



PROBABILITY THEORY

CS6140



Predrag Radivojac

KHOURY COLLEGE OF COMPUTER SCIENCES

NORTHEASTERN UNIVERSITY

Spring 2021

BRIEF INTRODUCTION

Probability theory

- branch of mathematics
- part of measure theory

Important concepts

- experiment (coin toss, roll of dice, ...)
- outcome (one of predefined options)

A way to formalize this

- define sample space, event space
- introduce P : assignment of numbers in $[0, 1]$ to groups of outcomes.

AXIOMS OF PROBABILITY

Ω = sample space, all outcomes of the experiment

\mathcal{A} = event space, set of subsets of Ω

Ω and \mathcal{A} must be non-empty

If the following conditions hold:

$$1. A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$$

$$2. A_1, A_2, \dots \in \mathcal{A} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$$

\mathcal{A} is called a sigma field (sigma algebra)

(Ω, \mathcal{A}) = a measurable space

EXERCISE: SIGMA ALGEBRA

Ω = non-empty set

\mathcal{A} = non-empty set of subsets of Ω

$$1. A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$$

$$2. A_1, A_2, \dots \in \mathcal{A} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$$

Example:

$$\Omega = \mathbb{R};$$

Let \mathcal{A} contain \emptyset , \mathbb{R} and all sets $(-\infty, a]$, $(a, b]$, (b, ∞) , for all $a, b \in \Omega$.

Is (Ω, \mathcal{A}) a measurable space?

$$\lim_{i \rightarrow \infty} \left(0, \frac{i-1}{i} \right] = (0, 1) \neq \mathcal{A}$$

$$i \in \{2, 3, \dots\}$$

AXIOMS OF PROBABILITY

(Ω, \mathcal{A}) = a measurable space

Any function $P : \mathcal{A} \rightarrow [0, 1]$ such that

1. $P(\Omega) = 1$

2. $A_1, A_2, \dots \in \mathcal{A}, A_i \cap A_j = \emptyset \forall i, j \Rightarrow P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$

is called a probability measure (probability distribution)

(Ω, \mathcal{A}, P) = a probability space

CONSEQUENCES OF THE AXIOMS OF PROBABILITY

(Ω, \mathcal{A}, P) = a probability space

1. $P(\emptyset) = 0$
2. $P(A^c) = 1 - P(A)$
3. $P(A) = \sum_{i=1}^k P(A \cap B_i)$, where $\{B_i\}_{i=1}^k$ is a partition of Ω
4. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

... and everything else.

EXERCISE: SMALLEST SIGMA ALGEBRA

Ω = non-empty set

\mathcal{A} = non-empty set of subsets of Ω

$$1. A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$$

$$2. A_1, A_2, \dots \in \mathcal{A} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$$

Example:

$$\Omega = \mathbb{R}$$

What is the smallest \mathcal{A} we can think of?

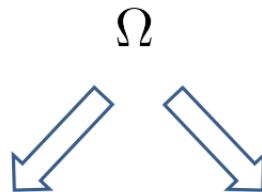
$$\mathcal{A} = \{\emptyset, \Omega\}$$

How can we choose P ?

$$P(\emptyset) = 0$$

$$P(\Omega) = 1 \quad \leftarrow \text{the only possible assignment!}$$

SAMPLE SPACES



discrete (countable)

continuous (uncountable)

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

$$\Omega = \mathbb{N}$$

$$\Omega = [0, 1]$$

$$\Omega = \mathbb{R}$$

Typically: $\mathcal{A} = \mathcal{P}(\Omega)$

Power set

Typically: $\mathcal{A} = \mathcal{B}(\Omega)$

Borel field

$$\Omega = [0, 1] \cup \{2\} = \text{mixed space}$$

EXERCISE: FINDING PROBABILITY DISTRIBUTIONS

(Ω, \mathcal{A}) = a measurable space

Example: $\Omega = \{0, 1\}$
 $\mathcal{A} = \{\emptyset, \{0\}, \{1\}, \Omega\}$

$$P(A) = \begin{cases} 1 - \alpha & A = \{0\} \\ \alpha & A = \{1\} \\ 0 & A = \emptyset \\ 1 & A = \Omega \end{cases} \quad \alpha \in [0, 1]$$

How can we choose P in practice?

Clearly, we cannot do it arbitrarily.

How can we satisfy all constraints?

PROBABILITY MASS FUNCTIONS (PMFs)

Ω = discrete sample space

$$\mathcal{A} = \mathcal{P}(\Omega)$$

Probability mass function:

$$1. \ p : \Omega \rightarrow [0, 1]$$

$$2. \ \sum_{\omega \in \Omega} p(\omega) = 1$$

The probability of any event $A \in \mathcal{A}$ is defined as

$$P(A) = \sum_{\omega \in A} p(\omega)$$

WELL-KNOWN PMFs

Bernoulli distribution: $\Omega = \{S, F\}$ $\alpha \in (0, 1)$

$$p(\omega) = \begin{cases} \alpha & \omega = S \\ 1 - \alpha & \omega = F \end{cases}$$

Alternatively, $\Omega = \{0, 1\}$

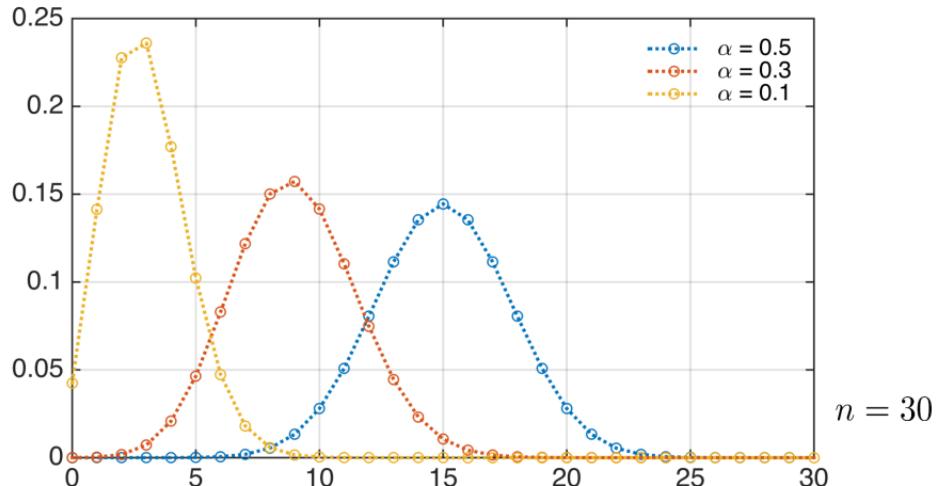
$$p(k) = \alpha^k \cdot (1 - \alpha)^{1-k} \quad \forall k \in \Omega$$

WELL-KNOWN PMFs

Binomial distribution:

$$\Omega = \{0, 1, \dots, n\} \quad \alpha \in (0, 1)$$

$$p(k) = \binom{n}{k} \alpha^k (1 - \alpha)^{n-k} \quad \forall k \in \Omega$$

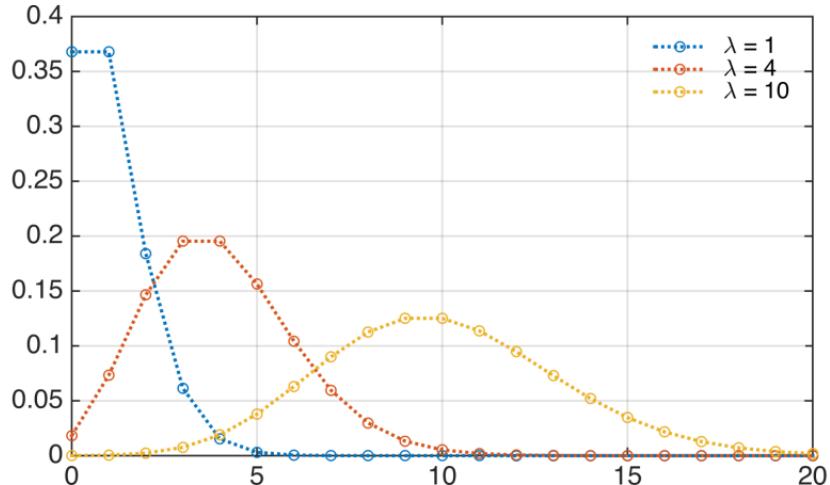


WELL-KNOWN PMFs

Poisson distribution:

$$\Omega = \{0, 1, \dots\} \quad \lambda \in (0, \infty)$$

$$p(k) = \frac{\lambda^k e^{-\lambda}}{k!} \quad \forall k \in \Omega$$

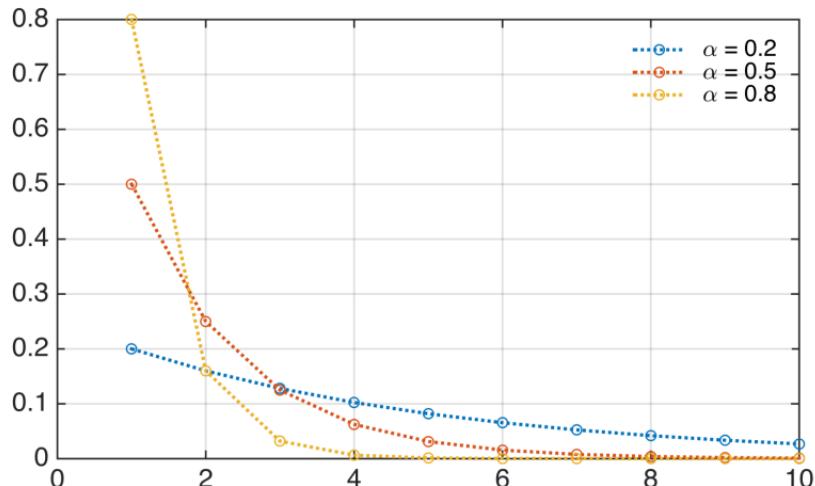


WELL-KNOWN PMFs

Geometric distribution:

$$\Omega = \{1, 2, \dots\} \quad \alpha \in (0, 1)$$

$$p(k) = (1 - \alpha)^{k-1} \alpha \quad \forall k \in \Omega$$



EXERCISE: CALCULATING PROBABILITIES OF EVENTS

$$\Omega = \{1, 2, \dots\}$$

$$\mathcal{A} = \mathcal{P}(\Omega)$$

P = induced by a geometric distribution (pmf) with parameter α

Consider the following event $A \in \mathcal{A}$:

$$A = \{k | k \text{ is odd}\}$$

$$P(A) = ?$$

PROBABILITY DENSITY FUNCTIONS (PDFs)

Ω = continuous sample space

$$\mathcal{A} = \mathcal{B}(\Omega)$$

Probability density function:

1. $p : \Omega \rightarrow [0, \infty)$

2. $\int_{\Omega} p(\omega) d\omega = 1$

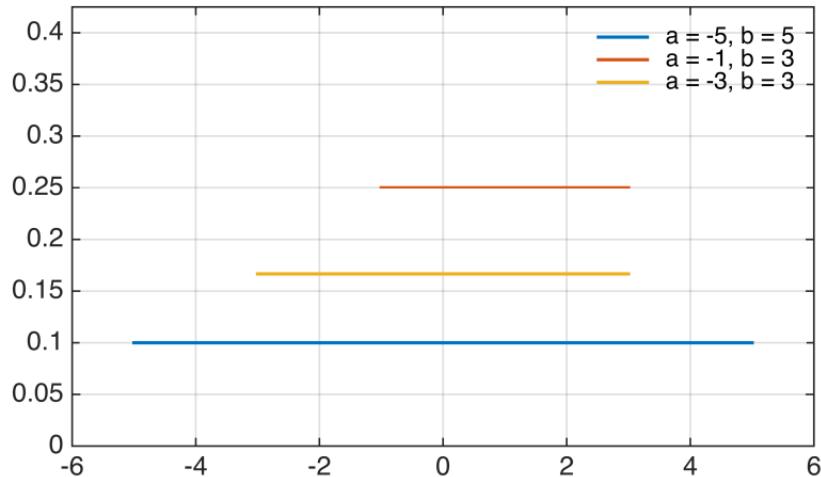
The probability of any event $A \in \mathcal{A}$ is defined as

$$P(A) = \int_A p(\omega) d\omega.$$

WELL-KNOWN PDFs

Uniform distribution: $\Omega = [a, b]$

$$p(\omega) = \frac{1}{b-a} \quad \forall \omega \in [a, b]$$

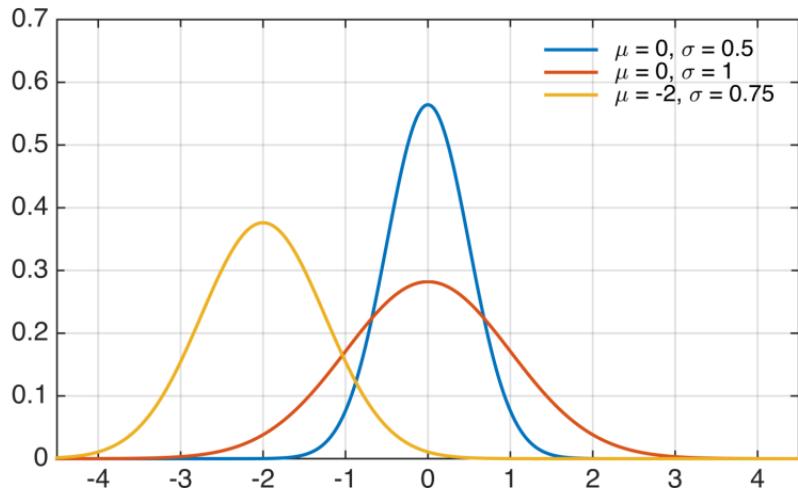


WELL-KNOWN PDFs

Gaussian distribution:

$$\Omega = \mathbb{R} \quad \mu \in \mathbb{R}, \sigma \in \mathbb{R}^+$$

$$p(\omega) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(\omega-\mu)^2} \quad \forall \omega \in \mathbb{R}$$

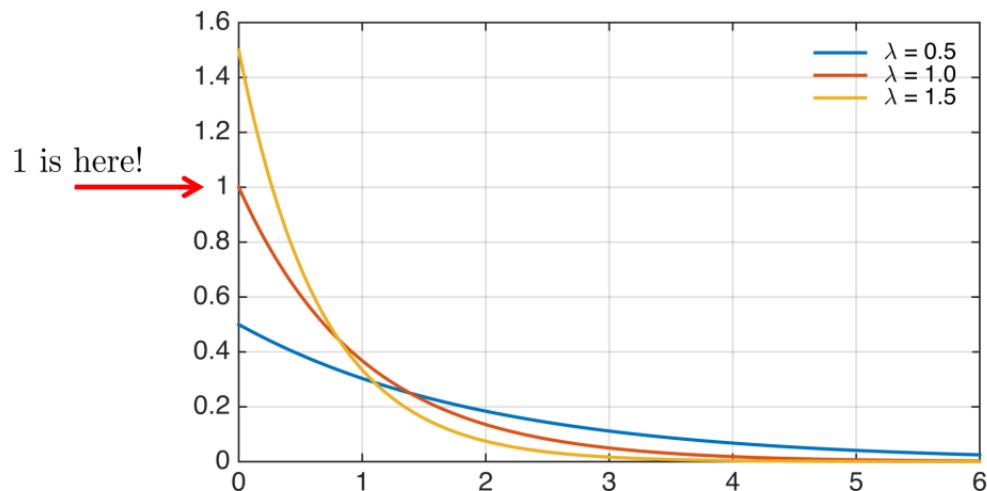


WELL-KNOWN PDFs

Exponential distribution:

$$\Omega = [0, \infty) \quad \lambda > 0$$

$$p(\omega) = \lambda e^{-\lambda\omega} \quad \forall \omega \geq 0$$



PMFs vs. PDFs

Ω = discrete sample space

Consider a singleton event $\{\omega\} \in \mathcal{A}$, where $\omega \in \Omega$

$$P(\{\omega\}) = p(\omega)$$

Ω = continuous sample space

Consider an interval event $A = [x, x + \Delta x]$, where Δ is small

$$\begin{aligned} P(A) &= \int_x^{x+\Delta x} p(\omega) d\omega \\ &\approx p(x)\Delta x \end{aligned}$$

MULTIDIMENSIONAL PMFs

$$\Omega = \Omega_1 \times \Omega_2 \times \dots \times \Omega_d$$

$$\mathcal{A} = \mathcal{P}(\Omega)$$

Probability mass function:

$$1. \ p : \Omega_1 \times \Omega_2 \times \dots \times \Omega_d \rightarrow [0, 1]$$

$$2. \ \sum_{\omega_1 \in \Omega_1} \cdots \sum_{\omega_d \in \Omega_d} p(\omega_1, \omega_2, \dots, \omega_d) = 1$$

The probability of any event $A \in \mathcal{A}$ is defined as

$$P(A) = \sum_{\omega \in A} p(\omega) \quad \omega = (\omega_1, \omega_2, \dots, \omega_d)$$

MULTIDIMENSIONAL PDFs

$$\Omega = \mathbb{R}^d$$

$$\mathcal{A} = \mathcal{B}(\mathbb{R})^d$$

Probability density function:

$$1. \ p : \mathbb{R}^d \rightarrow [0, \infty)$$

$$2. \ \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(\omega_1, \omega_2, \dots, \omega_d) d\omega_1 \cdots d\omega_d = 1$$

The probability of any event $A \in \mathcal{F}$ is defined as

$$P(A) = \int_{\omega \in A} p(\omega) d\omega.$$
$$\omega = (\omega_1, \omega_2, \dots, \omega_d)$$

MULTIDIMENSIONAL GAUSSIAN

$$\Omega = \mathbb{R}^d$$

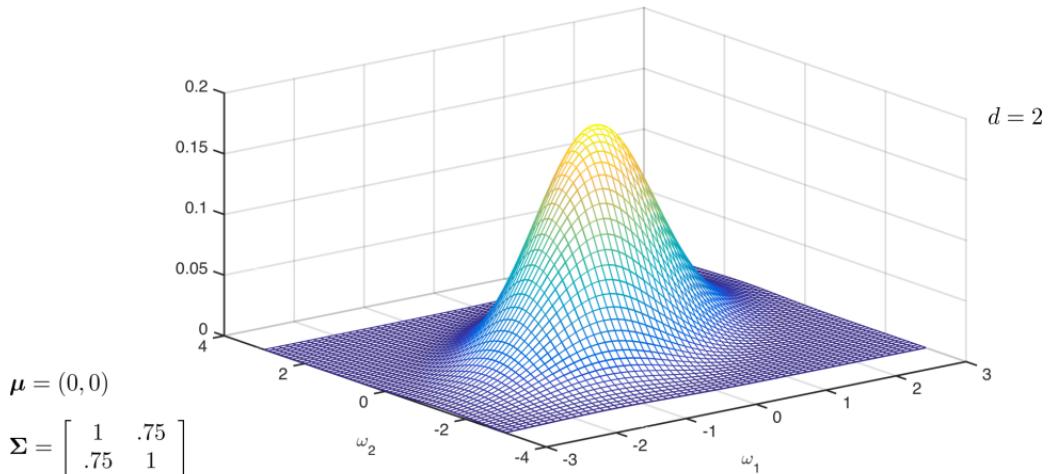
$$\mathcal{A} = \mathcal{B}(\mathbb{R})^d$$

$$\boldsymbol{\mu} \in \mathbb{R}^d$$

Σ = positive definite d -by- d matrix

$|\Sigma|$ = determinant of Σ

$$p(\boldsymbol{\omega}) = \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} \exp\left(-\frac{1}{2}(\boldsymbol{\omega} - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{\omega} - \boldsymbol{\mu})\right)$$



ELEMENTARY CONDITIONAL PROBABILITIES

(Ω, \mathcal{A}, P) = a probability space

B = event that already occurred

The probability that any event $A \in \mathcal{A}$ has also occurred is

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

where $P(B) > 0$.

Bayes' rule:

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

EXERCISE: CONDITIONAL PROBABILITIES

(Ω, \mathcal{A}, P) = probability space

A, B = events from \mathcal{A}

Derive $P(A|B)$

SUM RULE, PRODUCT RULE

(Ω, \mathcal{A}, P) = a probability space

Sum rule:

$$P(A) = \sum_{i=1}^k P(A \cap B_i)$$

where $\{B_i\}_{i=1}^k$ is a partition of Ω

Product rule:

$$P(A \cap B) = P(A|B) \cdot P(B)$$

where $P(B) > 0$

CHAIN RULE

(Ω, \mathcal{A}, P) = a probability space

Chain rule

$$P(A_1 \cap A_2 \dots \cap A_d) = P(A_1)P(A_2|A_1) \dots P(A_d|A_1 \cap A_2 \dots \cap A_{d-1})$$

where $\{A_i\}_{i=1}^d$ is a collection of d events

INDEPENDENCE OF EVENTS

(Ω, \mathcal{A}, P) = a probability space

Events A and B are **independent** if:

$$P(A \cap B) = P(A) \cdot P(B)$$

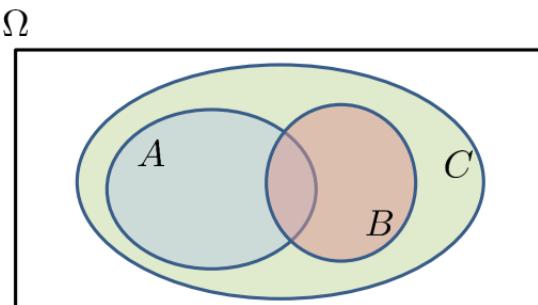
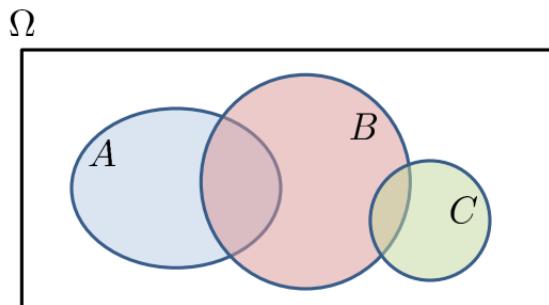
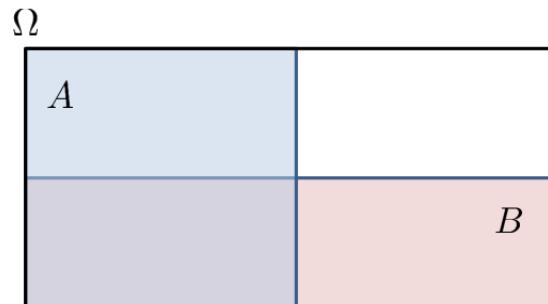
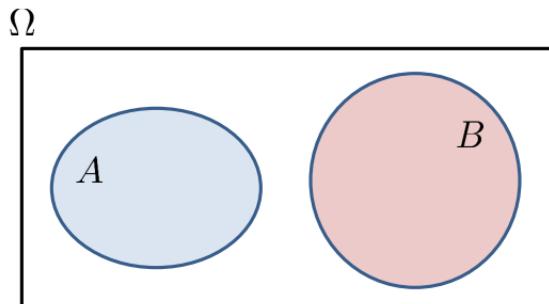
Events A and B are **conditionally independent** given C if:

$$P(A \cap B|C) = P(A|C) \cdot P(B|C)$$

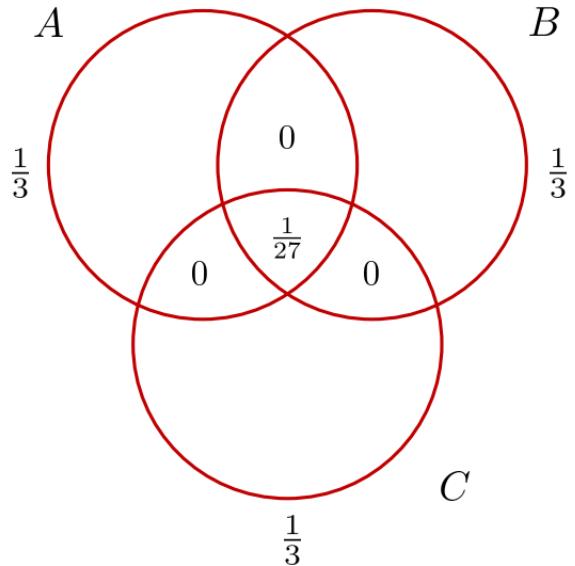
What if we had multiple events?

INDEPENDENCE EXAMPLES

(Ω, \mathcal{A}, P) = a probability space



EXERCISE: INDEPENDENCE OF EVENTS



Are A , B , and C collectively independent?