LECTURE 01 24-09-2024

Information Theory for Data Science

Master of Science in Data Science and Engineering

Academic Year 2024/2025

Introduction and rules

Vers. 1.0 24/09/2024

Politecnico di Torino

Prof. Roberto Garello (roberto.garello@polito.it)

Information Theory for Data Science

- Master of Science in Data Science and Engineering
- First year, first semester
- 8 credits
- Third Edition

Course Description 1/2

- **Information Theory** studies how to quantify, represent, store, compress, transmit, protect information.

- It is a **fundamental tool for a data scientist** and has many important applications in Classifiers, Machine Learning, Artificial Intelligence, Data compression, Data transmission and Cryptography.

- The course aims to provide the **basis of Information Theory** and show some of the most important **applications for Data Science**, including Classifiers, Data Compression and Cryptography.

Course Description 2/2

- The course has a **learning-by-doing approach** where for each topic an **assignment** is proposed that must be solved by writing **Matlab** (or Python) programs and writing a **report**.

- About half of the hours are devoted to **tutoring**, where the teachers are available to help students with the assignments.

Pre-requirements

- Basic notions of probability theory
- Very basic programming skills.

Course Structure

- The course is organized into **three sections**.

Course Topics

1. Introduction to Information Theory and application to Classifiers (Garello, 25 hours)

2. Application of Information Theory to Data Compression (Taricco, 30 hours)

3. Information Theory and Cryptography. Algorithms for Cryptography (Garello, 25 hours)

Exam

- Exam 1:
  - 3 Assignments (max = 32, 70%)
  - Written exam A (4/20 questions, max = 30, 30%)

- Exam 2:
  - Written exam B (4/50 questions, max = 25)
  - optional oral exam (entire program, may increase or decrease the grade)

Exam 1: Assignments

- 3 assignments (one for each section)
- Groups of 1 to 6 students
- You must deliver:
  1. A presentation (slides, A1 and A3) or a report (A2) containing all the results, the figures and the required answers.
  2. All the written Matlab programs.
- Grade: max = 30
- Deadlines:
  - 2 weeks: +2 → max = 32
  - 3 weeks: +1 → max = 31
  - 4 weeks: +0 → max = 30
  - **Later: assignment is not accepted**
- Assignment grade = weighted average (number of hours/80) of the 3 grades (max = 32)
- 70% of exam 1 grade

Exam 1: Written exam A

- 4 questions selected from a list of approximately 20 (time available = 1.5 hours, closed-book exam, maximum grade = 30)
- 30% of exam 1 grade

Exam 1: Final grade

- 70% average assignment grade, max = 32
- 30% written exam (A), max = 30
- Max grade = 30L (for students who get at least 30.5)

Exam 2: Written exam B

- 4 questions selected from a list of approximately 50 (time available = 1.5 hours, closed-book exam, maximum grade = 25)

Exam 2: optional oral exam

- Some questions on the entire program
- May increase or decrease the grade
- If an assignment had been submitted and received a grade of X/30, the written exam grade is increased by X/30

Exam 2: final grade

- Written exam (B), max = 25
- Optional oral exam
- Points for assignments
- Max grade = 30L

Groups 1/2

- You can work in groups of 1 to 6 students
- Ideal size = 4 students. Smaller groups are discouraged, as for some assignments it is suggested to divide the workload among multiple students. Additionally, one of the course objectives is to get students used to teamwork.
- Students who are having trouble finding others to form groups can post a message on the course forum on the portal

Groups 2/2

- The group composition must be finalized within the first two weeks and **cannot be changed** afterward.

- Once the group is formed, you must elect a **representative**, which will be the student whose last name comes first alphabetically.

- This student must send an email to roberto.garello@polito.it and giorgio.taricco@polito.it, CC'ing all other members, with the group composition. Additionally, this student will be responsible for uploading the assignment solutions to the portal and sending the PDF via email.

Language

- Matlab is preferred.
- Python is accepted only as a backup option and only in the format “Jupyter Notebook Format”.

How to deliver your assignments 1/2

- An assignment is made by some exercises.

- For each assignment you must prepare:
  1. A separate Matlab program file for each exercise (not a single file for the entire assignment).
  2. A presentation (slides) or a report containing all the answers, the figures and the comments for all the exercises of that assignment. Write your name and student number on the first page of the report. Presentations/Reports must be delivered in pdf.
  3. A zip archive containing the pdf report and the program files
  4. File names: using 'SURNAME' as the last name of the representative student, name the files as SURNAME_A1.pdf and SURNAME_A1.zip

How to deliver your assignments 2/2

To deliver your assignment, the representative student must:

- Upload the zip archive on the portal (elaborati section).

- Send an email to roberto.garello@polito.it (for topics 1 and 3) or giorgio.taricco@polito.it (for topic 2) with the pdf report only (**no zip, no programs**) in attachment. Include all other members of the group in cc.

Written exams: instructions

1. **Registration for the exam is mandatory.**
2. Students taking **Exam Type B** must also send an email (at least one week before the exam) to roberto.garello@polito.it.
3. Bring only plain white sheets (such as those used in printers).
4. Bring your student ID card.
6. No cheat sheets.
7. Remember to write your name and student code on all used sheets and number them.
8. Solve each question on separate sheets.
9. At the end of the exam, you will be required to scan each used sheet and upload the PDF file to the "elaborati" section of the portal. We recommend testing this process with your phone the day before.
10. Upload two separate PDFs: one for sections 1 & 3 (named **exam0802_yoursurname_garello.pdf**) and one for section 2 (named **exam0802_yoursurname_taricco.pdf**).

Very Important 1/3

- **The presentation (or report) and the programs of an assignment can be delivered only once per academic year (you cannot change them).**

Very Important 2/3: Ethic Code

- You must solve the assignments individually or in groups or 2/3/4/5/6.
- You are not allowed to share your programs with other students/groups and/or to use part of the programs written by other students.
- We will compare the programs you deliver looking for similarities with state-of-the-art anti-plagiarism programs (check against previous years, too).
- If we conclude that an assignment is not original, trespassers of the ethic code (both the author of the original code and who partially copied the program) will be referred to the disciplinary committee.
- The committee may decide to prevent the students to take the exam during the current semester or the current academic year.

Very Important 3/3

An important rule that also applies to those who have passed the exam.

- The publication of the solution of the assignments on any websites or repositories is prohibited.
- In fact, at this university, the publication of individual exam solutions is prohibited, and the solutions of the assignments are an integral part of the exam.

Course’s organization

- Schedule:
  - Tuesday 16.00/19.00 room 7t
  - Wednesday 10.00/13.00 room 11i
  - (Rooms may change)

Material

- Notes written by the teachers, uploaded on Dropbox
- Whenever possible (unless there are portal issues, errors, or other problems), the recording of the lesson will be made available. (Tutoring is never registered.)

Tutoring

- The teacher will answer your questions on the assignments.
- Offered on the Zoom Virtual Classroom, too, where the students will be admitted on an individual basis.

Teachers

- Roberto Garello: responsible of the course, teacher for the topics of assignments 1 and 3
- Giorgio Taricco: teacher and tutor for the topics of assignment 2

Roberto Garello
roberto.garello@polito.it

Affiliation:

- Department of Electronics and Telecommunications
- Associate Professor in Communications Engineering

Research:

- Error Correcting Codes, Space Communication Systems

Teaching 2024/2025

- Information Theory for Data Science
- Communication Systems
- Applied Signal Processing Laboratory
- Space technologies for application and services

Giorgio Taricco

giorgio.taricco@polito.it

Affiliation:

- Department of Electronics and Telecommunications
- Full Professor in Communications Engineering

Research:

- Information Theory, MIMO wireless systems, space communications

Teaching 2024/2025

- Information Theory for Data Science
- Advanced wireless communications and coding

Introduction to Probability Theory

Roberto Garello

Politecnico di Torino

roberto.garello@polito.it

24/09/2024

Discrete random variable

Discrete random variable X

$$(X, \Omega_X, P(x))$$

Alphabet Ω_X = discrete set of possible outcomes

$$\Omega_X = \{x_1, ..., x_i, ..., x_M\}$$

Probability Mass Function $P(X)$ = probability of each outcome

$$P(X) = \{p(x_1), ..., p(x_i), ..., p(x_M)\}$$

$$p(x_i) = P(X = x_i) \in \mathbb{R}$$

$$0 \leq p(x_i) \leq 1 \qquad \sum_{x_i \in \Omega_X} p(x_i) = 1$$

Example

Dice

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

$$P(X) = \{p(1), p(2), p(3), p(4), p(5), p(6)\}$$

$$p(1) = ... = p(6) = \frac{1}{6}$$

Example

Coin toss

Coin: H/T $N = 3, N_T$	
HHH	0
HHT	1
HTH	1
HTT	2
THH	1
THT	2
TTH	2
TTT	3

$$X = N_T$$

$$\Omega_X = \{0, 1, 2, 3\}$$

$$P(X = i) = \frac{\binom{N}{i}}{2^N}$$

$$P(X) = \left\{ \frac{1}{8}, \frac{3}{8}, \frac{3}{8}, \frac{1}{8} \right\}$$

Binomial distribution

$$P(X = i) = \binom{N}{i} p^i (1-p)^{N-i}$$

In our example $p = 1/2$

Example

Geometric distribution

Probability of failure $= p$

Probability of success $= 1 - p$

$X =$ number of times until a success occurs

$$P(X = 1) = (1 - p)$$

$$P(X = 2) = p(1 - p)$$

$$P(X = 3) = p^2(1 - p)$$

$$\ldots$$

$$P(X = i) = p^{(i-1)}(1 - p)$$

Event

Given

$$(X, \Omega_X, P(x))$$

an **event** A is any subset of Ω_X $\qquad A \subseteq \Omega_X$

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

Example

Dice

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

$$P(X) = \left\{ p(1) = \frac{1}{6}, p(2) = \frac{1}{6}, p(3) = \frac{1}{6}, p(4) = \frac{1}{6}, p(5) = \frac{1}{6}, p(6) = \frac{1}{6} \right\}$$

$$A = \{2, 4, 6\}$$

$$P(A) = p(2) + p(4) + p(6) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$$

Special Events

Empty subset

$$A = \{\} \rightarrow P(A) = 0$$

Full subset

$$A = \Omega_X \rightarrow P(A) = 1$$

Intersection of events

$$P(A,B) = P(A \cap B) = \sum_{x_i \in A \text{ and } x_i \in B} p(x_i)$$

Example

Dice

$$A = \{2, 4, 6\}\ P(A) = 1/2 \quad B = \{1, 3, 5\}\ P(B) = 1/2$$

$$P(A \cap B) = P(\{\}) = 0 < P(A)P(B) = 1/4$$

Dice

$$A = \{2, 4, 6\}\ P(A) = 1/2 \quad B = \{4\}\ P(B) = 1/6$$

$$P(A \cap B) = P(4) = 1/6 > P(A)P(B) = 1/12$$

Dice

$$A = \{2, 4, 6\}\ P(A) = 1/2 \quad B = \{3, 6\}\ P(B) = 1/3$$

$$P(A \cap B) = P(6) = 1/6 = P(A)P(B) = 1/6$$

Union of events

$$P(A \cup B) = \sum_{x_i \in A \text{ or } x_i \in B} p(x_i) = P(A) + P(B) - P(A \cap B)$$

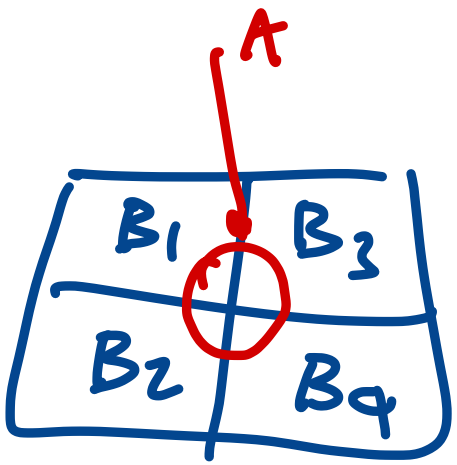
Total probability law

Ω_X decomposed as union of disjoint events

$$\Omega_X = \bigcup_i B_i$$

$$B_i \cap B_j = \Phi \ \forall i, j$$

$$P(A) = \sum_{B_i} P(A, B_i)$$



Example

Dice

$$B_1 = \{2, 4, 6\}\ B_2 = \{1, 3, 5\}$$

$$A = \{2, 3\}$$

$$P(A) = P(A, B_1) + P(A, B_2) = P(2) + P(3) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$$

Conditional probability

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

Example

Dice

$$B = \{2, 4, 6\}$$

$$A = \{4\}$$

$$P(A, B) = P(4) = \frac{1}{6}$$

$$P(B) = \frac{1}{2}$$

$$P(A|B) = \frac{P(A, B)}{P(B)} = \frac{1/6}{1/2} = \frac{1}{3}$$

Joint probability

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

$$\downarrow$$

$$P(A,B) = P(A|B)P(B)$$

$$\downarrow$$

$$P(A) = \sum_{B_i} P(A,B_i) = \sum_{B_i} P(A|B_i)P(B_i)$$

Bayes theorem

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

$P(A|B)$

$P(A,B) = P(A|B)\,P(B)$

$= P(B|A)\,P(A)$

Expectation

$$X \quad \Omega_X = \{x_1, ..., x_i, ...x_M\}$$

real function $f(X)$: $f(x_i) \in \mathbb{R}$

$$\{f(x_1), ..., f(x_i), ..., f(x_M)\}$$

$$\mathbb{E}[f(X)] = \sum_{x_i \in \Omega_X} p(x_i) f(x_i)$$

Moments

X with real outcomes

$$X \quad \Omega_X = \{x_1, ..., x_i, ...x_M\} x_i \in \mathbb{R}$$

mean value

$$\mu \equiv \mu_1 = \mathbb{E}[X] = \sum_{x_i} p(x_i) x_i$$

second order moment

$$\mu_2 = \mathbb{E}[X^2] = \sum_{x_i} p(x_i) x_i^2$$

variance

$$\sigma^2 = \mathbb{E}[(X - \mu)^2] = \mu_2 - \mu^2$$

Example

Dice

$$\mu = \mathbb{E}[X] = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3.5$$

$$\mu_2 = \mathbb{E}[X^2] = 1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + 3^2 \cdot \frac{1}{6} + 4^2 \cdot \frac{1}{6} + 5^2 \cdot \frac{1}{6} + 6^2 \cdot \frac{1}{6} = 15.17$$

$$\sigma^2 = \mu_2 - \mu^2 = 2.92$$

Joint Probability Mass Function

$$X, Y$$

$$p(x, y) = P(X = x, Y = y)$$

$$\sum_{x \in \Omega_X \times y \in \Omega_Y} p(x, y) = 1$$

Example

Weather

	Temp < 25	Temp ≥ 25
Sunny	0.4	0.2
Cloudy	0.35	0.05

Marginalization

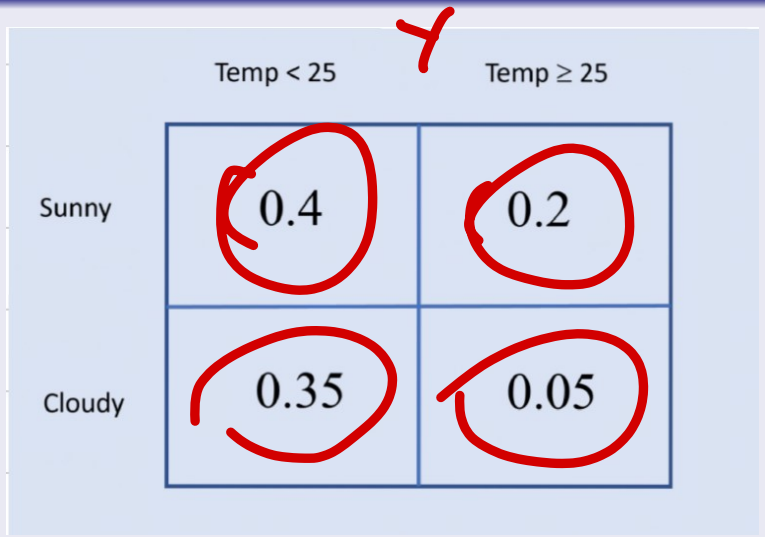
$$P(X, Y) \rightarrow P(X), P(Y)$$

$$p(x_i) = \sum_{y \in \Omega_Y} p(x_i, y)$$

$$p(y_i) = \sum_{x \in \Omega_x} p(x, y_i)$$

Example

Weather



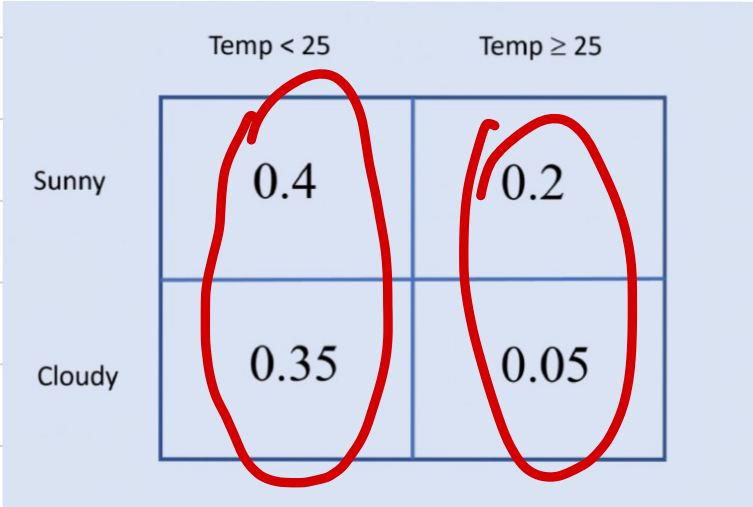
$$p(x_i) = \sum_{y \in \Omega_Y} p(x_i, y)$$

$$P(X = Sunny) = P(Sunny, Temp < 25) + P(Sunny, Temp \geq 25) = 0.4 + 0.2 = 0.6$$

$$P(X = Cloudy) = P(Cloudy, Temp < 25) + P(Cloudy, Temp \geq 25) = 0.35 + 0.05 = 0.4$$

Example

Weather



$$p(y_i) = \sum_{x \in \Omega_X} p(x, y_i)$$

$$P(Temp < 25) = P(Temp < 25, Sunny) + P(Temp < 25, Cloudy) = 0.4 + 0.35 = 0.75$$

$$P(Temp \geq 25) = P(Temp \geq 25, Sunny) + P(Temp \geq 25, Cloudy) = 0.2 + 0.05 = 0.25$$

Statistical independence

X,Y are statistically independent if and only if

$$\forall x, y \in \Omega_X \times \Omega_Y \;\; p(x, y) = p(x)p(y)$$

$$P(X, Y) = P(X)P(Y)$$

Conditional Probability Mass Function

Fix $y = y_i$

$$p(x|y_i) = P(X = x|Y = y_i) = \frac{p(x, y_i)}{p(y_i)}$$

$$\sum_{x \in \Omega_X} p(x|y_i) = 1$$

Example

Weather

	Temp < 25	Temp ≥ 25
Sunny	0.4	0.2
Cloudy	0.35	0.05

$$y_i = (Temp < 25)$$

$$P(X = Sunny | Temp < 25) = \frac{P(Sunny, Temp < 25)}{P(Temp < 25} = \frac{0.4}{0.75}$$

$$P(X = Cloudy | Temp < 25) = \frac{P(Cloudy, Temp < 25)}{P(Temp < 25} = \frac{0.35}{0.75}$$

INFORMTION CONTENT

ENTROPY

$H \geq 0$

$H = 0$

BINARY H

TERNARY H

LOG INEQUALITY

$H \leq \log_2 n$

LAGRANGE OPTIMITATION

$P_i = 1/M$

PRINCIPLE OF MAXIMUM ENTROPY

ASSIGNMENT 1

EXERCISE 1

EXERCISE 2

EXERCISE 3

INFORMATION CONTENT

$X \qquad \Omega_X = \{ x_1 \quad x_i \quad x_m \}$

$P(x) = \{ p_1 \quad p_i \quad p_m \}$

$A \subseteq \Omega_X \qquad P(A)$

WE WANT TO MEASURE AMOUNT OF INFO.
WE GAIN WHEN WE OBSERVE A

- INVERSELY PROPORTIONAL TO P(A)

- P(A) = 1 → NO INFO. CONTENT

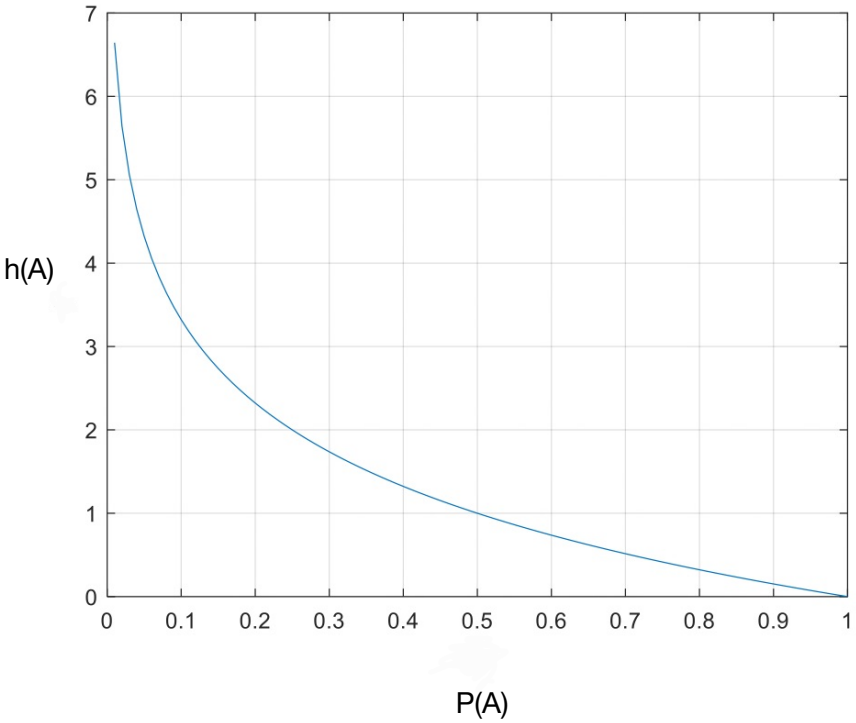
- A AND B STAT. IND → SUM OF THEIR INFO CONTENT

SHANNON'S INFO. CONTENT

$P(A)$

$$h(A) = \log_2 \frac{1}{P(A)}$$

h(A)



3 PROPERTIES

INV. PROPORTIONAL

$$P(A) = 1 \rightsquigarrow h(A) = 0$$

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

$$\log_2 \frac{1}{P(A,B)} = \log_2 \frac{1}{P(A)} + \log_2 \frac{1}{P(B)}$$

ENTROPY

X $\qquad \Omega_X = \{ x_1 \quad x_i \quad x_n \}$

$P(x) = \{ p_1 \quad p_i \quad p_n \}$

EXPECTATION OF INFO. CONTENT

$$H(X) = \sum_i p_i \, h(x_i)$$

$$= \sum_i p_i \log_2 \frac{1}{p_i}$$

$$H(x) = \sum_i p_i \log_2 \frac{1}{p_i}$$

$$= -\sum_i p_i \log_2 p_i$$

$p_i = 0$

$$\lim_{p_i \to 0} \frac{-\log_2 p_i}{1/p_i} =$$

$$= \lim_{p_i \to 0} \frac{-\log_2 e \; 1/p_i}{-1/p_i^2} = -\log_2 e \lim_{p \to 0} p_i = 0$$

LIMIT IS ZERO

$$H(x) \geq 0$$

$$H(x) = \sum_i p_i \log_2 \frac{1}{p_i}$$

$$\geq 0 \qquad \geq 0$$

$$H(x) = 0$$

$$H(x) = p_1 \log_2 \frac{1}{p_1} + p_2 \log_2 \frac{1}{p_2} + \cdots$$

$$+ p_i \log_2 \frac{1}{p_i} + \cdots$$

$$p_i = 1$$

$$p_j = 0 \qquad j \neq i$$

BINARY RANDOM VARIABLE

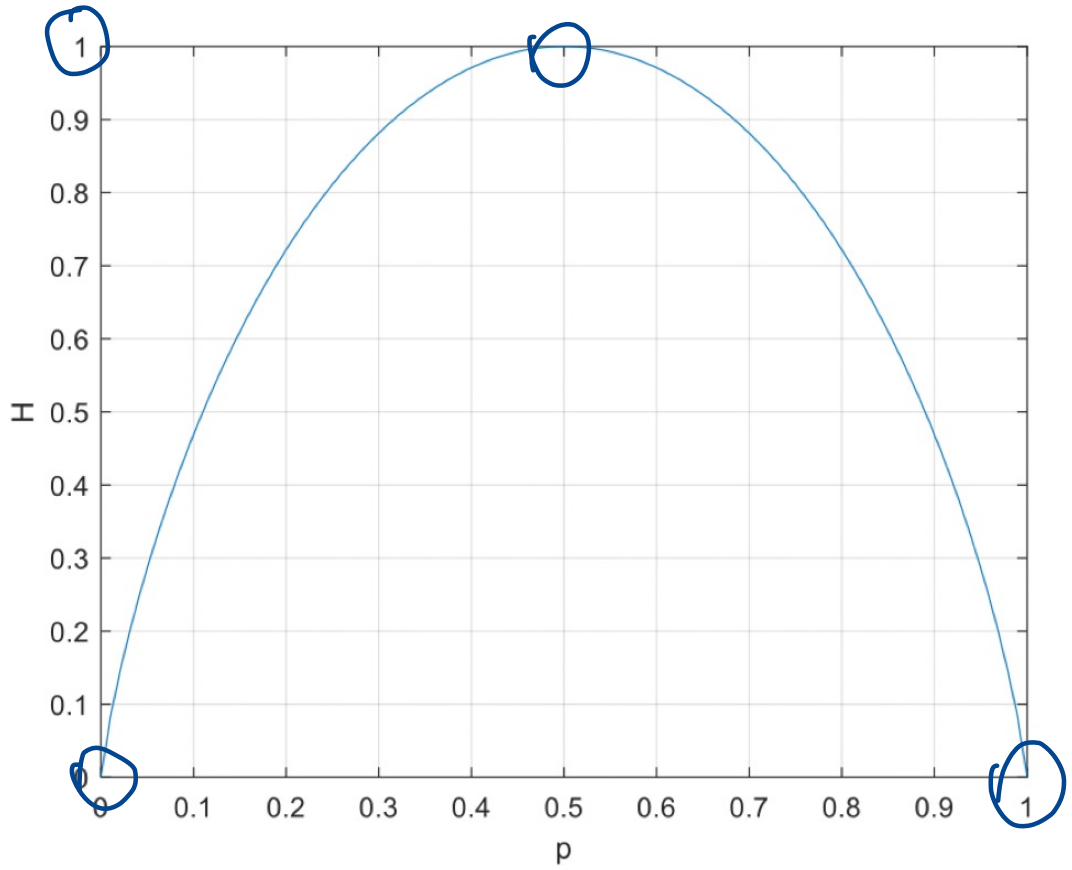
$$\Omega_X = \{ x_1 \quad x_2 \}$$

$$P(x) = \{ p_1 \quad p_2 \} \qquad p_1 + p_2 = 1$$

$$p \quad 1-p$$

$$H(x) = p_1 \log_2 \frac{1}{p_1} + p_2 \log_2 \frac{1}{p_2}$$

$$= p \log \frac{1}{p} + (1-p) \log_2 \frac{1}{1-p}$$


H
p
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

$$p = 1/2$$

$$H(x) = p \log_2 \frac{1}{p} + (1-p) \log_2 \frac{1}{1-p}$$

$$\frac{1}{2} \log_2 2 + \frac{1}{2} \log_2 2$$

$$= 1 \quad (\text{bit})$$

TERNARY RANDOM VARIABLE

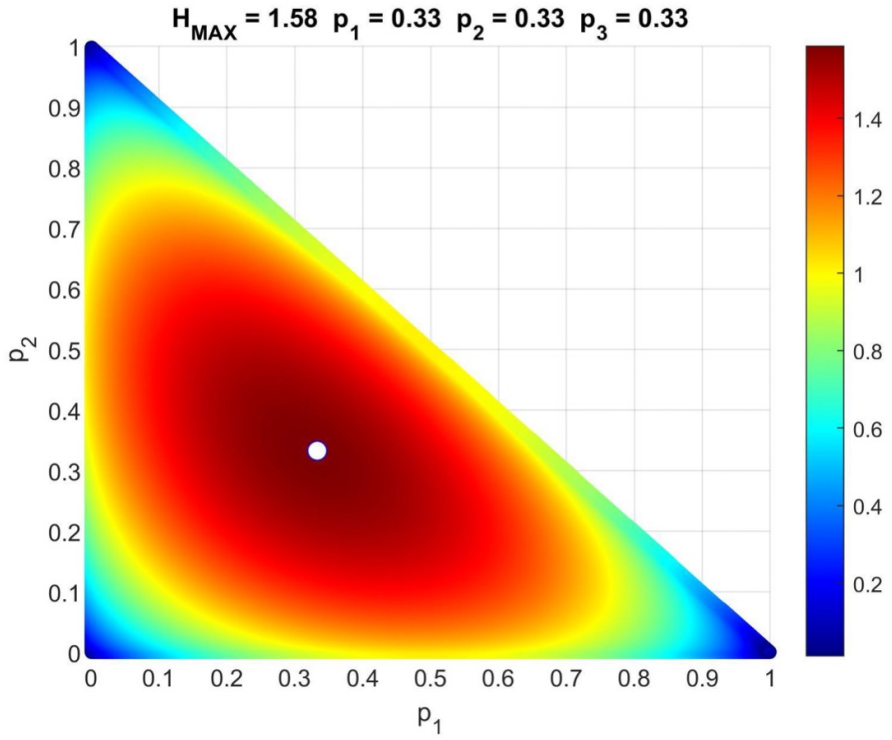
$$\Omega_X = \{ x_1 \quad x_2 \quad x_3 \}$$

$$P(x) = \{ p_1 \quad p_2 \quad p_3 \}$$

$$p_3 = 1 - p_1 - p_2$$

$$H(x) = p_1 \log_2 \frac{1}{p_1} + p_2 \log_2 \frac{1}{p_2} +$$

$$+ p_3 \log_2 \frac{1}{p_3}$$



$\frac{1}{3} \quad \frac{1}{3} \quad \frac{1}{3}$

$$H(x) = \frac{1}{3}\log_2 3 + \frac{1}{3}\log_2 3 + \frac{1}{3}\log_2 3$$

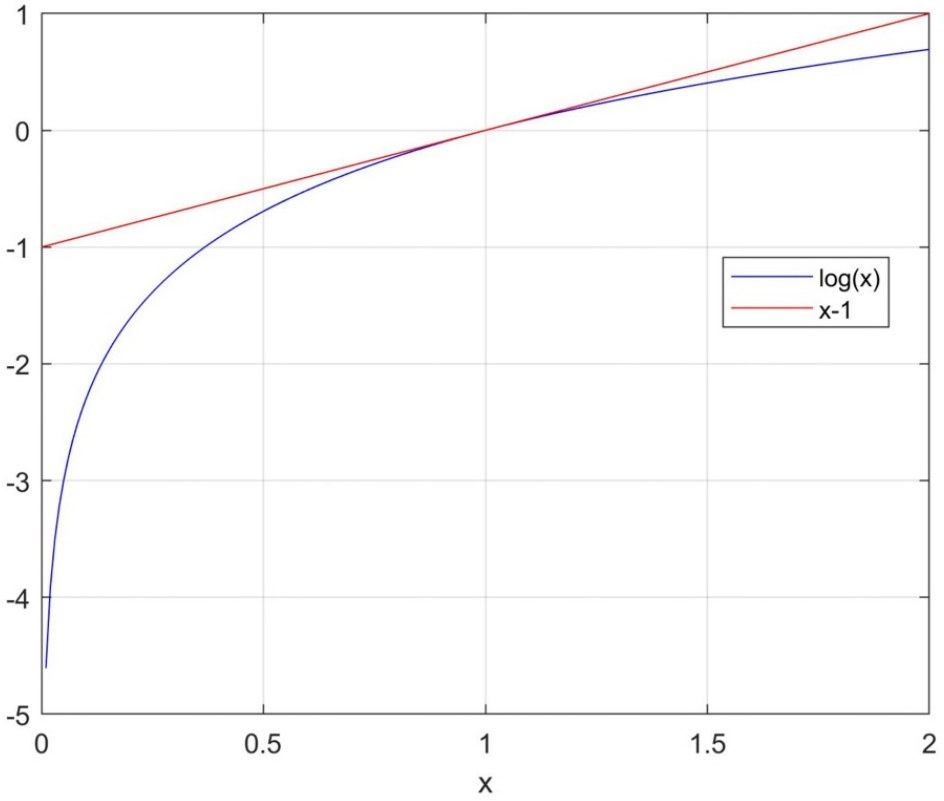
$$= \log_2 3 = 1.58$$

```
figure

colormap jet
spheresize = 30;
scatter(x,y,spheresize,h); hold on
plot(XT,YT,'bo','Markersize',8,'MarkerFaceColor','w');
colorbar
xticks([0:0.1:1])
yticks([0:0.1:1])
xlabel('p_1')
ylabel('p_2')
grid on
tit=sprintf('H_{MAX} = %.2f  p_1 = %.2f  p_2 = %.2f  p_3 = %.2f',maxx,XT,YT,ZT);
title(tit);
```

LOG INEQUALITY

$$\log_e x \leq x - 1$$



$$H(x) \leq \log_2 M$$

PROOF (ONLY FOR EXAM B)

$$H(x) - \log_2 M = \qquad \sum_i p_i = 1$$

$$= \sum_i p_i \log_2 \frac{1}{p_i} - \sum_i p_i \log_2 M$$

$$\sum_i p_i \log_2 \frac{1}{p_i M} \leq \sum_i p_i \log_2 e \left(\frac{1}{p_i M} - 1 \right)$$

$$= \log_2 e \sum_i p_i \left(\frac{1}{p_i M} - 1 \right)$$

$$\sum_i p_i \left(\frac{1}{p_i n} - 1 \right) = \sum_i \frac{1}{n} - \sum_i p_i$$

$$1 - 1 = 0$$

$$H(x) - \log_2 n \leq 0$$

$$H(x) \leq \log_2 n$$

CONSTRAINED OPTIMIZATION

FUNCTION $f(x_1, \dots, x_i, \dots, x_n)$

$g(x_1, \dots, x_i, \dots, x_n) = 0$

WE LOOK FOR MAX / MIN OF f

UNDER THE CONSTRAINT g

$$f(x_1 \quad x_i \; x_n)$$

$$g(x_1 \quad x_i \; x_n)$$

$$\Delta(x_1 \quad x_i \; x_n, \lambda_0) = f(x_1 \quad x_i \; x_n)$$

$$+ \lambda_0 g(x_1 \; x_i \; x_n)$$

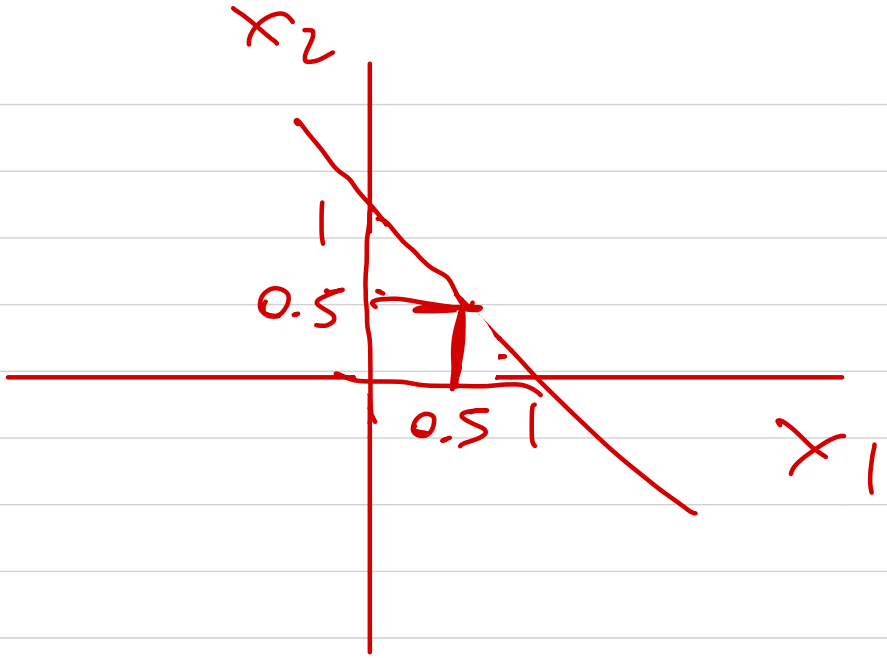
$$\Delta(x_1 \; x_i \; x_n, \lambda_0) = f(x_1 \; x_i \; x_n)$$

$$+ \lambda_0 g(x_1 \; x_i \; x_n)$$

$$\begin{cases} \dfrac{\partial \Delta}{\partial x_1} = 0 \\ \dfrac{\partial \Delta}{\partial x_i} = 0 \\ \dfrac{\partial \Delta}{\partial x_n} = 0 \\ \dfrac{\partial \Delta}{\partial \lambda_0} = 0 \end{cases}$$

$$\Delta = f + \lambda_0 g$$

$$g(x_1 \; x_i \; x_n) = 0$$



$$f(x_1, x_2) = x_1^2 + x_2^2$$

$$g(x_1, x_2) = x_1 + x_2 - 1 = 0$$

$$\Delta(x_1, x_2, \lambda_0) = (x_1^2 + x_2^2) + \lambda_0 (x_1 + x_2 - 1)$$

$$\begin{cases} \dfrac{\partial \Delta}{\partial x_1} = 2x_1 + \lambda_0 = 0 \\ \dfrac{\partial \Delta}{\partial x_2} = 2x_2 + \lambda_0 = 0 \\ \dfrac{\partial \Delta}{\partial \lambda_0} = x_1 + x_2 - 1 = 0 \end{cases} \rightarrow x_1 = x_2$$

$$x_1 = x_2 = 1/2$$

OUR PROBLEM.

$F \equiv$ ENTROPY $\quad \| \quad g$ CONSTRAINT

$$-\sum_i p_i \log_2 p_i \qquad \sum_i p_i = 1$$

$$\Delta = f + \lambda_0 g \qquad \sum_i p_i - 1 = 0$$

$$\Delta(p_1\ p_i\ p_n,\ \lambda_0)$$

$$= -\sum_i p_i \log_2 p_i + \lambda_0 \left(\sum_i p_i - 1 \right)$$

$$\Delta(p_1 \; p_i \; p_m \; \lambda_0)$$

$$= -\sum_i p_i \log_2 p_i + \lambda_0 \left(\sum_i p_i - 1 \right)$$

$$\begin{cases} \dfrac{\partial \Delta}{\partial p_1} = 0 \\ \dfrac{\partial \Delta}{\partial p_i} = -\log_2 p_i - \cancel{p_i} \log_2 e \dfrac{1}{\cancel{p_i}} + \lambda_0 = 0 \\ \qquad -\log_2 p_i + c + \lambda_0 = 0 \\ \dfrac{\partial \Delta}{\partial p_m} \\ \dfrac{\partial \Delta}{\partial \lambda_0} = 0 \qquad \sum_i p_i - 1 = 0 \qquad \sum_i p_i = 1 \end{cases}$$

$$-\log_2 p_i + c + \lambda_0 = 0$$

$$p_i = 2^{c+\lambda_0}$$

$$p_1 = \cdots = p_i = \cdots = p_M$$

$$\sum_{i=1}^{M} p_i = 1$$

$$p_i = \frac{1}{M} \quad \text{—} \quad H(x) = \log_2 M$$

ENTROPY MAXIMIZATION

$X \quad \Omega_X = (x_1 \quad x_i \quad x_n)$

$P(x) = (p_1 \quad p_i \quad p_n)$ UNKNOWN

μ KNOWN

WE WANT TO GUESS $P(x)$

WE SELECT THE PROB. DISTRIBUTION $P(x)$

THAT MAXIMIZES THE ENTROPY

$\longrightarrow$ MAX ENTROPY $\equiv$ MAX INFORMATION

WE SUPPOSE THAT

x_1 x_i x_n HAVE NUMERICAL VALUES

OR THEY ARE NUMERICAL VALUES (e.g., COSTS) ASSOCIATED TO THE OUTCOMES

PRINCIPLE OF MAXIMUM ENTROPY

WE APPLY LAGRANGE OPTIMIZATION WITH 2 CONSTRAINTS

$$-\sum_i p_i \log_2 p_i \qquad \left| \sum_i p_i - 1 = 0 \right. \qquad \left| \sum_i p_i x_i - \mu = 0 \right.$$

$$f \qquad\qquad g \qquad\qquad h$$

$$\Delta = f + \lambda_0 g + \lambda_1 h$$

$$\Delta = -\sum_i p_i \log_2 p_i + \lambda_0 \left(\sum_i p_i - 1 \right) + \lambda_1 \left(\sum_i p_i x_i - \mu \right)$$

$$\Delta = -\sum_i p_i \log_2 p_i + \lambda_0 \left(\sum_i p_i - 1 \right)$$

$$+ \lambda_1 \left(\sum_i p_i x_i - \mu \right)$$

$$\frac{\partial \Delta}{\partial p_i} = -\log_2 p_i - \underbrace{p_i \log_2 e \frac{1}{p_i}}_{c}$$

$$+ \lambda_0 + \lambda_1 x_i = 0$$

$$\log_2 p_i = c + \lambda_0 + \lambda_1 x_i$$

$$\log_2 p_i = c + \lambda_0 + \lambda_1 x_i$$

$$p_i = 2^{c+\lambda_0} \cdot 2^{\lambda_1 x_i}$$

$$= \alpha \beta^{x_i}$$

$$p_i = \alpha \beta^{x_i}$$

$$\sum_i p_i = 1 \longrightarrow \sum_i \alpha \beta^{x_i} = 1$$

$$\alpha = \frac{1}{\sum_i \beta^{x_i}}$$

$$p_i = \frac{\beta^{x_i}}{\sum_j \beta^{x_j}}$$

$$p_i = \frac{\beta^{x_i}}{\sum_j \beta^{x_j}}$$

$$\sum_i p_i x_i = \mu$$

$$\sum_i \frac{\beta^{x_i} \cdot x_i}{\sum_j \beta^{x_j}} = \mu$$

$$\sum_i \left[\beta^{x_i} \cdot x_i\right] = \mu \sum_j \beta^{x_j}$$

$$\sum_i \left[\beta^{x_i} \cdot x_i \right] = \mu \sum_i \beta^{x_i}$$

$n = 2$

$$\beta^{x_1} \cdot x_1 + \beta^{x_2} \cdot x_2 = \mu \left(\beta^{x_1} + \beta^{x_2} \right)$$

$x_1 = 0 \quad x_2 = 1 \qquad \mu = 1/2$

$$0 + \beta = \mu(1 + \beta)$$

$$\beta = \frac{\mu}{1 - \mu} = \frac{1/2}{1/2} = 1$$

$$p_i = \frac{\beta^{x_i}}{\sum_i \beta^{x_i}} = \frac{1}{2}$$

$p_1 = 1/2$ $\qquad$ $p_2 = 1/2$

Information Theory for Data Science

Assignment 1

Introduction to Information Theory and application to Classifiers

Draft version 0.1

Exercises:

1. Entropy of a random variable with 3 outcomes (pt. X)
2. Entropy of a random variable from a data series (pt. X)
3. Application of the principle of maximum entropy (pt. X)

4. …

Exercise 1 - Entropy of a random variable with 3 outcomes

1. Given a random variable with 3 outcomes, write a program to plot the entropy as a function of all possible probability vectors
2. Start with a probability vector where one of the elements is significantly higher than the others. Apply an iterative averaging procedure (for example, replace each element with the average of itself and its neighbors, followed by normalization). For each updated vector, compute the entropy and plot its value on the figure generated in step 1. Show that as the probability distribution approaches the uniform distribution, the entropy approaches its maximum value. Finally, discuss the results.

$$(p_1 \quad p_2 \quad p_3)$$

$$\left(\underbrace{\frac{p_1+p_2}{2}} \quad \underbrace{\frac{p_1+p_2+p_3}{2}} \quad \underbrace{\frac{p_2+p_3}{2}}\right)$$

Exercise 2 - Entropy of a random variable from a data series

1. Identify a data series and estimate the probabilities of the outcomes based on their occurrences, updating the probabilities at each time step.
2. At each time step, compute the entropy, plot its behavior, and discuss the results

Note: In the presentation, include a link to the source of the data series

Exercise 3 - Application of the principle of maximum entropy

Exercise 2.a

1. Invent an exercise where you have a random variable X with an alphabet Ω_X with 2 outcomes with integer values.
2. Show some examples of the the probability distribution P(X) for different values of the mean value.
3. Discuss the results

Exercise 2.b

1. Invent an exercise where you have a random variable X with alphabet Ω_X with at least 4 outcomes, where each outcome has an inteher value (“cost”).
2. Fix the mean value bigger than the arithmetic average of the costs, and apply the principle of maximum entropy to find the probability distribution P(X)
3. Plot P(X)
4. Repeat with a mean value equal to the arithmetic average and plot the result
5. Repeat with other values of the mean value and plot the results
6. Comment the results

You must numerically solve the equation generated by the Lagrange optimization.

As an example , for Matlab you can use

```
syms x
eqn = ( . . .  )*mu == ( . . . );
V = vpasolve(eqn,x,[0 10])
```

Important

Final version assigned on XX/10/2024

Delivery by

- XX/10/2024, 11.59 PM: +2 points
- XX/XX/2024, 11.59 PM: +1 point
- XX/XX/2024, 11.59 PM: 0 points
- **Later: not accepted**