CSE 840: Computational Foundations of Artificial Intelligence March 24, 2025

Probability Measure

Instructor: Vishnu Boddeti *Scribe: Atharva Kirkole*

1 Introduction

This lecture dives into probability measures, their mathematical foundations, and key examples.

1.1 Basic Definitions

Definition 1 (Probability Measure) *A probability measure involves:*

- *An abstract space* Ω *(the sample space)*
- *A corresponding* σ*-algebra* $\mathcal{A}$ *containing measurable events*

The σ-algebra $\mathcal{A}$ must satisfy:

- Closure under complements: If $A \in \mathcal{A}$, then $A^c \in \mathcal{A}$
- Closure under countable unions: If $\{A_i\}_{i=1}^{\infty} \subseteq \mathcal{A}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$

Definition 2 (Measure) *A measure* μ *is a function* $\mu : \mathcal{A} \to [0, \infty]$ *that is countably additive. For any sequence* $\{A_i\}_{i=1}^{\infty}$ *of pairwise disjoint sets:*

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i)$$

Definition 3 (Probability Space) *The triple* $(\Omega, \mathcal{A}, P)$ *constitutes a probability space when:*

- $P(\Omega) = 1$
- $P : \mathcal{A} \to [0, 1]$ *is a probability measure*

2 Examples of Probability Measures

2.1 Discrete Probability Measures

Single Die:

- Sample space: $\Omega = \{1, 2, 3, 4, 5, 6\}$

- σ-algebra: $\mathcal{A} = \mathcal{P}(\Omega)$
- Probability measure: $P(\{i\}) = \frac{1}{6}$ for each $i \in \Omega$

Two Dice:

- Sample space: $\Omega = \{(i, j) : 1 \leq i, j \leq 6\}$
- σ-algebra: $\mathcal{A} = \mathcal{P}(\Omega)$
- Probability measure: $P(\{(i, j)\}) = \frac{1}{36}$ for each $(i, j) \in \Omega$

2.2 Continuous Probability Measures

Normal Distribution

- Sample space: $\Omega = \mathbb{R}$
- σ-algebra: Borel σ-algebra $\mathcal{B}(\mathbb{R})$
- Probability density function:

$$f_X(x|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- Probability measure: For $A \in \mathcal{B}(\mathbb{R})$,

$$P(A) = \int_A f_X(x|\mu, \sigma)\, dx$$

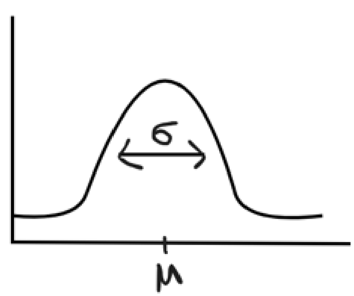

Figure 1: Fig.1 - Normal distribution

3 Types of Probability Measures

3.1 Discrete Measures

Definition 4 (Discrete Measure) *A discrete measure on countable space* $\Omega = \{x_1, x_2, \ldots\}$ *with* $\mathcal{A} = \mathcal{P}(\Omega)$ *satisfies:*

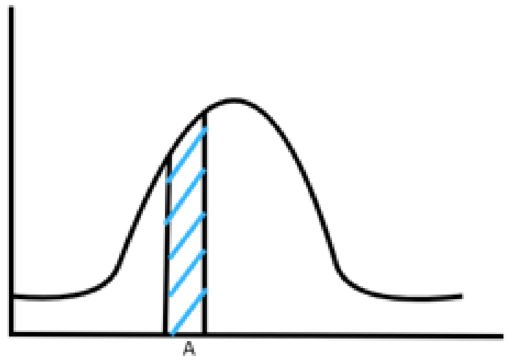

Figure 2: Fig.2 - Normal distribution Integration

- $P(\{x_i\}) = p_i$ *with* $0 \leq p_i \leq 1$
- $\sum_{i=1}^{\infty} p_i = 1$

For any $A \in \mathcal{A}$*:*

$$P(A) = \sum_{\{i | x_i \in A\}} p_i$$

Coin Toss: $\Omega = \{\text{H}, \text{T}\}$ with $P(\{\text{H}\}) = p$ and $P(\{\text{T}\}) = 1 - p$

3.2 Dirac Measures

Definition 5 (Dirac Measure) *For* $x \in \mathbb{R}$*, the Dirac measure* δ_x *on* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *is:*

$$\delta_x(A) = \begin{cases} 1 & \textit{if } x \in A \\ 0 & \textit{otherwise} \end{cases}$$

Proposition 6 *Any discrete measure on* $\mathbb{R}$ *can be expressed as a sum of Dirac measures:*

$$P = \sum_{i=1}^{n} p_i \delta_{x_i}$$

The die measure from Example 2.1 can be written as:

$$P = \frac{1}{6}(\delta_1 + \delta_2 + \cdots + \delta_6)$$

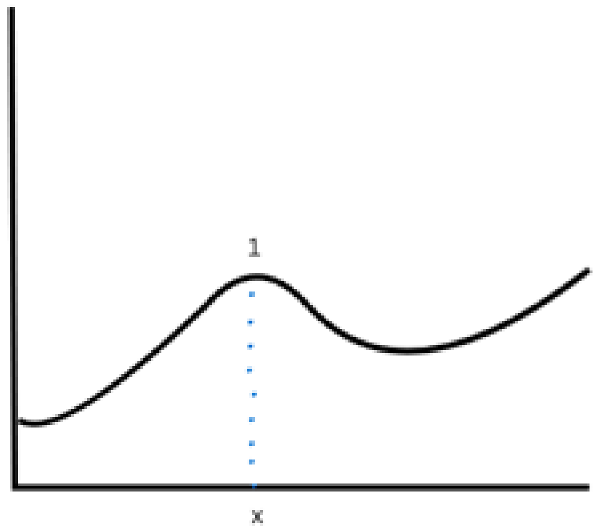

Figure 3: Fig.3 - Dirac Measure Figure

4 Probability Measures with Density Functions

Definition 7 (Measure with Density) *In the context of the measure space* $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$*, where* λ *is the Lebesgue measure, a measurable function* $f : [0, \infty]$ *satisfying:*

$$\int_{\mathbb{R}} d\lambda = 1 \quad \text{which is} \int_A f(x)dx = 1$$

defines a probability measure γ *on* $\mathbb{R}^n$ *via:*

$$\gamma(A) = \int_A f(x)dx \quad \text{for all } A \in \mathcal{A}$$

γ *is the probability measure on* $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ *with density f. We denote this relationship as* $\gamma = f\lambda$*.*

Remark 8 *Not all probability measures on* $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ *admit density functions. For example:*

- *The Dirac measure* δ_x *cannot be expressed as* $\delta_x = f\lambda$ *for any f*
- *Discrete measures like* $\sum_{i=1}^n p_i \delta_{x_i}$ *lack density functions*

5 Absolute Continuity of Measures

Definition 9 (Absolute Continuity) *A probability measure* γ *on* $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ *is* absolutely continuous *with respect to* μ $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ *if for every* μ*-null set is also a* γ*-null set :* $\forall\ B \in \mathcal{B}(\mathbb{R}^n)$ *: if* $\mu(B) = 0$ *then* $\gamma(B) = 0$

$$\mu(A) = 0 \implies \gamma(A) = 0 \quad \forall A \in \mathcal{B}(\mathbb{R}^n)$$

Notation; $\gamma << \mu$

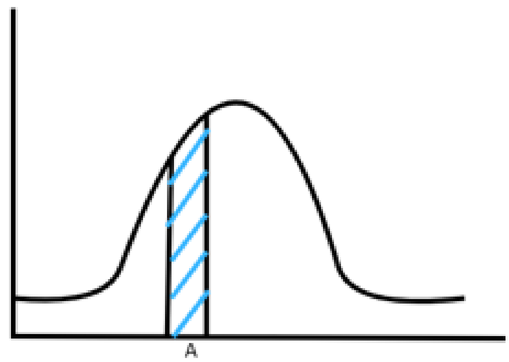

Figure 4: Fig.4 - Integral of continuous distribution

6 Radon-Nikodym Theorem

Theorem 10 (Radon-Nikodym) *For σ-finite measures μ and γ on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, the following are equivalent:*

1. *γ has a density with respect to μ (i.e., $\exists f$ measurable with $\gamma(A) = \int_A f d\mu$)*
2. *$\gamma \ll \mu : \gamma$ is absolutely continuous with respect to μ*

If $\gamma \ll \mu$, Then $\exists \phi$ such that, $\gamma(A) = \int_A \phi d\mu$ Then, ϕ exists and is unique.

6.1 Proof:

Step 1: Define the Set of Dominating Functions

Define the set G of all measurable functions $g : X \to [0, \infty)$ satisfying

$$\int_A g \, d\mu \leq \gamma(A) \quad \forall A \in \mathcal{B}.$$

- G is non-empty because the zero function $g \equiv 0$ belongs to G.
- If $g_1, g_2 \in G$, then $\max(g_1, g_2) \in G$ since for any A, we can split A into $A_1 = \{g_1 \geq g_2\}$ and $A_2 = \{g_1 < g_2\}$, leading to
$$\int_A \max(g_1, g_2) \, d\mu \leq \gamma(A).$$

Step 2: Define the Supremum and Construct a Maximizing Sequence

Define

$$X = \sup_{g \in G} \int_X g \, d\mu.$$

Since γ is σ-finite, we have $X < \infty$. Choose a sequence $\{g_n\} \subset G$ such that

$$\lim_{n\to\infty} \int_X g_n \, d\mu = X.$$

Define the increasing sequence

$$f_n = \max(g_1, \ldots, g_n),$$

which satisfies $f_n \in G$ and $\lim_{n\to\infty} \int_X f_n \, d\mu = X$ by construction.

Step 3: Construct the Density Function

Define

$$f = \sup_n g_n.$$

By the monotone convergence theorem,

$$\int_A f \, d\mu = \sup_n \int_A g_n \, d\mu \leq \gamma(A) \quad \forall A \in \mathcal{B}.$$

However, since $f_n \in G$ and $\int_X f_n d\mu$ approaches the supremum X, we conclude

$$\int_A f \, d\mu = \gamma(A),$$

ensuring f is the required Radon–Nikodym derivative.

Step 4: Uniqueness

Suppose there exist two such functions f_1 and f_2. Then

$$\int_A (f_1 - f_2) \, d\mu = 0 \quad \forall A \in \mathcal{B}.$$

Let $A = \{f_1 > f_2\}$. Then $\mu(A) = 0$ since $\int_A f_1 \, d\mu = \int_A f_2 \, d\mu$. Similarly, $\mu(\{f_2 > f_1\}) = 0$, implying $f_1 = f_2$.

Thus, the function f is unique up to sets of measure zero, completing the proof.

7 Bertrand paradox

Probability that a random chord exceeds $\sqrt{3}$ Consider the unit circle and the length of a random chord. The probability that the chord length $L > \sqrt{3}$ depends on how we define "random":

1. **Uniform endpoint method (Probability $= 1/3$)**: Fix one endpoint and choose the other uniformly. The chord length exceeds $\sqrt{3}$ when the angle $\theta \in (2\pi/3, 4\pi/3)$.
2. **Uniform radius method (Probability $= 1/2$)**: Choose a random radius and a point on it uniformly for the chord. The condition holds when the distance $d < 1/2$ from center.
3. **Uniform midpoint method (Probability $= 1/4$)**: Choose the midpoint uniformly in the disk. The condition holds when the midpoint lies in a circle of radius $1/2$.

We calculate each case:

1. **Endpoint method**:

$$P(L > \sqrt{3}) = P(\cos\theta < -1/2) = \frac{4\pi/3 - 2\pi/3}{2\pi} = \frac{1}{3}$$

2. **Radius method**: The length condition becomes $2\sqrt{1-d^2} > \sqrt{3} \Rightarrow d < 1/2$. Thus:

$$P = \frac{1/2}{1} = \frac{1}{2}$$

3. **Midpoint method**: The area where $L > \sqrt{3}$ is a circle of radius 1/2:

$$P = \frac{\pi(1/2)^2}{\pi(1)^2} = \frac{1}{4}$$

Discussion

This example illustrates the **Bertrand paradox**, showing how probability results depend critically on the the selected defined sample space. The different probabilities (1/2, 1/3, 1/4) correspond to different natural σ-algebras and measures on the space of chords.

8 Applications in Artificial Intelligence

Probability measures play a crucial role in AI, particularly in probabilistic modeling, Bayesian inference, and deep learning. Many AI systems rely on probability distributions to model uncertainty and make informed decisions.

Fact: Gaussian Processes in Machine Learning: Gaussian Processes (GPs) leverage probability measures to define distributions over functions. This enables AI models to perform regression and classification tasks with built-in uncertainty. GPs are widely used in Bayesian optimization, reinforcement learning, and time-series forecasting.