Lebesgue Decomposition and Probability Theory

Instructor: Vishnu Boddeti *Scribes: Preston Korytkowski, Shashank Jayaram*

1 Lebesgue Decomposition

1.1 Definitions

Definition 1 *A measure μ is called **absolutely continuous** with respect to λ if $\lambda(A) = 0 \Rightarrow \mu(A) = 0$ for all $A \in \mathcal{B}(\mathbb{R})$.*

Definition 2 *A measure μ is called **singular** (w.r.t. λ) if there exists $N \in \mathcal{B}(\mathbb{R})$ with $\lambda(N) = 0$ and $\mu(N^{\complement}) = 0$ ($\mu \perp \lambda$).*

Example: Dirac measure δ_0, where $\delta_0(\{0\}) = 1$ and $\delta_0(\mathbb{R} \setminus \{0\}) = 0$, is singular with respect to λ.

1.2 Theorem (Lebesgue Decomposition)

Theorem 3 *Let μ, ρ be a probability measure on $(\mathbb{R}, \mathcal{B})$. Then there exists a unique decomposition*

$$\rho = \rho_a + \rho_s$$

such that ρ_a is absolutely continuous w.r.t. μ, and ρ_s is singular w.r.t. μ.

Example: $\rho = \frac{1}{2}(\mathcal{N}(0,1), \delta_0)$ decomposes into absolutely continuous part $\frac{1}{2}\mathcal{N}(0,1)$ and singular part $\frac{1}{2}\delta_0$.

1.3 Cantor Distribution

The Cantor distribution is a non-trivial distribution that is singular w.r.t. λ.

Construct the Cantor set:

- Start with $C_0 = [0, 1]$
- Remove the middle third: $C_1 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$
- Repeat: $C_2 = [0, \frac{1}{9}] \cup [\frac{2}{9}, \frac{1}{3}] \cup [\frac{2}{3}, \frac{7}{9}] \cup [\frac{8}{9}, 1]$

- The Cantor set $C = \bigcap_{n=1}^{\infty} C_n$

The Cantor function can be used to define a probability distribution that is singular w.r.t. λ.

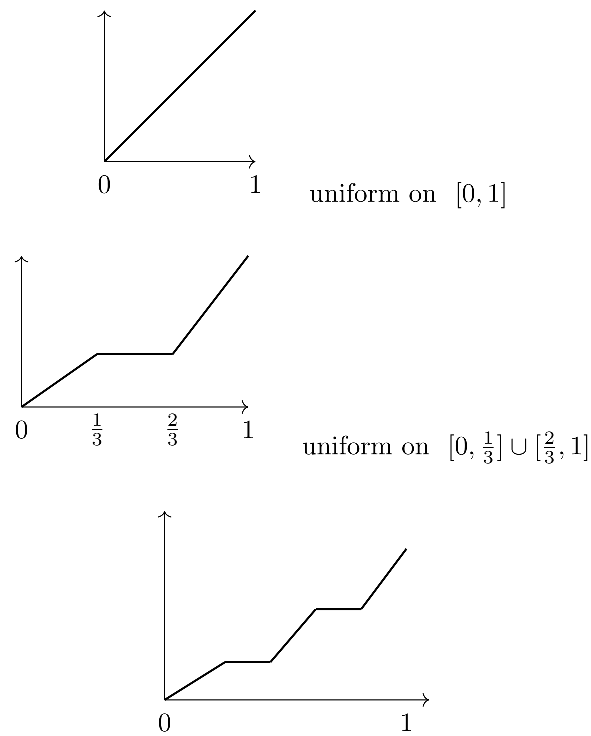


Figure 1: CDFs of the uniform distributions on $C_0, C_1, C_2, \ldots$. Taking the limit yields a distribution supported on the Cantor set.

Proposition 4 *The Cantor set has the following properties:*

- *It is compact.*
- *It is non-empty and has no interior points.*
- *Every point is a boundary point.*

Proposition 5 *The associated Cantor function has the following properties:*

- *It is continuous.*
- *It defines a valid probability measure.*
- *Its absolutely continuous component is zero.*

2 Cumulative Distribution Function (CDF)

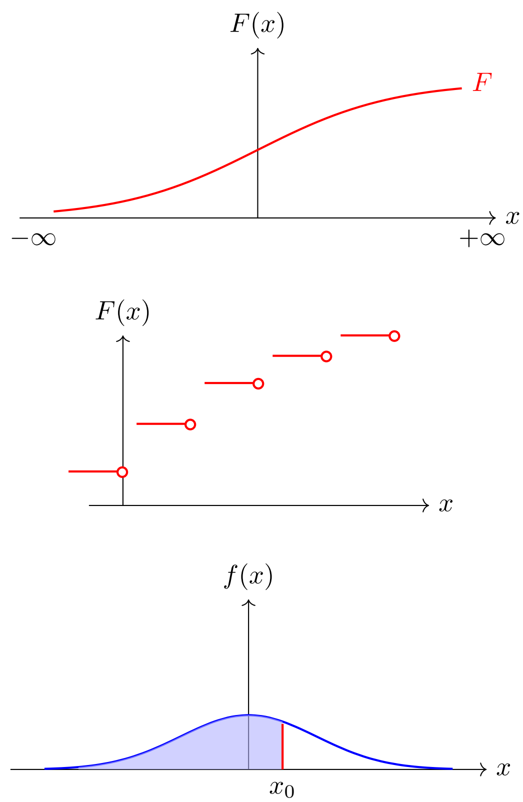
Definition 6 *Let P be a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. The function:*

$$F(x) = P((-\infty, x])$$

is called a ***cumulative distribution function (CDF)*** *that satisfies the following properties:*

(i) *F is monotonically increasing;* $\lim_{x\to-\infty} F(x) = 0$, *and* $\lim_{x\to\infty} F(x) = 1$.

(ii) *F is right-continuous: If $x_n \downarrow x$, then $F(x_n) \to F(x)$.*



PDF of a normal distribution; shaded region = $P(X \leq x_0)$

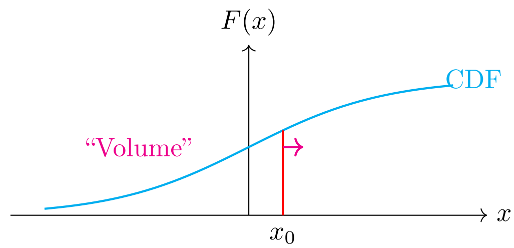


Figure 2: Top: Typical CDFs — smooth (continuous) and stepwise (discrete). Middle: PDF of a continuous distribution and the shaded probability up to x_0. Bottom: The corresponding CDF value is the accumulated area under the PDF.

2.1 Existence and Uniqueness

Theorem 7 *Let $F : \mathbb{R} \to \mathbb{R}$ be a function satisfying properties (i) and (ii) above. Then there exists a unique probability measure P on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that*

$$P((-\infty, x]) = F(x)$$

Remark 8 *This relationship works both ways — given a PDF, we can construct the CDF, and given a valid CDF, we can construct a unique corresponding PDF.*

3 Random Variables

Definition 9 *Let $(\Omega, \mathcal{A}, P)$ be a probability space and $(\tilde{\Omega}, \tilde{\mathcal{A}})$ a measurable space. A mapping $X : \Omega \to \tilde{\Omega}$ is called a* ***random variable*** *if X is measurable, i.e., for all $\tilde{A} \in \tilde{\mathcal{A}}$:*

$$X^{-1}(\tilde{A}) := \{\omega \in \Omega | X(\omega) \in \tilde{A}\} \in \mathcal{A}$$

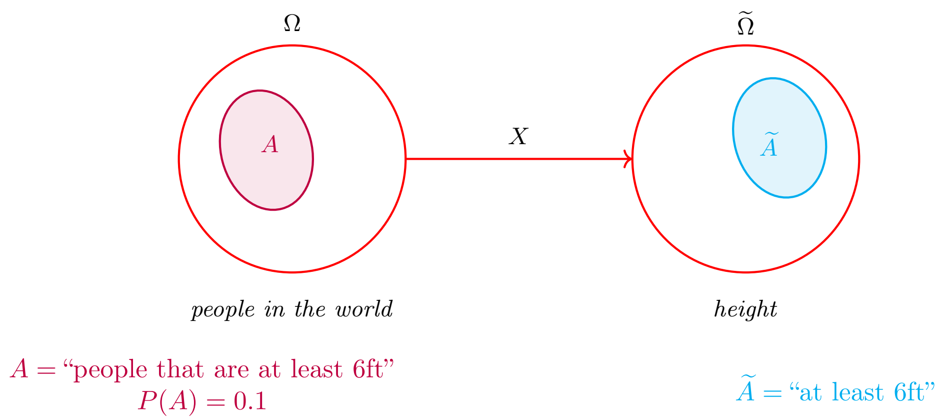


Figure 3: A random variable $X : \Omega \to \widetilde{\Omega}$ maps people to their heights. Sets like A in Ω and $\widetilde{A}$ in $\widetilde{\Omega}$ are linked through $X^{-1}(\widetilde{A}) = A$.

Example: Sum of two dice:

- Sample space: $\Omega = \{(i, j) \mid 1 \leq i, j \leq 6\}$
- Event A: sum $= 3 \Rightarrow A = \{(1, 2), (2, 1)\}$
- $P(A) = \frac{2}{36} = \frac{1}{18}$
- Define random variable $X(\omega) = i + j$ for $\omega = (i, j)$

3.1 Combining Random Variables

Theorem 10 *Suppose X_i, $i = 1, 2, \ldots$ are all measurable random variables. Then the following are also measurable random variables:*

1. $X_1 + X_2 + X_3 + \cdots + X_n$
2. X_1^2
3. cX_1 *for any* $c \in \mathbb{R}$
4. $X_1 X_2$
5. $\inf\{X_n\ ;\ n \geq 1\}$
6. $\liminf_{n\to\infty} X_n$
7. $\sup\{X_n\ ;\ n \geq 1\}$
8. $\limsup_{n\to\infty} X_n$

3.2 Induced Measure / Distribution of a Random Variable

Definition 11 *Given* $X : \Omega \to \tilde{\Omega}$*, for* $\tilde{A} \in \tilde{\mathcal{A}}$*, the* ***distribution*** *of* X *is defined as:*

$$P_X(\tilde{A}) = P(X^{-1}(\tilde{A}))$$

This defines a probability measure on $(\tilde{\Omega}, \tilde{\mathcal{A}})$*.*

3.3 Sigma-algebra Induced by a Random Variable

Definition 12 *Let* $X : (\Omega, \mathcal{A}, P) \to (\tilde{\Omega}, \tilde{\mathcal{A}})$*. The* σ***-algebra induced by*** X *is defined as:*

$$\sigma(X) := \{X^{-1}(\tilde{A}) \mid \tilde{A} \in \tilde{\mathcal{A}}\}$$

This is the smallest σ*-algebra that makes* X *measurable.*

4 Conditional Probability

Basic probability operations:

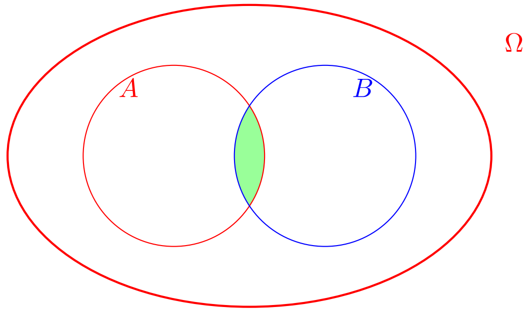
- $P(A \cap B)$: probability of A and B
- $P(A \cup B)$: probability of A or B

Definition 13 *Let* $(\Omega, \mathcal{A}, P)$ *be a probability space and* $A, B \in \mathcal{A}$ *with* $P(B) > 0$*. The* ***conditional probability*** *of* A *given* B *is defined as:*

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

Theorem 14 *The mapping* $P_B : \mathcal{A} \to [0, 1]$ *defined by* $P_B(A) = P(A \mid B)$ *is a probability measure on* $(\Omega, \mathcal{A})$*. It is called the* ***conditional distribution*** *of* P *with respect to* B*.*

$$P(A \cap B) = P(\text{“A and B”})$$



$$P(A \cup B) = P(\text{“A or B”})$$

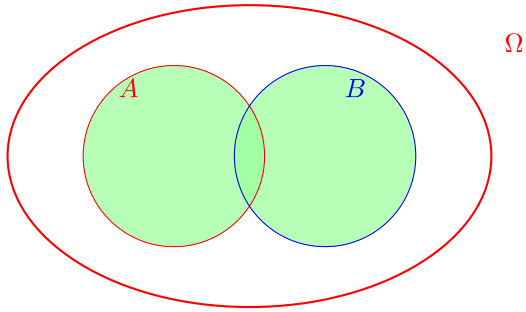


Figure 4: Venn diagrams illustrating $P(A \cap B)$ (overlap only) and $P(A \cup B)$ (union of both).

4.1 Examples

Example: Example with two dice:

$$P(\text{Sum is 7} \mid \text{First die is 2}) = \frac{P(\text{Sum is 7 and first die is 2})}{P(\text{First die is 2})}$$

Example: Let Ω be all people on Earth.

- A: person has a disease
- B: person is vaccinated

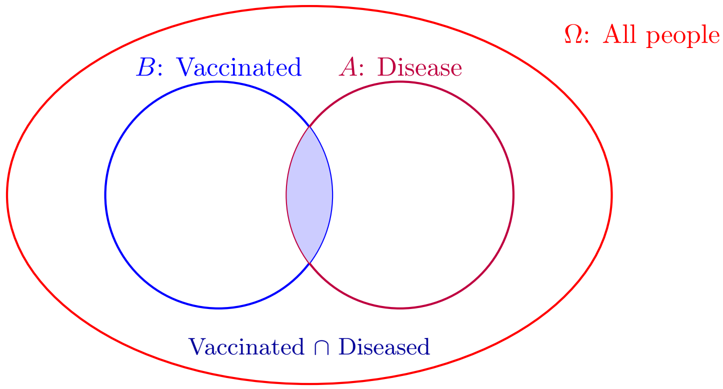


Figure 5: Conditional probability illustrated: $P(\text{Disease} \mid \text{Vaccinated}) = \frac{P(A \cap B)}{P(B)}$

Then the conditional probability:

$$P(\text{Disease} \mid \text{Vaccinated}) = \frac{P(\text{Vaccinated and Disease})}{P(\text{Vaccinated})}$$

4.2 Application: Naive Bayes Classifier (NBC)

The **Naive Bayes Classifier** is a probabilistic model commonly used in machine learning for classification tasks. It is based on applying **Bayes' Rule** with the *naive assumption* that features are conditionally independent given the class.

Definition 15 (Bayes' Rule) *For events A and B with $P(B) > 0$,*

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

Example Task: Given the data below, predict the ailment of a `sneezing builder`:

SYMPTOM	OCCUPATION	AILMENT
sneezing	nurse	flu
sneezing	farmer	hayfever
headache	builder	concussion
headache	builder	flu
sneezing	teacher	flu
headache	teacher	concussion
sneezing	builder	???

Goal: Predict $P(\text{flu} \mid \text{sneezing, builder})$

Step 1: Use Bayes' Rule

$$P(\text{flu} \mid \text{sneezing, builder}) = \frac{P(\text{flu}) \cdot P(\text{sneezing} \mid \text{flu}) \cdot P(\text{builder} \mid \text{flu})}{P(\text{sneezing, builder})}$$

Step 2: Estimate Probabilities from Data

- $P(\text{flu}) = 0.5$
- $P(\text{sneezing} \mid \text{flu}) = 0.66$
- $P(\text{builder} \mid \text{flu}) = 0.33$
- $P(\text{sneezing}, \text{builder} \mid \text{flu}) = 0.66 \cdot 0.33 = 0.22$
- $P(\text{sneezing}) = 0.5$
- $P(\text{builder}) = 0.33$
- $P(\text{sneezing}, \text{builder}) = 0.5 \cdot 0.33 = 0.165$

Step 3: Compute Final Probability

$$P(\text{flu} \mid \text{sneezing, builder}) = \frac{0.5 \cdot 0.22}{0.165} = \frac{0.11}{0.165} \approx 0.66$$

So, the **sneezing builder has flu with probability 0.66**.

Key Assumption

The Naive Bayes Classifier assumes:

$$P(\text{symptom, occupation} \mid \text{ailment}) = P(\text{symptom} \mid \text{ailment}) \cdot P(\text{occupation} \mid \text{ailment})$$

This assumption of conditional independence rarely holds in practice, but the NBC often performs surprisingly well regardless.

References

[1] Random Variables and Measurable Functions. Available at: `https://sas.uwaterloo.ca/~dlmcleis/s901/chapt3.pdf`. Accessed March 2025.

[2] Machine Learning - Lecture 4: The Naive Bayes Classifier. Available at: `https://users.sussex.ac.uk/~christ/crs/ml/lec02b.html`. Accessed March 2025.