CSE 840: Computational Foundations of Artificial Intelligence November 08, 2023

Definition of a probability measure, discrete, density; Radon-Nikodym

Instructor: Vishnu Boddeti *Scribe: Xinnan Dai, Jay Revolinsky, Shenglai Zeng*

1 Probability Measure

Definition 1

- Given space Ω ("abstract space")
- Need a r-algebra A_R on $Omega$. ("measurable events")
  - $A \in A_r \implies A^C \in A_r$
  - $(A_i)_{i\in\mathbb{N}} \subset A_r \implies \bigcup\limits_{i=1}^{\infty} A_i \in A_r$ ("countable unions")
  - $\varnothing, \Omega \in A_r$
  - countable intersections
- A measure μ on (Ω, A_r) is a function $\mu : A_r \to [0, \infty]$ that is countably additive: If $(A_i)_{i\in\mathbb{N}}$ is a sequence of pairwise disjoint sets, then $\mu(\bigcup\limits_{i=1}^{\infty} A_i) = \sum\limits_{i=1}^{\infty} \mu(A_i)$

A measure P on a measurable space (Ω, A_r) is called a *probability measure* if $P(\Omega) = 1$. The elements of A_r are called events. Then (Ω, A_r, P) is called a *probability space*.

Example (1):

Throw a die

$\Omega = 1, 2, ..., 6, A_r = P(\Omega)$ (r-algebra generated by the "elementary events" $\{1\}, \{2\}...\{6\}$).

P can be defined uniquely by assigning $P(\{1\}) = P(\{2\}) = ... = P(\{6\}) = \frac{1}{6}$

For example $P(\{1, 5\}) = P(\{1\}) + P(\{5\}) = \frac{1}{3}$

Throw two dice:

$\Omega = \{1, 2, ..., 6\} \times \{1, 2, ...6\} = \{\underbrace{(1, 1)}_{first\ die, second\ die}, (1, 2)....\}$ all of which are elementary events

$A_r = P(\Omega)$

$P(\{(i, j)\}) = \frac{1}{36}$

Example (2): Normal distribution

$\Omega = \mathbb{R}$

$A_r =$ Borel-r-algebra

$f_{\mu,r} : \mathbb{R} \implies \mathbb{R}$

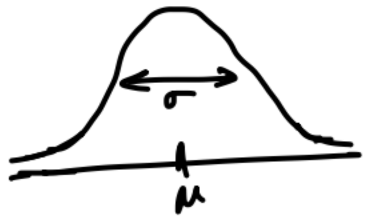

Figure 1: μ

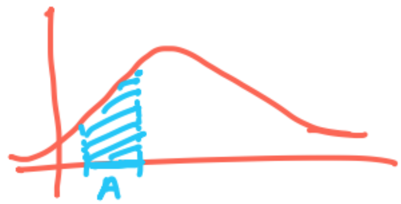

Figure 2: A

$x \rightarrowtail \frac{1}{\sqrt{2\pi r^2}} \exp\left(\frac{-(x-\mu)^2}{2r^2}\right)$

$P : A_r \to [0,1], P(A) := \int_A f_{\mu,r}(x)dx$

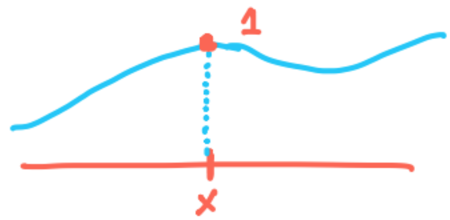

Figure 3: Dirac measure

2 Different Types of Probability Measures

Definition 1 Discrete measure:

$\Omega = \{x_1, x_2, ...\}$ finite and countable

$A_r = P(\Omega)$

We define a probability measure $P : A_r \to [0,1]$ by assigning probabilities to the "elementary events":

$$P(\{x_i\}) =: P_i$$

with $0 \leq P_i \leq 1, \Sigma_i P_i = 1$

For $A \in A_r$ we assign

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

Examples: a coin toss, distribution on Q

Definition 2 Dirac measure:

For $x \in \mathbb{R}$, we define the Dirac measure δ_x on $(\mathbb{R}, B(\mathbb{R}))$ by setting $\delta_x(A) = \begin{cases} 1 & x \in A \\ 0 & otherwise \end{cases}$ sometimes this is called a point mass at a point x. A discrete measure on $\mathbb{R}$ can be written as a sum of Dirac measures. For example, throwing a die can be considered as

$$\tfrac{1}{6}(\delta_1 + \delta_2 + ... + \delta_6)$$

Measures with a density

Consider $(\mathbb{R}^n, B(\mathbb{R}^n))$ and the Lebesque measure λ. Consider a function $f : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ that is measurable and satisifies $\int f d\lambda = 1 \implies \int f(x)dx = 1$.

Then we define a measure γ on $\mathbb{R}^n$ by setting, for all $A \in A_r$,

$$\gamma(A) := \int_A f(x)dx$$

γ is the probability measure on $(\mathbb{R}^n, B(\mathbb{R}^n))$ with density f.

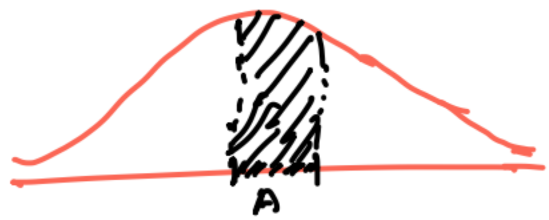

Figure 4: $\mu(A) = 0 \implies \int_A f d\mu \equiv \gamma(A) = 0$

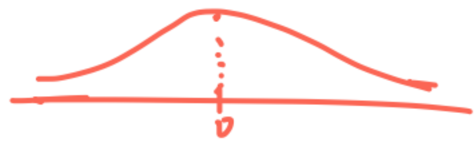

Figure 5: $\gamma_A = \int_A f d\lambda$

Notation: $\gamma = f * \lambda$

Question: Can we describe every probability measure on $(\mathbb{R}^n, B(\mathbb{R}^n))$ in terms of density?

Answer: no!

Counterexample: δ_0 Dirac measure.

On the same measure space $(\mathbb{R}^n, B(\mathbb{R}^n))$, if we have two measures λ, γ.

Question: $\gamma(A) = \int_A \varnothing d\lambda$

Does $\varnothing$ exist?

Answer: No!

Definition 1. *A probability measure on γ on $(\mathbb{R}^n, B(\mathbb{R}^n))$ is called absolutely continuous with respect to another measure μ on $(\mathbb{R}^n, B(\mathbb{R}^n))$ if every μ-null set is also a γ-null set*

$$\forall B \in B(\mathbb{R}^n) : \mu(B) = 0 \implies \gamma(B) = 0.$$

Notation: $\gamma \ll \mu$

$$\mu(A) = 0 \implies \int_A f d\mu \equiv \gamma(A) = 0$$

Example: $N(0,1) \ll \lambda$

$$\gamma_A = \int_A f d\lambda \ 5$$

Example: $\delta_0 \not\ll \lambda$ because

$$\lambda(0) = 0 \text{ but } \delta_0(0) = 1$$

Theorem 2. *(Radon-Nikodym): Consider two probability measures* γ, μ *on* $(\mathbb{R}^n.B(\mathbb{R}^n))$*. Then the following two statements are equivalent:*

(1) γ *has a density w.r.t* μ*.*

(2) γ *is absolutely continuous w.r.t* μ*.*

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

Proof idea:

(1) $\implies$ (2) easy

(2) $\implies$ (1) We need to construct a density!

Consider the set G of all functions g with the following properties:

$$\circledast \begin{cases} \bullet\ g \text{ is measurable, } g \geq 0 \\ \bullet\ g * \mu \leq \gamma \text{, that is } \forall A \in B(\mathbb{R}^n : \int_A g d\mu \leq \gamma(A). \end{cases}$$

- Observe: $g = 0$ satisfies $\circledast$, so G is not empty.
- If g,h both satisfy $\circledast$, then $sup(g, h)$ satisfies $\circledast$.
- Define := $sup_{g \in G} \int g d\xi$ and construct a sequence $(g_n)_{n \in \mathbb{N}}$ such that $lim \int g_n d\mu = \xi$.
- Define "density" $f := sup g_n$.
- Now prove: f is the density that we are looking for. $\square$