

MEASURES & DENSITIES

PROBABILISTIC DECISION MAKING VU – PRACTICALS

(REINFORCEMENT LEARNING KU)

Thomas Wedenig

Oct 17, 2025

Institute of Machine Learning and Neural Computation
Graz University of Technology, Austria

A probability space is a triple

$$(\Omega, \mathcal{F}, \mathbb{P})$$

σ -Algebra

Let $\Omega \neq \emptyset$ be a set and let $\mathcal{F} \subseteq \mathcal{P}(\Omega)$. Then \mathcal{F} is a σ -algebra over Ω if

- $\Omega \in \mathcal{F}$
- $A \in \mathcal{F} \implies A^c := \Omega \setminus A \in \mathcal{F}$
(closed under complement)
- $A_1, A_2, \dots \in \mathcal{F} \implies \bigcup_{i \geq 1} A_i \in \mathcal{F}$
(closed under countable union)

Probability Measure

Let (Ω, \mathcal{F}) be a measurable space. A map

$$\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$$

is a **probability measure** if

- $\mathbb{P}(\Omega) = 1$
- for pairwise disjoint $A_1, A_2, \dots \in \mathcal{F}$:
 $\mathbb{P}\left(\bigcup_{i \geq 1} A_i\right) = \sum_{i \geq 1} \mathbb{P}(A_i)$

THE NEED FOR MEASURE THEORY

A probability space is a triple

$$(\Omega, \mathcal{F}, \mathbb{P})$$

Hold on! Why even care about \mathcal{F} ?

Proposal: Just set $\mathcal{F} := \mathcal{P}(\Omega)$ and stop talking about it 😂

For **countable** Ω , $\mathcal{F} := \mathcal{P}(\Omega)$ works just fine 🥰

For **uncountable** Ω , things become ... tricky 😬

- Let $(\mathbb{R}^d, \mathcal{P}(\mathbb{R}^d))$ be a measurable space
- We want to define a **measure** $\lambda : \mathcal{P}(\mathbb{R}^d) \rightarrow [0, \infty]$
- ... that **generalizes the notion of volume**
 - **Length** in \mathbb{R} , **Area** in \mathbb{R}^2 , **Volume** in \mathbb{R}^3 , ...
- λ is called the **Lebesgue measure**
- For e.g. $d = 1$, we demand
 - $\lambda(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \lambda(A_i)$ for all pairwise disjoint $A_i \in \mathcal{P}(\mathbb{R})$.
 - $\lambda([a, b]) = (b - a)$ for all $a, b \in \mathbb{R}$
 - $\lambda(A + t) = \lambda(A)$ for all $t \in \mathbb{R}, A \in \mathcal{P}(\mathbb{R})$

It turns out there exists no such λ on $\mathcal{P}(\mathbb{R})$ 🤔

- We can't consistently measure length for all possible subsets of \mathbb{R}
- ... because $\mathcal{P}(\mathbb{R})$ contains **nasty sets**
- If we would e.g. use $\mathcal{F} := \{\emptyset, \mathbb{R}\}$, we don't have this problem 🤔
- ... but \mathcal{F} is **useless** because we can't measure anything 😅

We need something **between** $\{\emptyset, \mathbb{R}\}$ and $\mathcal{P}(\mathbb{R})$!

- The **sweet spot** is $\mathcal{F} := \mathcal{B}(\mathbb{R}^d)$, the **Borel σ -algebra** on \mathbb{R}^d
- Contains **all sets we will ever need**, while **excluding the nasty sets** $\in \mathcal{P}(\mathbb{R}^d)$ 🤖
- When we don't explicitly define \mathcal{F} we **typically assume** $\mathcal{F} := \mathcal{B}(\mathbb{R}^d)$

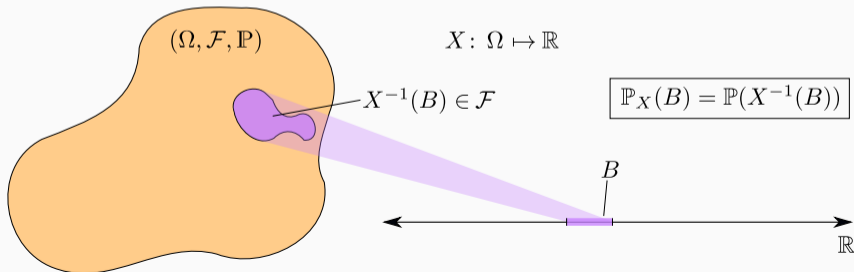
PROBABILITY DENSITY FUNCTIONS

Random Variables (RVs)

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability triple with $\mathcal{F} := \mathcal{B}(\mathbb{R})$. A **random variable (RV)** is a measurable function $X : \Omega \rightarrow \mathbb{R}$. The **distribution** of X is defined as

$$\mathbb{P}_X(B) := \mathbb{P}(X^{-1}(B))$$

for any $B \in \mathcal{B}$ and with $X^{-1}(B) := \{\omega \in \Omega : X(\omega) \in B\}$

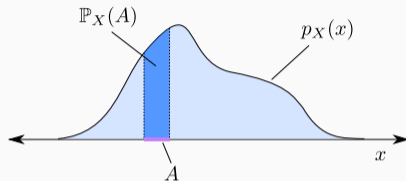


Probability Density Function (PDF)

Let $X : \Omega \rightarrow \mathbb{R}$ be a random variable (RV).

If it exists, a function $p_X : \mathbb{R} \rightarrow [0, \infty]$ is a **density** of X if for any event A ,

$$\int_A p_X(x) dx = \mathbb{P}_X(A).$$



- Note that $\int_{\mathbb{R}} p_X(x) dx = 1$ since

$$\mathbb{P}_X(\mathbb{R}) = \mathbb{P}(X^{-1}(\mathbb{R})) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in \mathbb{R}\}) = \mathbb{P}(\Omega) = 1.$$

- **Density \neq probability !**
- Technical Remark: Integration is w.r.t. the Lebesgue measure, i.e., $\int_A p_X(x) dx = \int_A p_X(x) d\lambda(x)$

Definition

Let $\Omega := \mathbb{R}$, $\mathcal{F} := \mathcal{B}(\mathbb{R})$. The **Dirac measure** at 0 is defined as

$$\delta_0(A) := \begin{cases} 1, & 0 \in A, \\ 0, & 0 \notin A \end{cases}$$

and $(\Omega, \mathcal{F}, \delta_0)$ is a **valid probability space**.

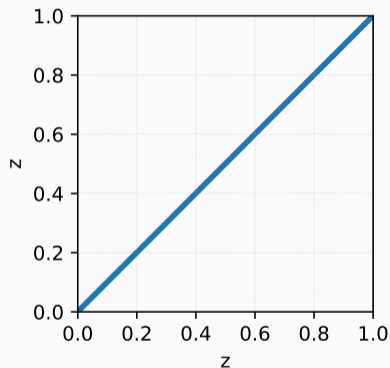
Does this admit a density? 🤔

Can we find $p : \mathbb{R} \rightarrow [0, \infty]$ such that $\int_A p(x) \, dx = \delta_0(A)$ for all $A \in \mathcal{F}$?

$$\delta_0(A) := \begin{cases} 1, & 0 \in A, \\ 0, & 0 \notin A \end{cases} \quad \int_A p(x) dx \stackrel{!}{=} \delta_0(A)$$

- No such integrable p exists !
 - (w.r.t. the Lebesgue measure)
- Intuition: $\delta_0(\{0\}) = 1$, but $\{0\}$ has no “length”
 - i.e., $\lambda(\{0\}) = \lambda([0, 0]) = 0$
- Hence, no matter which p , we always have $\int_{\{0\}} p(x) dx = 0$ 😞

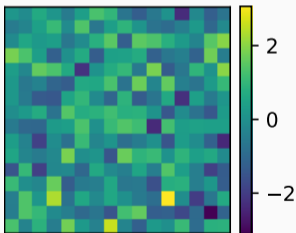
Let $Z \sim \text{Unif}([0, 1])$ and define $X := (Z, Z)$.



Again, \mathbb{P}_X does not admit a density (w.r.t. λ), since a line has no “area”.

- Popular generative model: **Generative Adversarial Networks (GANs)**
- We have a (smooth) **neural network** $G_{\theta} : \mathbb{R}^n \rightarrow \mathbb{R}^d$ with $n < d$
- **Idea:** Sampling from the model amounts to drawing

$\mathbf{z} \sim \mathcal{N}(\mathbf{z}; \mathbf{0}, I) \in \mathbb{R}^n$ followed by computing $\mathbf{x} = G_{\theta}(\mathbf{z}) \in \mathbb{R}^d$



Source: thispersondoesnotexist.com

GAN

$$\mathbf{z} \sim \mathcal{N}(\mathbf{z}; \mathbf{0}, I) \in \mathbb{R}^n, \quad \mathbf{x} = G_{\boldsymbol{\theta}}(\mathbf{z}) \in \mathbb{R}^d, \quad n < d$$

- The “model” is the **pushforward** probability measure

$$\mathbb{P}_{\mathbf{x}}(A) = \mathbb{P}_{\mathbf{z}}(G_{\boldsymbol{\theta}}^{-1}(A))$$

for any $A \in \mathcal{B}(\mathbb{R}^d)$

- Since G is smooth, all generated \mathbf{x} lie on a n -dimensional **manifold** $\mathcal{M} \subset \mathbb{R}^d$

Since $\lambda(\mathcal{M}) = 0$, **no density $p_{\mathbf{x}}$ exists** (w.r.t. λ) **!**

Thus, no **maximum likelihood training** possible 😬

GAN

$$\mathbf{z} \sim \mathcal{N}(\mathbf{z}; \mathbf{0}, I) \in \mathbb{R}^n, \quad \mathbf{x} = G_{\theta}(\mathbf{z}) \in \mathbb{R}^d, \quad n < d$$

How could we fix this? 🤔

- Solution 1: Set $n = d$ and make G_{θ} invertible → **Normalizing Flow**
- Solution 2: Keep $n < d$, but add (Gaussian) noise to output:

$$\mathbf{x} = G_{\theta}(\mathbf{z}) + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, I)$$

→ model of a **Variational Autoencoder (VAE)**

QUIZ TIME!
fbr.io/pdmp2

