Probability Theory

CS6140

Predrag Radivojac
Khoury College of Computer Sciences
Northeastern University

Fall 2024

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

Consider non-empty Ω and $\mathcal{A}$. If the following conditions hold:

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

$\mathcal{A}$ is called a sigma field or sigma algebra.

$$(\Omega, \mathcal{A}) = \text{a measurable space}$$

Example: Sigma Algebra

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

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

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 $(\Omega, \mathcal{A})$ a measurable space?

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

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

Axioms of Probability

$(\Omega, \mathcal{A})$ = a measurable space

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

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

is called a probability measure or **probability distribution**.

$$(\Omega, \mathcal{A}, P) = \text{a probability space}$$

Consequences of the Axioms of Probability

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

1. $P(\emptyset) = 0$
2. $P(A^c) = 1 - P(A)$
3. $P(A) = \sum_{i=1}^{k} P\left(A \cap B_i\right)$, 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.

Example: Smallest Sigma Algebra

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

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

Example:

$\Omega = \mathbb{R}$

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

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

How can we choose P?

$P(\emptyset) = 0$
$P(\Omega) = 1$

$\longleftarrow$ the only possible assignment!

Sample Spaces

$$\Omega$$

discrete (countable)

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

$\Omega = \mathbb{N}$

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

Power set

continuous (uncountable)

$\Omega = [0, 1]$

$\Omega = \mathbb{R}$

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

Borel field

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

Example: Finding Probability Distributions

$(\Omega, \mathcal{A})$ = a measurable space

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

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

How can we choose P in practice?

Clearly, we cannot do it arbitrarily.

How can we satisfy all constraints?

Probability Mass Functions (PMFs)

$\Omega =$ discrete sample space
$\mathcal{A} = \mathcal{P}(\Omega)$

Probability mass function:

1. $p : \Omega \to [0, 1]$
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: $\Omega = \{S, F\} \quad \alpha \in (0,1)$

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

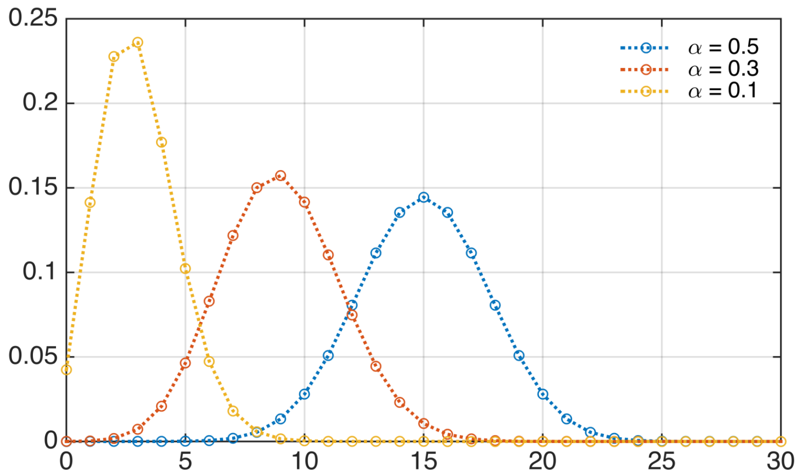
Alternatively, $\Omega = \{0, 1\}$

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

WELL-KNOWN PMFS

Binomial distribution: $\Omega = \{0, 1, \ldots, n\} \quad \alpha \in (0, 1)$

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

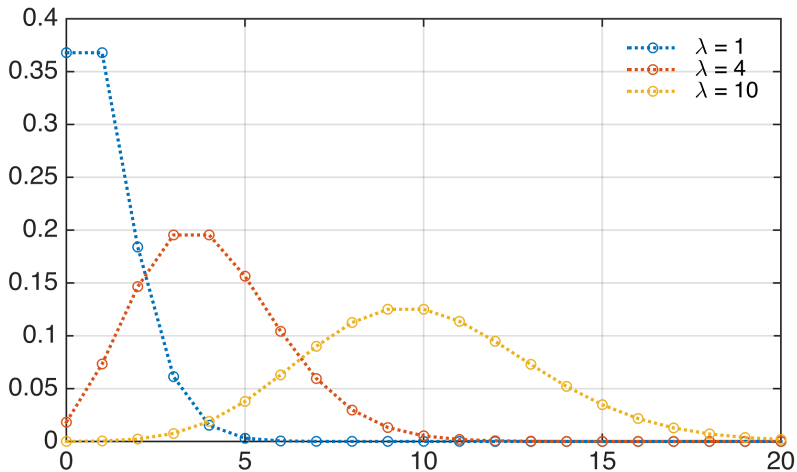


$n = 30$

Well-Known PMFs

Poisson distribution: $\Omega = \{0, 1, \ldots\}$ $\lambda \in (0, \infty)$

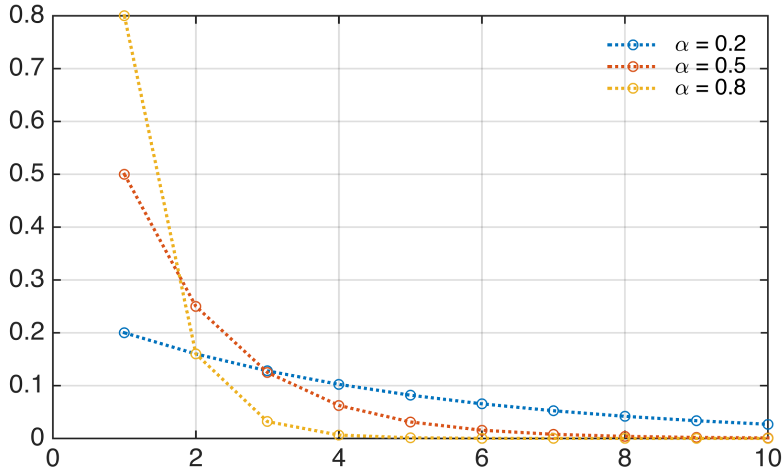
$$p(k) = \frac{\lambda^k e^{-\lambda}}{k!} \qquad \forall k \in \Omega$$



Well-Known PMFs

Geometric distribution: $\Omega = \{1, 2, \ldots\}$ $\alpha \in (0, 1)$

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



Exercise: Calculating Probabilities of Events

$\Omega = \{1, 2, \ldots\}$
$\mathcal{A} = \mathcal{P}(\Omega)$
$P =$ 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. $p : \Omega \to [0, \infty)$
2. $\int_\Omega p(\omega) d\omega = 1$

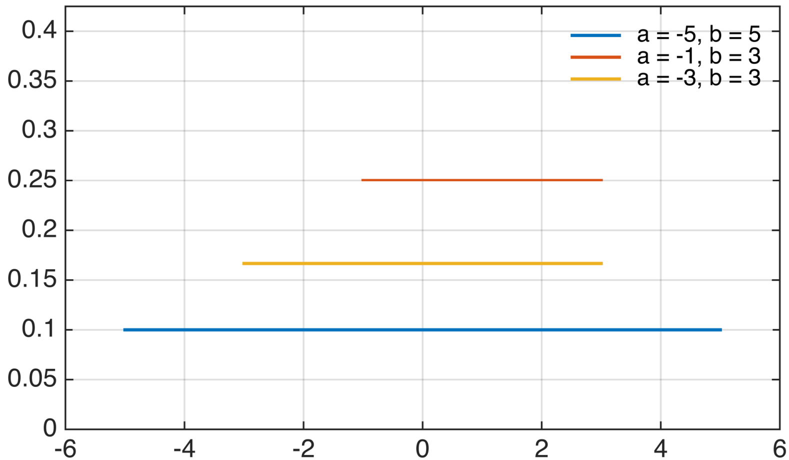
The probability of any event $A \in \mathcal{A}$ is defined as

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

Well-Known PDFs

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

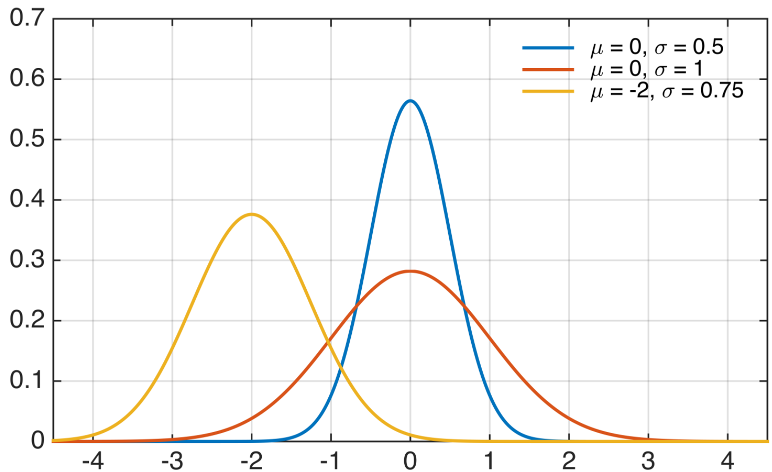
$$p(\omega) = \frac{1}{b - a} \qquad \forall \omega \in [a, b]$$



WELL-KNOWN PDFS

Gaussian distribution: $\Omega = \mathbb{R}$ $\mu \in \mathbb{R},\ \sigma \in \mathbb{R}^+$

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

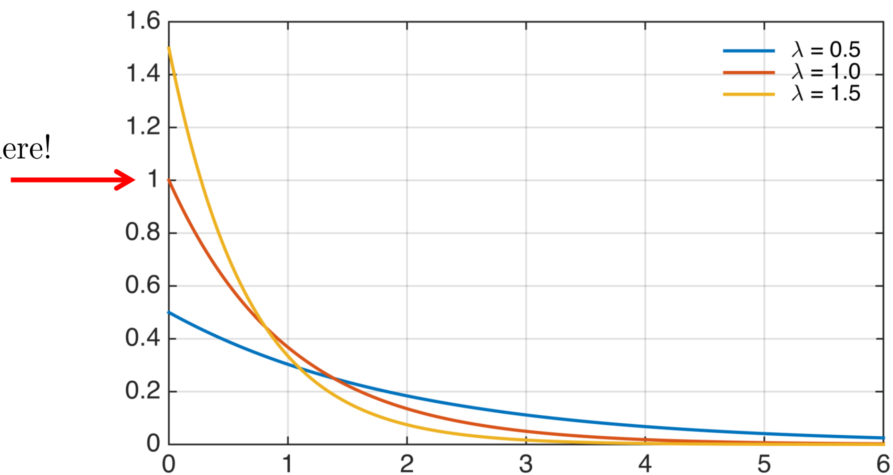


Well-Known PDFs

Exponential distribution: $\Omega = [0, \infty)$ $\lambda > 0$

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

1 is here!



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

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

Multidimensional PMFs

$\Omega = \Omega_1 \times \Omega_2 \times \ldots \times \Omega_d$
$\mathcal{A} = \mathcal{P}(\Omega)$

Probability mass function:

1. $p : \Omega_1 \times \Omega_2 \times \ldots \times \Omega_d \to [0, 1]$
2. $\sum_{\omega_1 \in \Omega_1} \cdots \sum_{\omega_d \in \Omega_d} p\left(\omega_1, \omega_2, \ldots, \omega_d\right) = 1$

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

$$P(A) = \sum_{\boldsymbol{\omega} \in A} p(\boldsymbol{\omega}) \qquad \boldsymbol{\omega} = (\omega_1, \omega_2, \ldots, \omega_d)$$

Multidimensional PDFs

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

Probability density function:

1. $p : \mathbb{R}^d \to [0, \infty)$
2. $\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(\omega_1, \omega_2, \ldots, \omega_d) d\omega_1 \cdots d\omega_d = 1$

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

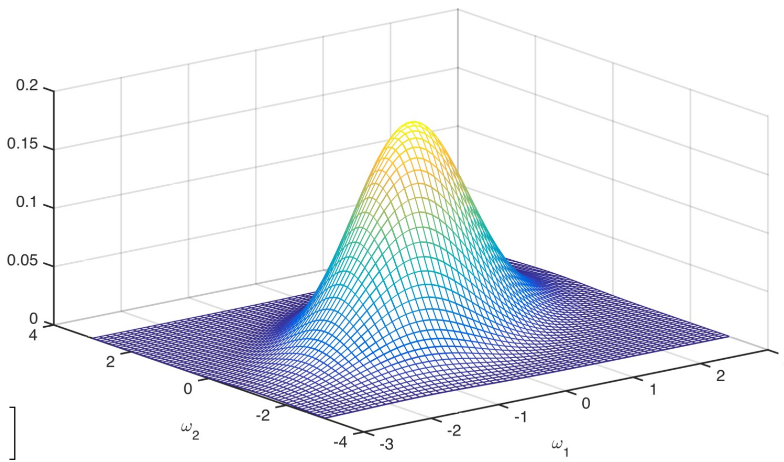
$$P(A) = \int_{\boldsymbol{\omega} \in A} p(\boldsymbol{\omega}) d\boldsymbol{\omega} \qquad \boldsymbol{\omega} = (\omega_1, \omega_2, \ldots, \omega_d)$$

Multidimensional Gaussian

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

$\boldsymbol{\mu} \in \mathbb{R}^d$
$\boldsymbol{\Sigma}$ = positive definite d-by-d matrix
$|\boldsymbol{\Sigma}|$ = determinant of $\boldsymbol{\Sigma}$

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



$d = 2$

$\boldsymbol{\mu} = (0, 0)$

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & .75 \\ .75 & 1 \end{bmatrix}$$

Elementary Conditional Probabilities

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

B = event from $\mathcal{A}$ 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

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

A, B = events from $\mathcal{A}$

Derive $P(A|B)$

Sum Rule, Product Rule

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

Sum rule:

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

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

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

Chain rule:

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

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

Independence of Events

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

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

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

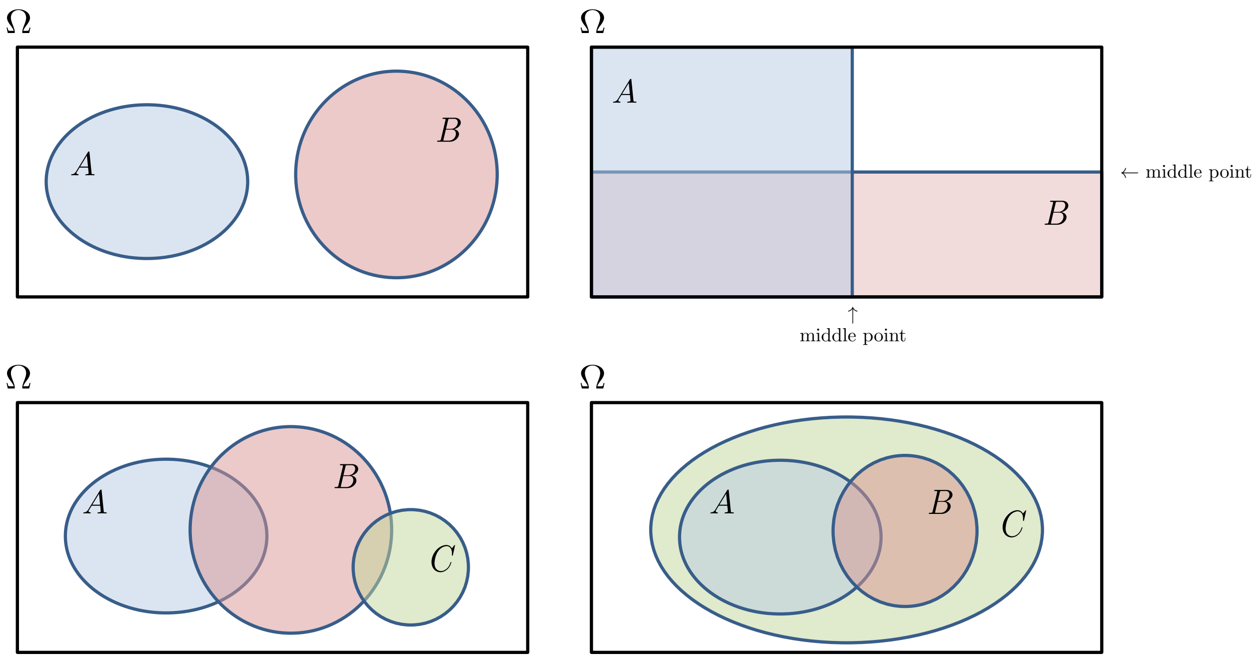
Events A and B are **conditionally independent** given C if:

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

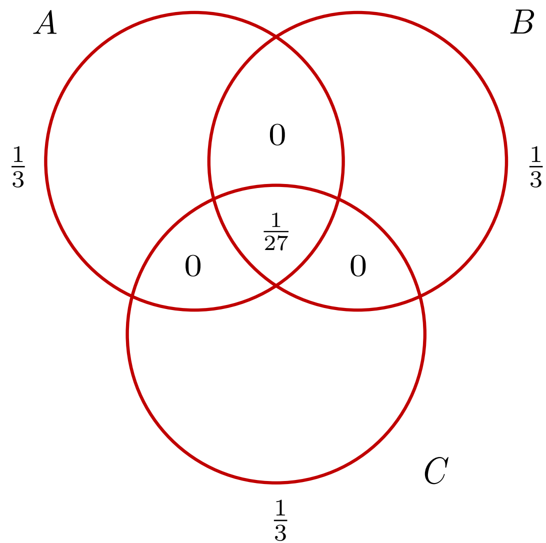
What if we had multiple events?

Exercise: Independence of Events

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



Exercise: Independence of Events



Are A, B, and C collectively independent?