Multinomial Distribution Notes

1. Introduction & Intuition

- **Multinomial** = "multi-trial" version of the Categorical distribution.
- K categories with probabilities $\boldsymbol{p} = (p_1, \ldots, p_K)^\top$ satisfying $\sum_i p_i = 1$.
- Conduct n *independent* trials and *count* how many times each category appears.
- Examples:
  - Roll a six-sided die n times $\rightarrow$ counts of each face.
  - n website visits $\rightarrow$ counts of pages viewed (home, about, blog, ...).

Define the count vector

$$\boldsymbol{X} = (X_1, \ldots, X_K)^\top, \qquad X_i \in \{0, 1, \ldots, n\}, \qquad \sum_{i=1}^{K} X_i = n.$$

2. Probability Mass Function (PMF)

$$P(\boldsymbol{X} = \boldsymbol{x}) \;=\; \frac{n!}{x_1!\, x_2! \ldots x_K!} \prod_{i=1}^{K} p_i^{x_i}, \qquad \text{where } \sum_i x_i = n.$$

Intuition.

1. *Product term* $\prod p_i^{x_i}$: probability of *one particular* ordering (independent trials).
2. *Coefficient* $\dfrac{n!}{x_1! \ldots x_K!}$: number of distinct reorderings of those n outcomes (shuffle n slots, divide by internal duplicates).

3. Worked Example ("Blue–Green–Red" urn)

Three colours, $\boldsymbol{p} = (0.5,\ 0.3,\ 0.2)$, four draws ($n = 4$). Suppose $\boldsymbol{x} = (2, 1, 1)$ (blue twice, green once, red once):

$$P(\boldsymbol{X} = \boldsymbol{x}) = \frac{4!}{2!\,1!\,1!}\, 0.5^2\, 0.3^1\, 0.2^1 = 12 \times 0.015 = 0.18.$$

4. Expected Value

Each category behaves like $\text{Binomial}(n, p_i)$ inside the vector:

$$\mathbb{E}[X_i] = np_i \quad \Longrightarrow \quad \mathbb{E}[\boldsymbol{X}] = n\boldsymbol{p}.$$

5. Covariance Matrix

$$\text{Cov}(\boldsymbol{X}) = n\big[\text{diag}(\boldsymbol{p}) - \boldsymbol{p}\boldsymbol{p}^{\top}\big].$$

- Diagonal entries $np_i(1 - p_i) \to$ variance of each count.
- Off diagonals $-np_ip_j \to$ negative covariance (raising one count lowers the others to keep total n).

6. Why the Multinomial Coefficient? (Quick Derivation)

1. Choose positions for outcome 1: $\binom{n}{x_1}$.
2. Then outcome 2: $\binom{n - x_1}{x_2}$.
3. Continue until outcome $K - 1$; the last count is forced.
4. Multiply all choices and use $\binom{a}{b} = \frac{a!}{b!(a-b)!}$. Cancellations collapse to $\frac{n!}{x_1! \dots x_K!}$.

7. Big Picture – Four Sibling Distributions

Distribution	**# Outcomes (K)**	**# Trials (n)**
Bernoulli	2	1
Binomial	2	n (count successes)
Categorical	K	1
Multinomial	K	n (count each category)

8. Summary

- Multinomial extends Categorical to n trials.
- PMF combines an arrangement count and independent probabilities.
- Mean vector $n\boldsymbol{p}$; covariance $n\big[\text{diag}(\boldsymbol{p}) - \boldsymbol{p}\boldsymbol{p}^{\top}\big]$.
- Commonly used in NLP (bag-of-words counts), contingency tables, and Dirichlet–Multinomial Bayesian modelling.