) \cdot \mathbb{P}(A_3|A_1 \cap A_2) \cdots \mathbb{P}(A_n|A_1 \cap \ldots \cap A_{n-1})$

Conditional Independence
We say A and B are conditionally independent on C

$$\mathbb{P}(A \cap B|C) = \mathbb{P}(A|C) \cdot \mathbb{P}(B|C)$$

Pairwise Independence
Events $A_1, A_2, ..., A_n$ are pairwise independent if for all i, j

$$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i) \cdot \mathbb{P}(A_j)$$

Mutual Independence
Events $A_1, A_2, ..., A_n$ are mutually independent if

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \cdot \mathbb{P}(A_{i_2}) \cdots \mathbb{P}(A_{i_k})$$

For every subset $\{i_1, i_2, ..., i_k\}$ of $\{1, 2, ..., n\}$.

Random Variables (Discrete and Continuous)

Random Variable (RV)
A random variable (RV) X is a numeric function of the outcome $X : \Omega \to \mathbb{R}$. $X(\omega)$ is the summary of the outcome ω. The set of possible values X can take on is its range/support, denoted Ω_X.

Probability Mass Function (PMF) (Discrete)
For a discrete RV X, it assigns probabilities to values in its range. That is $p_X : \Omega_X \to [0,1]$, where $p_X(k) = P(X = k)$.

Expectation (Discrete)
The "expectation" (or "expected value") of a random variable X is:

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

LOTUS - Expectation of A Function (Discrete)
For a RV X and function g:

$$\mathbb{E}\left[g(X)\right] = \sum_{b \in \Omega_X} g(b) \cdot p_X(b)$$

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

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

Variance

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

Standard Deviation (SD)

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

Property of Variance

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

Probability Density Function (PDF) (Continuous)
The PDF of a continuous RV X is the function $f_X : \mathbb{R} \to \mathbb{R}$, such that

1. $f_X(z) \geqslant 0$ for all $z \in \mathbb{R}$
2. $\int_{-\infty}^{\infty} f_X(t)\, dt = 1$

Furthermore, $\mathbb{P}(a \leqslant X \leqslant b) = \int_a^b f_X(w)\ \mathrm{d}w$

Cumulative Distribution Function (CDF)
The CDF of ANY random variable (discrete or continuous) is $F_X(t) = \mathbb{P}(X \leqslant t)$.
If X is *continuous*, $F_X(t) = \mathbb{P}(X \leqslant t) = \int_{-\infty}^{t} f_X(w)\, dw$ for all $t \in \mathbb{R}$. Further, $\frac{d}{du}F_X(u) = f_X(u)$

Expectation (Continuous)

$$\mathbb{E}\left[X\right] = \int_{-\infty}^{\infty} x\ f_X\left(x\right) \mathrm{d}x$$

LOTUS - Expectation of A Function (Continuous)
For a RV X and function g:

$$\mathbb{E}\left[g(X)\right] = \int_{-\infty}^{\infty} g(x) f_X\left(x\right) \mathrm{d}x$$

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

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

Independence (Random Variables)
Random variables X and Y are independent if for all $x \in \Omega_X$ and all $y \in \Omega_Y$:

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

Variance Adds for Independent RVs
If X, Y are independent, then

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$

Zoo of Random Variables

Bernoulli/Indicator Random Variable
$X \sim$ Ber(p) iff X has PMF:

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

$\mathbb{E}[X] = p$ and $\mathrm{Var}(X) = p(1-p)$.

Binomial Random Variable
$X \sim$ Bin(n,p) iff X has PMF

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

$\mathbb{E}[X] = np$ and $\mathrm{Var}(X) = np(1-p)$.
X is the number of successes in n independent trials,

Uniform Random Variable (Discrete)
$X \sim$ Unif(a,b), for integers $a \leqslant b$, iff X has PMF:

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

$\mathbb{E}[X] = \frac{a+b}{2}$ and $\mathrm{Var}(X) = \frac{(b-a)(b-a+2)}{12}$
X is an integer between a and b uniformly at random.

Geometric Random Variable
$X \sim$ Geo(p) iff X has PMF:

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

$\mathbb{E}[X] = \frac{1}{p}$ and $\mathrm{Var}(X) = \frac{1-p}{p^2}$ and $F_X(k) = 1 - (1-p)^k$
X is the number of independent trails until the first success.

Hypergeometric Random Variable
$X \sim$ HypGeo(N, K, n) iff X has PMF:

$$p_X(k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}},$$

$$k \in \Omega_X = \{\max\{0, n+K-N\}, \ldots, \min\{K, n\}\}$$

$\mathbb{E}[X] = \frac{nK}{N}$ and $\mathrm{Var}(X) = n\frac{K(N-K)(N-n)}{N^2(N-1)}$.
X is the number of success in a sample of size n from a set of N elements, K of which are successes

Negative Binomial Random Variable
$X \sim$ NegBin(r,p) iff X has PMF:

$$p_X(k) = \binom{k-1}{r-1}(1-p)^{k-r}p^r, k \in \Omega_X = \{r, r+t, ...\}$$

$\mathbb{E}[X] = \frac{r}{p}$ and $\mathrm{Var}(X) = n\frac{r(1-p)}{p^2}$.
X is the number of independent trials till the r'th success.

Poisson Random Variable
$X \sim$ Poi(λ) iff X has PMF:

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

$\mathbb{E}[X] = \lambda$ and $\mathrm{Var}(X) = \lambda$
X is the number of successes seen in a time interval where λ is the average number of successes in the time interval.

Uniform Random Variable (Continuous)
$X \sim$ Unif(a,b) iff X has PDF:

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{if } x \in \Omega_X = [a, b] \\ 0 & \text{otherwise} \end{cases}$$

$\mathbb{E}[X] = \frac{a+b}{2}$ $\mathrm{Var}(X) = \frac{(b-a)^2}{12}$ $F_X(k) = \frac{k-a}{b-a}$ if $a \leqslant k < b$
X is a real number uniformly at random between a and b

Exponential Random Variable
$X \sim$ Exp(λ) iff X has PDF:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \in \Omega_X = [0, \infty) \\ 0 & \text{otherwise} \end{cases}$$

$\mathbb{E}[X] = \frac{1}{\lambda}$ and $\mathrm{Var}(X) = \frac{1}{\lambda^2}$
$F_X(x) = 1 - e^{-\lambda x}$ for $x \geqslant 0$.
X is the time till the first success (λ is same as in Poisson)

Normal (Gaussian, "bell curve") Random Variable
$X \sim \mathcal{N}(\mu, \ \sigma^2)$ iff X has PDF:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}, \quad x \in \Omega_X = \mathbb{R}$$

$\mathbb{E}[X] = \mu$ and $\mathrm{Var}(X) = \sigma^2$. random variable is typically denoted Z and has mean 0 and variance 1.
Scale/shift: If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$. E.g., standard normal: $\frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$.
Addition: If $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$, $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ are *independent*,

$$aX + bY + c \sim \mathcal{N}(a\mu_X + b\mu_Y + c, a^2\sigma_X^2 + b^2\sigma_Y^2)$$

Central Limit Theorem

The Central Limit Theorem (CLT)
Let $X_1, X_2, \ldots, X_n$ be i.i.d. RVs with mean μ and variance σ^2. Let $Y_n = \frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$. As $n \to \infty$, the CDF of Y_n converges to the CDF of $\mathcal{N}(0,1)$

Joint Distributions

Joint PMFs The joint PMF of discrete RVs X and Y is:
$$p_{X,Y}(a,b) = \mathbb{P}(X=a, Y=b)$$
Their joint range is
$$\Omega_{X,Y} = \{(c,d) : p_{X,Y}(c,d) > 0\} \subseteq \Omega_X \times \Omega_Y$$
Note that $\sum_{(s,t)\in\Omega_{X,Y}} p_{X,Y}(s,t) = 1$.

Joint PDFs The joint PDF of continuous RVs X and Y is:
$$f_{X,Y}(a,b) \geqslant 0$$
Their joint range is
$$\Omega_{X,Y} = \{(c,d) : f_{X,Y}(c,d) > 0\} \subseteq \Omega_X \times \Omega_Y$$
Note that $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(u,v)\,\mathrm{d}u\,\mathrm{d}v = 1$.

Multi-dimensional LOTUS (Discrete) If $g : \mathbb{R}^2 \to \mathbb{R}$ is a function, then
$$\mathbb{E}\left[g(X,Y)\right] = \sum_{x\in\Omega_X}\sum_{y\in\Omega_Y} g(x,y)\ p_{X,Y}(x,y)$$

Multi-dimensional LOTUS (Continuous) If $g : \mathbb{R}^2 \to \mathbb{R}$ is a function, then
$$\mathbb{E}\left[g(X,Y)\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(s,t)\ f_{X,Y}(s,t)\ \mathrm{d}s\,\mathrm{d}t$$

Marginal PMFs Let X, Y be discrete random variables. The marginal PMF of X is:
$$p_X(a) = \sum_{b\in\Omega_Y} p_{X,Y}(a,b)$$

Marginal PDFs Let X, Y be continuous random variables. The marginal PDF of X is:
$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,\mathrm{d}y$$

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

Independence of RVs (Continuous) Continuous RVs X, Y are independent, written $X \perp Y$, if for all $x \in \Omega_X$ and $y \in \Omega_Y$,
$$f_{X,Y}(x,y) = f_X(x)f_Y(y)$$

Probabilities
$$\mathbb{P}(a \leqslant X < b, c \leqslant Y \leqslant d) = \sum_{x=a}^{b}\sum_{y=c}^{d} p_{X,Y}(x,y)$$

Probabilities
$$\mathbb{P}(a \leqslant X < b, c \leqslant Y \leqslant d) = \int_a^b\int_c^d f_{X,Y}(x,y)\,\mathrm{d}y\,\mathrm{d}x$$

Conditional PMF Let X, Y be discrete random variables, then
$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$$

Conditional PDF Let X, Y be continuous random variables, then
$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

Conditional Expectation Let X, Y be discrete random variables, then we define the conditional expectation of $g(X)$ given (the event that) $Y = y$ as:
$$\mathbb{E}[X|Y=y] = \sum_{x\in\Omega_x} x \cdot p_{X|Y}(x|y)$$

Conditional Expectation Let X, Y be continuous random variables, then we define the conditional expectation of $g(X)$ given (the event that) $Y = y$ as:
$$\mathbb{E}[X|Y=y] = \int_{-\infty}^{\infty} x \cdot f_{X|Y}(x|y)dx$$

Law of Total Expectation (LTE)
Let X, Y be jointly distributed discrete random variables.
$$\mathbb{E}\left[\,g(X)\,\right] = \sum_{y\in\Omega_Y}\mathbb{E}\left[\,g(X)\mid Y=y\,\right]\ p_Y(y)$$

Law of Total Expectation (LTE)
Let X, Y be jointly distributed continuous RVs.
$$\mathbb{E}\left[\,g(X)\,\right] = \int_{-\infty}^{\infty}\mathbb{E}\left[\,g(X)\mid Y=y\,\right]\ f_Y(y)\,\mathrm{d}y$$

Covariance and Variance
$$\begin{aligned}\mathsf{Cov}(X,Y) &= \mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])] \\ &= \mathbb{E}[XY] - \mathbb{E}[X][Y]\end{aligned}$$
$$\mathsf{Var}(X+Y) = \mathsf{Var}(X) + \mathsf{Var}(Y) + 2\mathsf{Cov}(X,Y)$$

Covariance Patterns

1. $\mathsf{Cov}(X,X) = \mathsf{Var}(X)$
2. $\mathsf{Cov}(X,-Y) = -\mathsf{Cov}(X,Y)$
3. if X, Y independent $\mathsf{Cov}(X,Y) = 0$

Tail Bounds

Markov's Inequality Let X be a random variable supported (only) on **non-negative** numbers. For any $t > 0$:

$$\mathbb{P}(X \geqslant t) \leqslant \frac{\mathbb{E}[X]}{t}$$

Chebyshev's Inequality Let X be a random variable. For any $t > 0$:

$$\mathbb{P}(|X - \mathbb{E}[X]| \geqslant t) \leqslant \frac{\operatorname{Var}(X)}{t^2}$$

Chernoff Bound Let $X = X_1 + X_2 + \ldots + X_n$, where $X_1, X_2, \ldots, X_n$ are independent Bernoulli random variables and $\mu = \mathbb{E}[X]$. For any $0 \leqslant \delta \leqslant 1$:

$$\mathbb{P}\left(X \geqslant (1+\delta)\mu\right) \leqslant \exp\left(-\frac{\delta^2 \mu}{3}\right)$$

$$\mathbb{P}\left(X \leqslant (1-\delta)\mu\right) \leqslant \exp\left(-\frac{\delta^2 \mu}{2}\right)$$

The Union Bound Let $E_1, E_2, ..., E_n$ be a collection of events. Then: $\mathbb{P}(\bigcup_{i=1}^n E_i) \leqslant \sum_{i=1}^n \mathbb{P}(E_i)$.

Maximum Likelihood Estimation (MLE)

Likelihood (Discrete)
Let $x_1, ..., x_n$ be i.i.d. samples from PMF $p_X(x;\theta)$ where θ is a parameter (or vector of parameters). The **likelihood** of θ given the samples is

$$\mathcal{L}(x_1, ..., x_n; \theta) = \prod_{i=1}^n p_X(x_i; \theta).$$

Maximum Likelihood Estimator (MLE)
Let $x_1, ..., x_n$ be i.i.d. samples from probability distribution with parameter (or vectors of parameters) θ. The **maximum likelihood estimator (MLE)** $\hat{\theta}_{MLE}$ for θ is the value of θ that maximizes the likelihood/log-likelihood: $\hat{\theta}_{MLE} = \arg\max_\theta \mathcal{L}(x_1, ..., x_n; \theta) = \arg\max_\theta \ln \mathcal{L}(x_1, ..., x_n; \theta)$.

Likelihood (Continuous)
Let $x_1, ..., x_n$ be i.i.d. samples from PDF $f_X(x;\theta)$ where θ is a parameter (or vector of parameters). The **likelihood** of θ given the samples is

$$\mathcal{L}(x_1, ..., x_n; \theta) = \prod_{i=1}^n f_X(x_i; \theta).$$

Unbiased Estimators
An estimator $\hat{\theta}$ of a parameter θ is **unbiased** if $\mathbb{E}[\hat{\theta}] = \theta$.

Consistent Estimators
An estimator $\hat{\theta}$ of a parameter θ is **consistent** if $\lim_{n\to\infty} \mathbb{E}[\hat{\theta}] = \theta$. The MLE is consistent (under some very mild assumptions).