Conditional Probability, Dependence, Exchangeability & de Finetti's Theorem

Lecture Roadmap

- **Warm-up:** Conditional probability and dependence/independence
- **Main topic:** Exchangeability (a weaker symmetry property than i.i.d.)
- **Key goal:** Understanding why exchangeability matters for Bayesian modelling
- **Destination:** de Finetti's theorem (informal statement + interpretation)

Big picture:

$$\text{i.i.d.} \Rightarrow \text{exchangeable} \quad \text{and} \quad \text{exchangeable} \Rightarrow \text{mixture of i.i.d. (de Finetti)}$$

Conditional Probability (Definition)

Let A and B be events with $\mathbb{P}(B) > 0$.

Definition:

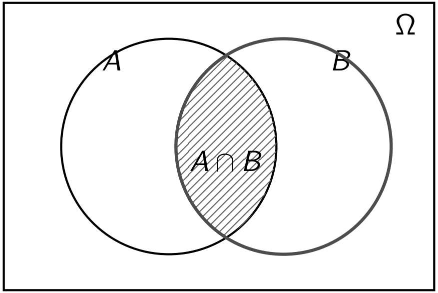
$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Interpretation:

- $\mathbb{P}(A \mid B)$ means: "the probability that A occurs *given* that we already know B occurred."
- Conditioning restricts the universe to B.
- Then we ask: what fraction of that restricted universe lies inside A?

Requirement: $\mathbb{P}(B) > 0$ (otherwise division is not defined).

Conditional Probability: Diagrammatic Intuition



Key idea: once we know B happened, the sample space becomes B.

$$\mathbb{P}(A \mid B) = \frac{\text{"size of } A \cap B\text{"}}{\text{"size of } B\text{"}}$$

In probability language: $\mathbb{P}(A \mid B) = \mathbb{P}(A \cap B)/\mathbb{P}(B)$.

Independence (and Conditional Independence)

(Unconditional) independence: A and B are independent if

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

Conditional independence: Let A, B, C be events with $\mathbb{P}(C) > 0$. We say

$$A \perp B \mid C \quad \iff \quad \mathbb{P}(A \cap B \mid C) = \mathbb{P}(A \mid C)\,\mathbb{P}(B \mid C).$$

Interpretation:

- Once we know C, the events A and B become “unrelated”.
- The information in A does not change the probability of B, given C.

Equivalent form (often useful):

$$A \perp B \mid C \quad \iff \quad \mathbb{P}(A \mid B, C) = \mathbb{P}(A \mid C),$$

whenever $\mathbb{P}(B \cap C) > 0$.

Why We Care: Moving Toward Bayesian Modelling

In Bayesian statistics, we allow ourselves to place a probability distribution on a **model parameter** θ:

$$\theta \sim \pi(\theta).$$

Frequentist perspective:

- The parameter θ is fixed (unknown but not random).
- Randomness lives only in the data X.

Bayesian perspective:

- θ represents our uncertainty / belief.
- $\pi(\theta)$ is a *subjective* probability distribution.

Question: When is it mathematically justified to write down $\pi(\theta)$ at all?

This is where **exchangeability** and **de Finetti's theorem** enter.

Exchangeability (Definition)

Let $(Y_1, \ldots, Y_N)$ have joint density / pmf

$$\pi(y_1, \ldots, y_N).$$

Definition (Exchangeability): $(Y_1, \ldots, Y_N)$ are exchangeable if for any permutation σ of $\{1, \ldots, N\}$,

$$\pi(y_1, \ldots, y_N) = \pi(y_{\sigma(1)}, \ldots, y_{\sigma(N)}).$$

Meaning:

- Reordering the labels does *not* change the joint distribution.
- The variables are "symmetric" in their roles.
- Exchangeability is weaker than i.i.d.

Quick fact: If $Y_1, \ldots, Y_N$ are i.i.d., then they are exchangeable.

Exchangeability vs Independence

Important: Exchangeability $\neq$ independence.

Independence

Means factorisation into a product:

$$\pi(y_1, \ldots, y_N) = \prod_{i=1}^{N} \pi(y_i).$$

Exchangeability

Means invariance under permutations:

$$\pi(y_1, \ldots, y_N) = \pi(y_{\sigma(1)}, \ldots, y_{\sigma(N)}).$$

Common misconception:

- Exchangeable variables can still be dependent.

Examples: Are They Independent? Are They Exchangeable?

Consider the following examples.

(1) Four Binomial variables

$$Y_i \sim \text{Binomial}(n, p), \quad i = 1, 2, 3, 4.$$

(2) Bivariate Normal

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma \right), \qquad \Sigma = \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix}.$$

(3) Trivariate Normal

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \Sigma \right), \qquad \Sigma = \begin{pmatrix} * & * & * \\ * & * & * \\ * & * & * \end{pmatrix}.$$

Task: For each case, decide:

Independent? Exchangeable?

Example (1): Binomial Case

$$Y_i \sim \text{Binomial}(n, p), \quad i = 1, 2, 3, 4, \quad \text{(assume i.i.d.)}$$

Independence: Yes, by assumption i.i.d.

Exchangeability: Yes, because the joint distribution factorises and each factor is identical:

$$\pi(y_1, y_2, y_3, y_4) = \prod_{i=1}^{4} \pi(y_i),$$

and reordering (y_1, y_2, y_3, y_4) does not change the product.

So:

$$(Y_1, Y_2, Y_3, Y_4) \text{ are independent and exchangeable.}$$

Example (2): Bivariate Normal

$$(X, Y) \sim \mathcal{N}(\mathbf{0}, \Sigma), \qquad \Sigma = \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix}.$$

Independence:

- X and Y are independent **iff** the covariance is 0.
- For a Gaussian, covariance 0 $\iff$ independence.
- So independence holds iff $\rho = 0$.

Exchangeability:

- Exchangeable means (X, Y) has same distribution as (Y, X).
- This is true if the joint distribution is symmetric under swapping.
- In many common cases (e.g. equal marginal variances), this holds.

Conclusion:

$$\text{Exchangeable: often yes (symmetry).} \qquad \text{Independent: only if } \rho = 0.$$

Example (3): Trivariate Normal

$$(X, Y, Z) \sim \mathcal{N}\left(\begin{pmatrix}1\\2\\3\end{pmatrix}, \Sigma\right).$$

Not independent (typically):

- In a multivariate normal, non-zero off-diagonal entries (covariances) usually imply dependence.
- Independence would require *all* off-diagonal covariances to be zero.

Exchangeability requires symmetry:

- If we reorder variables, we must reorder the mean vector and covariance structure accordingly.
- To compare $\pi(x, y, z)$ with $\pi(z, y, x)$, the mean vector would need to become $(3, 2, 1)^\top$ under relabelling, and Σ must be permuted.

Takeaway:

Exchangeability can hold even without independence,

From Conditional i.i.d. to Exchangeability (Proposition)

Proposition

If $\theta \sim \pi(\theta)$ and

$$Y_1, \dots, Y_N \text{ are conditionally i.i.d. given } \theta,$$

then marginally (unconditionally) the variables $Y_1, \dots, Y_N$ are exchangeable.

Meaning:

- Given the latent parameter θ, the data look i.i.d.
- But after we "integrate out" θ, dependencies appear.
- Despite possible dependence, we still retain symmetry under reordering:

$$(Y_1, \dots, Y_N) \text{ are exchangeable.}$$

Proof of the Proposition (Step-by-step)

Assume conditional i.i.d. given θ:

$$\pi(y_1, \dots, y_N \mid \theta) = \prod_{i=1}^{N} \pi(y_i \mid \theta).$$

Marginal joint distribution:

$$\pi(y_1, \dots, y_N) = \int \pi(y_1, \dots, y_N \mid \theta)\, \pi(\theta)\, d\theta.$$

Substitute the conditional i.i.d. factorisation:

$$\pi(y_1, \dots, y_N) = \int \left(\prod_{i=1}^{N} \pi(y_i \mid \theta) \right) \pi(\theta)\, d\theta.$$

Now apply any permutation σ:

$$\prod_{i=1}^{N} \pi(y_i \mid \theta) = \prod_{i=1}^{N} \pi(y_{\sigma(i)} \mid \theta),$$

What This Proposition Tells Us

Conditional i.i.d. $\Rightarrow$ exchangeability marginally.

This is a very common Bayesian modelling structure:

$$\theta \sim \pi(\theta), \qquad Y_i \mid \theta \overset{\text{i.i.d.}}{\sim} \pi(\cdot \mid \theta).$$

Key insights:

- Even if $Y_1, \ldots, Y_N$ are not independent marginally, they are symmetric.
- This symmetry is exactly the notion of exchangeability.
- This provides a natural bridge between *Bayesian latent-variable models* and *probabilistic symmetry assumptions*.

de Finetti's Theorem (Informal Statement)

Theorem (de Finetti, informal)

If $Y_1, \ldots, Y_N$ are exchangeable, then their joint distribution can be written as

$$\pi(y_1, \ldots, y_N) = \int \left(\prod_{i=1}^{N} \pi(y_i \mid \theta) \right) \pi(\theta)\, d\theta,$$

for some latent parameter θ and some distribution $\pi(\theta)$.

Interpretation:

- Exchangeability implies the data behave *as if* they were i.i.d. given a hidden θ.
- The distribution $\pi(\theta)$ can be viewed as a "prior".

So de Finetti gives (under suitable conditions) a **representation theorem**:

$$\text{Exchangeable} \iff \text{Mixture of i.i.d. models.}$$

Why de Finetti Matters for Bayesian Statistics

Bayesian philosophy: It is meaningful to assign probabilities to unknown parameters.

Frequentist objection:

- Parameters are fixed constants, not random quantities.
- Writing $\pi(\theta)$ seems unnatural or unjustified.

de Finetti's response:

- If you accept *exchangeability* of your observations, then there *exists* a latent parameter θ such that the data are i.i.d. given θ.
- Therefore a distribution $\pi(\theta)$ exists (mathematically).

Crucial nuance:

- de Finetti guarantees existence of some $\pi(\theta)$,
- but it does *not* guarantee your chosen prior is "good" or realistic.

Example: Normal Model with a Prior

Suppose a data model is

$$X_i \mid \mu \sim \mathcal{N}(\mu, \sigma^2), \qquad i = 1, \ldots, N.$$

Frequentist view: μ is fixed but unknown.

Bayesian view: we express uncertainty via a prior

$$\mu \sim \pi(\mu).$$

Then the marginal joint distribution is a mixture:

$$\pi(x_1, \ldots, x_N) = \int \left(\prod_{i=1}^{N} \mathcal{N}(x_i \mid \mu, \sigma^2) \right) \pi(\mu) \, d\mu.$$

Message: Exchangeability allows us to *interpret* this model as a coherent representation of beliefs.

Summary and What Comes Next

Today we covered:

- Conditional probability:

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

- Conditional independence:

$$A \perp B \mid C \iff \mathbb{P}(A \cap B \mid C) = \mathbb{P}(A \mid C)\mathbb{P}(B \mid C)$$

- Exchangeability:

$$\pi(y_1, \ldots, y_N) = \pi(y_{\sigma(1)}, \ldots, y_{\sigma(N)})$$

- Proposition: conditional i.i.d. given $\theta \Rightarrow$ exchangeable marginally
- de Finetti (informal): exchangeability $\Rightarrow$ mixture of i.i.d. (a "prior" exists)

Next lecture: Bayes' theorem

$$\pi(\theta \mid y) = \frac{\pi(y \mid \theta)\pi(\theta)}{\pi(y)}$$ and how to interpret posterior inference.