

# 1 Combinatorics

**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:

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

**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.

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

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

**Multinomial Coefficients:** If we have  $k$  distinct types of objects ( $n$  total), with  $n_1$  of the first type,  $n_2$  of the second, ..., 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

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

**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.

## 2 Discrete Probability

### 2.1 Discrete Probability

**Key Probability Definitions:** The **sample space** is the set  $\Omega$  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  $(\Omega, \mathbb{P})$ , where  $\Omega$  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  $\Omega$  is a sample space such that each of the unique outcome elements in  $\Omega$  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  $\Omega$  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  $\Omega$ .

**Law of Total Probability (LTP):** If events  $E_1, \dots, E_n$  partition  $\Omega$ , 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  $\Omega$  and let  $B$  be any event. Then  $\mathbb{P}(A_1|B) = \frac{\mathbb{P}(B | A_1)\mathbb{P}(A_1)}{\sum_{i=1}^n \mathbb{P}(B | A_i)\mathbb{P}(A_i)}$ . 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.  $\mathbb{P}(B | A \cap C) = \mathbb{P}(B | C)$

## 2.2 Random Variables

**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  $\Omega_X$ .

If  $\Omega_X$  is finite or countable infinite (typically integers or a subset),  $X$  is a **discrete RV**, otherwise it is a **continuous random variable**.

**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  $F_X(t) = \mathbb{P}(X \leq t)$ . If  $X$  is a *continuous* RV,  $F_X(t) = \mathbb{P}(X \leq t) = \int_{-\infty}^t f_X(w) dw$ .

**Probability Density Function (PDF):** For a continuous RV  $X$ , the probability density function  $f_X$  satisfies (1)  $f_X(x) \geq 0$  for all real valued  $x$ , and (2)  $\int_{-\infty}^{\infty} f_X(x) dx = 1$ .

**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:  

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

**Expectation (Continuous):** The **expectation** of a continuous RV  $X$  is:  

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x) dx.$$

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

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

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

**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  $X_i$ ,  $\mathbb{E}[X_i] = 1 \cdot \mathbb{P}(X_i = 1) + 0 \cdot \mathbb{P}(X_i = 0) = \mathbb{P}(X_i = 1)$ .

## 2.3 Variance

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

## 2.4 Some notes on the zoo (see sheet with distributions)

**Notes on discrete random variables:**

1. A Geometric r.v. counts the number of independent  $\text{Ber}(p)$  trials until an event/success happens (including the final event/success). It is memoryless.
2. If  $X_1, \dots, X_n$  are independent Poisson RV's, where  $X_i \sim \text{Poi}(\lambda_i)$ , then  $X = X_1 + \dots + X_n \sim \text{Poi}(\lambda_1 + \dots + \lambda_n)$ .
3. 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).

**Notes on continuous distributions:**

1. The exponential distribution is memoryless.
2. If  $Z \sim \mathcal{N}(0, 1)$ , then  $Z$  is called a standard normal random variable and has CDF  $\Phi(z) = \mathbb{P}(Z \leq z)$
3. If  $X \sim \mathcal{N}(\mu, \sigma^2)$ , then  $aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$ .  
 In particular, we can always scale/shift to get the standard Normal:  $\frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1)$ .
4. If  $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$  and  $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$  are *independent*, then  

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

## 2.5 Central Limit Theorem (CLT)

**Independent and Identically Distributed (i.i.d.):** We say  $X_1, \dots, X_n$  are said to be **independent and identically distributed (i.i.d.)** if all the  $X_i$ 's are independent of each other, and have the same distribution (PMF for discrete RVs, or CDF for continuous RVs).

**CLT:** Let  $X_1, \dots, X_n$  be iid random variables with  $\mathbb{E}[X_i] = \mu$  and  $\text{Var}(X_i) = \sigma^2$ . Let  $X = \sum_{i=1}^n X_i$ . Let  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  be the *sample mean*. Then

In the limit as  $n \rightarrow \infty$ ,  $\bar{X}$  approaches the normal distribution  $\mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$ .

In the limit as  $n \rightarrow \infty$ ,  $Y' = \frac{X - n\mu}{\sigma\sqrt{n}}$  approaches  $\mathcal{N}(0, 1)$ .

For large  $n$ ,  $X$  can be well approximated by  $\mathcal{N}(n\mu, n\sigma^2)$ .

When applying CLT, don't forget the continuity correction when  $X_1, \dots, X_n$  are discrete random variables,

## 2.6 Multivariate Probability

	Discrete	Continuous
<b>Joint PMF/PDF</b> must be $\geq 0$	$p_{X,Y}(x, y) = \mathbb{P}(X = x, Y = y)$	$f_{X,Y}(x, y) \neq \mathbb{P}(X = x, Y = y)$
<b>Joint range/support</b> $\Omega_{X,Y}$	$\{(x, y) \in \Omega_X \times \Omega_Y : p_{X,Y}(x, y) > 0\}$	$\{(x, y) \in \Omega_X \times \Omega_Y : f_{X,Y}(x, y) > 0\}$
<b>Joint CDF</b>	$F_{X,Y}(x, y) = \sum_{t \leq x, s \leq y} p_{X,Y}(t, s)$	$F_{X,Y}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(t, s) ds dt$
<b>Normalization</b>	$\sum_{(x,y) \in \Omega_{X,Y}} p_{X,Y}(x, y) = 1$	$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx dy = 1$
<b>Marginal PMF/PDF</b>	$p_X(x) = \sum_{y \in \Omega_Y} p_{X,Y}(x, y)$	$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy$
<b>Expectation</b>	$\mathbb{E}[g(X, Y)] = \sum_{(x,y) \in \Omega_{X,Y}} g(x, y) p_{X,Y}(x, y)$	$\mathbb{E}[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{X,Y}(x, y) dx dy$
$\mathbb{P}(a \leq X < b, c \leq Y \leq d)$	$\sum_{a \leq t < b} \sum_{c \leq s \leq d} p_{X,Y}(t, s)$	$\int_a^b \int_c^d f_{X,Y}(x, y) dy dx$
<b>Independence</b> must have	$\forall x, y, p_{X,Y}(x, y) = p_X(x)p_Y(y)$ $\Omega_{X,Y} = \Omega_X \times \Omega_Y$	$\forall x, y, f_{X,Y}(x, y) = f_X(x)f_Y(y)$ $\Omega_{X,Y} = \Omega_X \times \Omega_Y$

In the discrete column, whenever we sum over pairs  $(t, s)$ , they are in  $\Omega_{X,Y}$ .

## 2.7 Law of Total Probability + Law of Total Expectation

**Law of Total Probability (r.v. version):** If  $X$  is a discrete random variable, then  $\mathbb{P}(A) = \sum_{x \in \Omega_X} \mathbb{P}(A|X = x)p_X(x)$  If  $X$  is continuous,  $\mathbb{P}(A) = \int_{-\infty}^{\infty} \mathbb{P}(A|X = x)f_X(x) dx$ .

**Conditional Expectation:** The expected value of random variable  $X$  given that event  $A$  has occurred, written  $\mathbb{E}[X|A]$ , is defined as

$$\mathbb{E}[X|A] = \sum_{x \in \Omega_X} x \cdot \mathbb{P}(X = x|A).$$

For continuous r.v.s  $X$  and  $Y$ ,

$$\mathbb{E}[X|Y = y] = \int_{-\infty}^{\infty} x \cdot \frac{f_{X,Y}(x,y)}{f_Y(y)} dx$$

If  $X$  is discrete (and  $Y$  is either discrete or continuous), then

$$\mathbb{E}[g(X) | Y = y] = \sum_{x \in \Omega_X} g(x) \mathbb{P}(X = x | Y = y)$$

**Law of Total Expectation (LTE):** If  $Y$  is discrete (and  $X$  is either discrete or continuous), then:

$$\mathbb{E}[X] = \sum_{y \in \Omega_Y} \mathbb{E}[X|Y = y] \cdot \mathbb{P}(Y = y).$$

If  $Y$  is continuous, then

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} \mathbb{E}[X|Y = y] f_Y(y) dy$$

**Covariance:** The Covariance between random variables  $X$  and  $Y$  is:

$$\begin{aligned} \text{Cov}(X, Y) &= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \end{aligned}$$

## Maximum likelihood estimation (MLE)

**Likelihood function:** Let  $X_1 = x_1, \dots, X_n = x_n$  be iid samples of a random variable with probability mass function  $p_X(x; \theta)$  (if  $X$  discrete) or density  $f_X(x; \theta)$  (if  $X$  continuous), where  $\theta$  is a parameter (or a vector of parameters). The likelihood function is the probability of seeing the data (in the discrete case).

If  $X$  is discrete:

$$L(X_1 = x_1, \dots, X_n = x_n; \theta) = \prod_{i=1}^n p_X(x_i; \theta)$$

If  $X$  is continuous:

$$L(X_1 = x_1, \dots, X_n = x_n; \theta) = \prod_{i=1}^n f_X(x_i; \theta)$$

**Maximum likelihood estimator:** We denote the MLE of  $\theta$  as  $\hat{\theta}_{\text{MLE}}$  or simply  $\hat{\theta}$ , the parameter (or vector of parameters) that maximizes the likelihood function or equivalently, maximizes the log-likelihood. (We find it using calculus.) For a fixed set of realizations/samples  $x_1, \dots, x_n$ , the MLE of  $\theta$  is

$$\begin{aligned} \hat{\theta}_{\text{MLE}}(x_1, \dots, x_n) &= \arg \max_{\theta} L(x_1, \dots, x_n; \theta) = \\ &= \arg \max_{\theta} \ln L(x_1, \dots, x_n; \theta) \end{aligned}$$

Viewing the estimator  $\hat{\theta}_{\text{MLE}}$  as a function of the random variables  $X_1, \dots, X_n$ , the estimator  $\hat{\theta}_{\text{MLE}}(X_1, \dots, X_n)$  is itself a random variable.

**Unbiased estimator:** An estimator  $\hat{\theta}$  of  $\theta$  is unbiased iff  $\mathbb{E}[\hat{\theta}(X_1, \dots, X_n)] = \theta$ .