Lecture 01

Roberto Garello

Politecnico di Torino

roberto.garello@polito.it

30/09/2023

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

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

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

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

$$\dots$$

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

Event

Given

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

an **event** A is any subset of Ω_X

$A \subseteq \Omega_X$

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

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

Φ

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

Full subset

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

Intersection of events

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

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

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

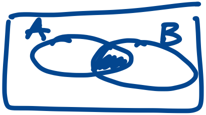
$$P(A \cap B) = P(4) = 1/6 > P(A)P(B) = 1/12$$

Dice

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

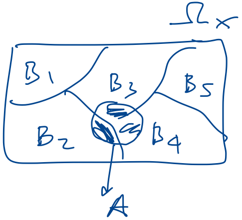
$$P(A \cap B) = P(6) = 1/6 = P(A)P(B) = 1/6$$

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

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

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

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

Expectation

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

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

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

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

$$p_1 \underline{f(x_1)} + p_2 \underline{f(x_2)} + p_3 \underline{f(x_3)} + \cdots$$

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

Y

X

	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

Y

	Temp < 25	Temp ≥ 25	
Sunny	0.4	0.2	0.6
Cloudy	0.35	0.05	0.4
			1.0

X

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

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

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

Example

Weather

Y

X	Temp < 25	Temp ≥ 25
Sunny	0.4	0.2
Cloudy	0.35	0.05
	0.75	0.25

= 1.0

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

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

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

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

X

0.75

$$y_i = (Temp < 25)$$

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

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

INFORMATION CONTENT

ENTROPY

$H \geq 0$

$H = 0$

BINARY H

TERNARY H

LOG INEQUALITY

$H \leq \log_2 n$

LAGRANGE OPTIMIZATION

$p_i = 1/M$

PRINCIPLE OF MAXIMUM ENTROPY

EXERCISE 1

EXERCISE 2

INFORMATION CONTENT

$X \quad \Omega_x \quad P(x)$

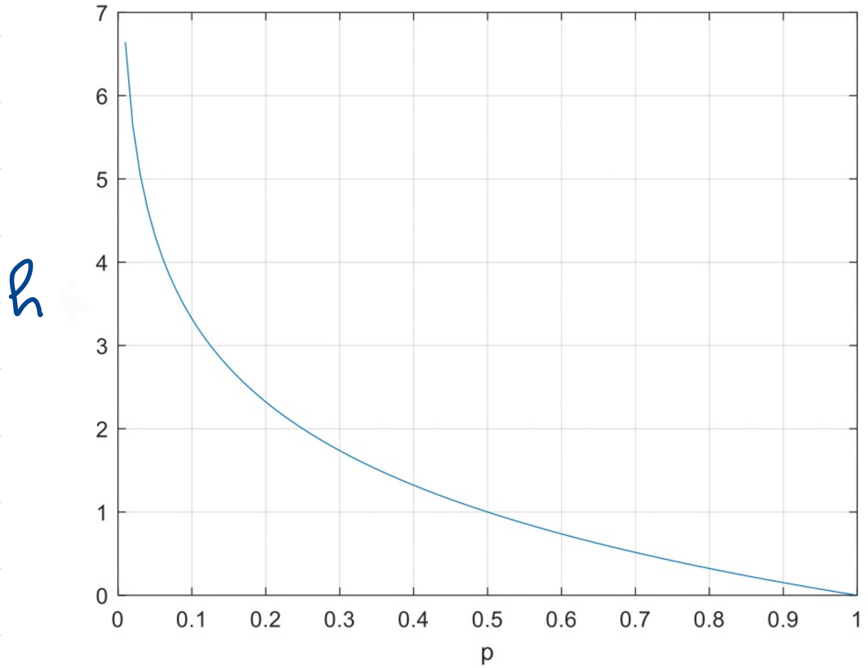
$A \subseteq \Omega_x \quad P(A)$

① INVERSELY PROPORTIONAL TO $P(A)$

② IF $P(A) = 1 \rightarrow$ ZERO INFO. CONTENT

③ INFO. CONTENT OF 2 EVENTS STATISTICALLY INDEPENDENT IS THE SUM OF THE TWO

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



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

$$p(A, B) = p(A)\, p(B)$$

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

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

ENTROPY

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

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

$$H(X) = E[h(X)] =$$

$$\boxed{= \sum_{x_i \in \Omega_X} p_i \log_2 \frac{1}{p_i}}$$

$$p_i = 0 \quad ?$$

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

$$0 \cdot \infty$$

$$\lim_{p \to 0} -p \log_2 p =$$

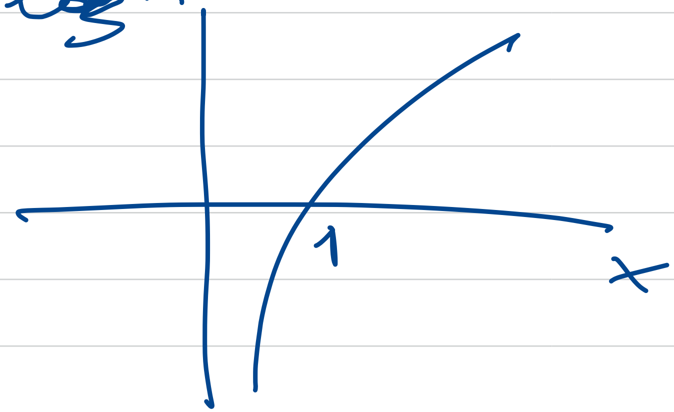
$$= \lim_{p \to 0} \frac{\log_2 p}{1/p} = \lim_{p \to 0} \frac{\log_2 e \; 1/p}{-1/p^2} = \lim_{p \to 0} -p \log_2 e = 0$$

$H(x) \geq 0$

$$H(x) = \sum_{x_i} \underbrace{p_i}_{\geq 0} \underbrace{\log_2 \frac{1}{p_i}}_{\geq 0}$$

$0 \leq p_i \leq 1$

$\log x$



$H(x) \geq 0$

$$H(x) = 0 \quad ?$$

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

$$\underbrace{p_1 \log_2 \frac{1}{p_1}}_{=0} + \underbrace{p_2 \log_2 \frac{1}{p_2}}_{=0} + \cdots$$

$$p_i \log_2 \frac{1}{p_i} = 0 \quad \begin{cases} p_i = 0 \\ p_i = 1 \end{cases}$$

$$\sum p_i = 1$$

$H(X) = 0$ IFF $\exists i$ $p_i = 1$

$j \neq i$ $p_j = 0$

(THERE IS A CERTAIN EVENT

$\longrightarrow$ NO INFO.)

IN ALL THE OTHER CASES

$H(X) > 0$

BINARY RANDOM VARIABLE

$X \qquad \Omega = \{ x_0 \quad x_1 \}$

$\{ p_0 \quad p_1 \} \qquad p_0 + p_1 = 1$

$\{ p, \quad 1-p \}$

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

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

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

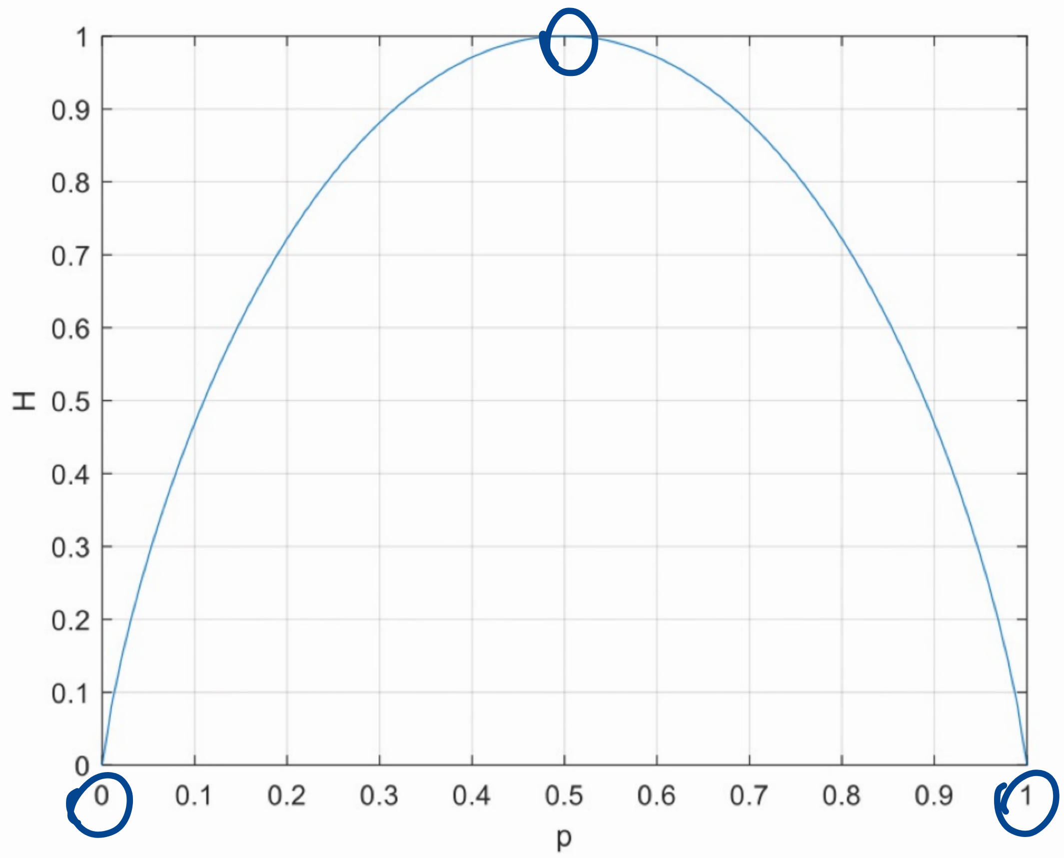
$p = 1/2$

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

$$= 1 \quad [\text{bit}]$$

$p = 0$ $\quad H = 0$

$p = 1$


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

TERNARY RANDOM VARIABLE

$X \qquad \Omega_x = \{ x_1 \quad x_2 \quad x_3 \}$

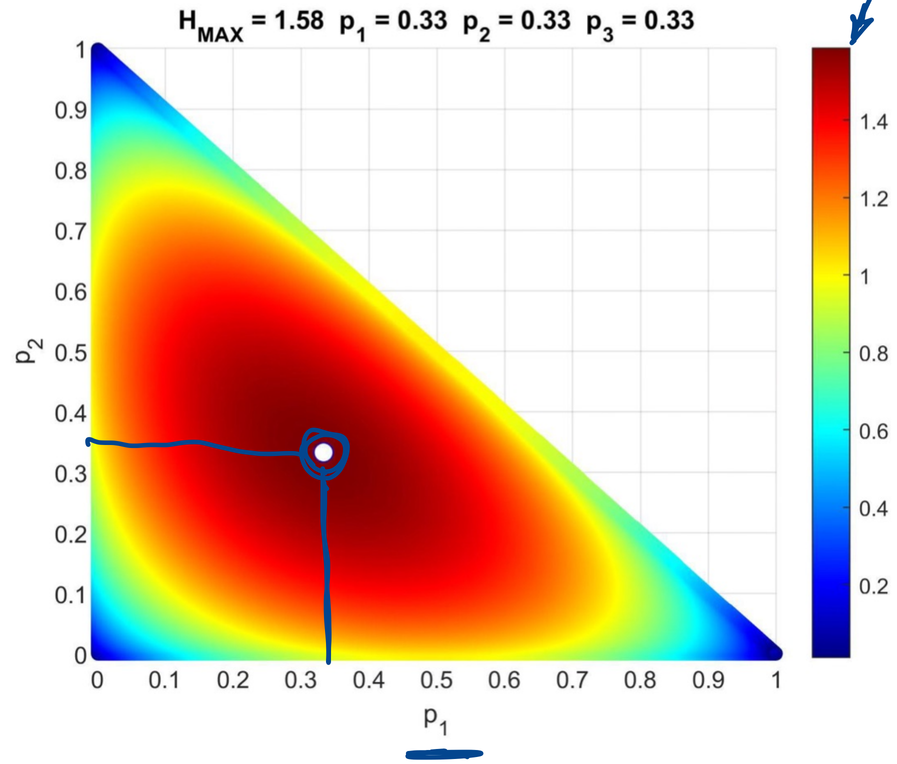
$$P(x) = \{ p_1 , p_2 , p_3 \}$$

$p_1 + p_2 + p_3 = 1$

$$H(x) = p_1 \log_2 \frac{1}{p_1} + p_2 \log_2 \frac{1}{p_2} + p_3 \log_2 \frac{1}{p_3}$$

$p_1 = p_2 = p_3$

$$= 3 \cdot \frac{1}{3} \log_2 \frac{1}{1/3} = \log_2 3 = 1.58$$



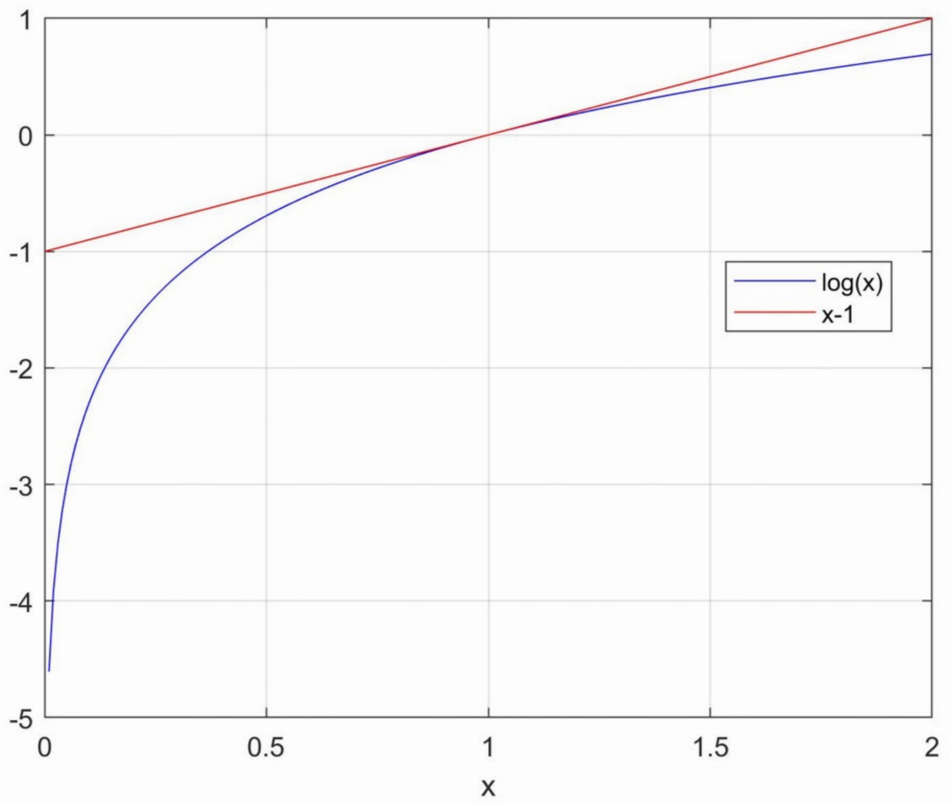
$H_{MAX} = 1.58$ $p_1 = 0.33$ $p_2 = 0.33$ $p_3 = 0.33$
p_2
p_1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1.4
1.2
1
0.8
0.6
0.4
0.2

```
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

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



$$H(X) \le \log_2 M$$

ALPHABET CARDINALITY

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

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

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

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

$$H(X) - \log_2 M$$

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

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

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

$$= \Big[\underbrace{\underline{\sum_i \frac{1}{M}} - \sum_i p_i}_{1 - 1 \,=\, 0} \Big] \log_2 e$$

$$H(X) - \log_2 M \leq 0 \rightarrow H(X) \leq \log_2 M$$

CONSTRAINED OPTIMIZATION

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

UNDER CONSTRAINT

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

$$\Delta(x_1 \quad x_i \quad x_n \quad \lambda_0) = f + \lambda_0 g$$

$$\Delta(x_1 \quad x_i \quad x_n \quad \lambda_0) = f + \lambda_0 g$$

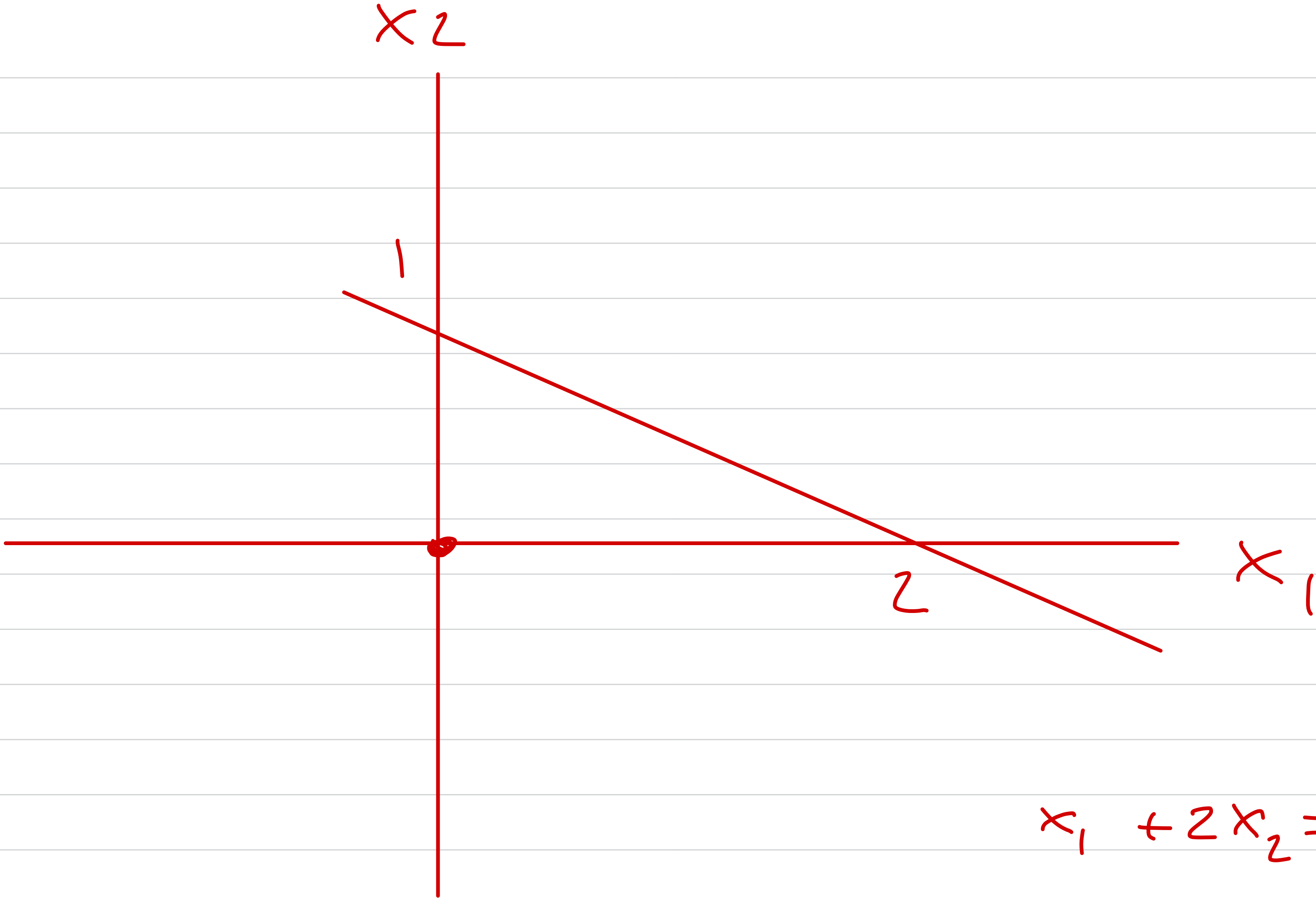
$$\frac{\partial \Delta}{\partial x_1} = 0$$

$$\vdots$$

$$\frac{\partial \Delta}{\partial x_n} = 0$$

A MAXIMUM FOR Δ IS A MAXIMUM FOR F UNDER CONSTRAINT $g = 0$

$$\rightarrow \frac{\partial \Delta}{\partial \lambda_0} = 0 \qquad g = 0$$



$x_1 + 2x_2 = 2$

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

$g(x_1, x_2) = x_1 + 2x_2 - 2 = 0$

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

$$g(x_1, x_2) = x_1 + 2x_2 - 2 = 0$$

$$\Delta(x_1, x_2, \lambda_0) = f + \lambda_0 g$$

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

$$\frac{\partial \Delta}{\partial x_1} = 2x_1 + \lambda_0 = 0$$

$$\frac{\partial \Delta}{\partial x_2} = 2x_2 + 2\lambda_0 = 0$$

$$\frac{\partial \Delta}{\partial \lambda_0} = x_1 + 2x_2 - 2 = 0$$

$$2x_1 + \lambda_0 = 0 \qquad 4x_1 + 2\lambda_0 = 0$$

$$2x_2 + 2\lambda_0 = 0 \qquad 2x_2 - 4x_1 = 0$$

$$x_1 + 2x_2 - 2 = 0 \qquad x_1 + 4x_1 = 2$$

$$x_1 = 2/5$$

$$x_2 = 4/5$$

OUR PROBLEM

F $\quad H(X) = \sum_i p_i \log_2 \frac{1}{p_i}$

g $\quad \sum_i p_i = 1 \qquad \sum_i p_i - 1 = 0$

$$\mathcal{L}(p_1 \; p_i \; p_n \; \lambda_0) = -\sum_i p_i \log_2 p_i$$

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

$$\Delta(p_1 \quad p_i \quad p_n \quad \lambda_0) = -\sum_i p_i \log_2 p_i$$

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

$$\frac{\partial \Delta}{\partial p_1} = 0$$

$\vdots$

$$\frac{\partial \Delta}{\partial p_i} = 0 \qquad -\log_2 \frac{1}{p_i} - \cancel{p_i} \log_2 e \frac{1}{\cancel{p_i}} + \lambda_0 = 0$$

$\vdots$

$$\frac{\partial \Delta}{\partial p_n} = 0$$

$$\frac{\partial \Delta}{\partial \lambda_0} = 0 \qquad \left(\sum_i p_i - 1\right) = 0$$

$$-\log_2 \frac{1}{p_i} \overbrace{-\log_2 e}^{C} + \lambda_0 = 0$$

$$-\log_2 \frac{1}{p_i} + C + \lambda_0 = 0$$

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

DOES NOT DEPEND ON INDEX i

MAX $\qquad p_1 = \dots = p_i = \dots = p_n$

UNDER

CONSTRAINT $\qquad \sum_i p_i = 1$

$$p_i = \frac{1}{M}$$

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

$$= \log_2 M$$

MAXIMUM OF ENTROPY

CORRESPONDS TO MAXIMUM UNCERTAINTY

$$p_1 = \dots = p_i = \dots = p_M = \frac{1}{M}$$

AND IS EQUAL TO $H(x) = \log_2 M$

PRINCIPLE OF MAXIMUM ENTROPY

X $\underbrace{\Omega_x}_{\text{KNOWN}}$ $P(x)$ UNKNOWN

BUT WE KNOW μ

WE WANT TO GUESS $P(x)$

WE CHOOSE THE ONE THAT MAXIMIZES ENTROPY

$\to$ MAXIMUM UNCERTAINTY

$\to$ MAX INFORMATION

WE APPLY LAGRANGE OPTIMIZATION

UNDER 2 CONSTRAINTS

$$\begin{cases} \sum_i p_i = 1 \\ \mu = \sum_i p_i x_i \end{cases}$$

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

$$g \rightarrow \sum_i p_i = 1 \rightarrow \sum_i p_i - 1 = 0$$

$$h \rightarrow \sum_i p_i x_i = \mu \rightarrow \sum_i p_i x_i - \mu = 0$$

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

$$g \rightarrow \sum_i p_i = 1 \rightarrow \sum_i p_i - 1 = 0$$

$$h \rightarrow \sum_i p_i x_i = \mu \rightarrow \sum_i p_i x_i - \mu = 0$$

$$\begin{aligned} \Delta(p_1 \; p_i \; p_n) = & -\sum_i p_i \log_2 p_i \\ & + \lambda_0 \left(\sum_i p_i - 1 \right) \\ & + \lambda_1 \left(\sum_i p_i x_i - \mu \right) \end{aligned}$$

$$\Delta(p_1 \quad p_i \quad p_n) = -\sum_i p_i \log_2 p_i$$

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

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

$$\frac{\partial \Delta}{\partial p_1} = 0$$

$$\frac{\partial \Delta}{\partial p_i} = 0 \longrightarrow -\log_2 p_i - p_i \frac{1}{p_i} \log_2 e + \lambda_0$$

$$+ \lambda_1 x_i$$

$$= 0$$

$$\frac{\partial \Delta}{\partial p_n} = 0$$

$$\frac{\partial \Delta}{\partial \lambda_0} = 0 \longrightarrow \sum_i p_i - 1 = 0$$

$$\frac{\partial \Delta}{\partial \lambda_1} = 0 \longrightarrow \sum_i p_i x_i - \mu = 0$$

$$\frac{\partial \Delta}{\partial p_i} = -\log_2 p_i - \log_2 e + \lambda_0 + \lambda_1 x_i = 0$$

c

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

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

$$\underline{\alpha} \quad \underline{\beta}^{x_i}$$

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

$$\sum_i p_i = 1$$

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

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

$$\boxed{p_i = \frac{\beta^{x_i}}{\sum_i \beta^{x_i}}}$$

$$\sum_i p_i x_i = \mu$$

$$\sum_i \frac{x_i \beta^{x_i}}{\sum_i \beta^{x_i}} = \mu$$

$$\sum_i \beta^{x_i} \cdot x_i = \mu \left(\sum_i \beta^{x_i} \right)$$

$$\sum_i \beta^{x_i} \cdot x_i = \mu \left(\sum_i \beta^{x_i} \right)$$

SOLVE FOR β

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

$$\sum_i \beta^{x_i} \cdot x_i = \mu \left(\sum_i \beta^{x_i} \right)$$

$M = 2$

$$\longrightarrow \boxed{x_0 \beta^{x_0} + x_1 \beta^{x_1} = \mu \left(\beta^{x_0} + \beta^{x_1} \right)}$$

$\longrightarrow$ SOLVE FOR β

$$\Omega_x = \{ \underset{x_0}{0}, \underset{x_1}{1} \} \qquad \mu$$

$$0 \cdot \beta^0 + 1 \cdot \beta^1 = \mu \left(\beta^0 + \beta^1 \right)$$

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

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

$$\mu = 1/2$$

$$\beta = \frac{1}{2}(1 + \beta)$$

$$\beta = 1$$

$$P_i = \frac{\beta^{x_i}}{\sum_i \beta^{x_i}}$$

$$P_0 = \frac{\beta^0}{\beta^0 + \beta^1} = \frac{1}{2}$$

$$P_1 = \frac{\beta^1}{\beta^0 + \beta^1} = \frac{1}{2}$$

Information Theory for Data Science

Assignment 1

Introduction to Information Theory and application to Classifiers

Draft version 0.1

Exercises:

1. Entropy of a binary random variable with 3 outcomes (pt. X)

2. Application of the principle of maximum entropy (pt. X)

3. ….

Exercise 1 - Entropy of a binary random variable

Exercise 1.A

1. Plot the entropy of a binary random variable
2. Discuss the result

$$H(x) = -p \log_2 p - (1-p) \log_2 (1-p)$$

Exercise 1.B – Renyi entropy

$$H_\alpha(X) = \frac{1}{1-\alpha}\log_2\left[\sum_i p_i^\alpha\right] \qquad \alpha \geq 0 \qquad \alpha \neq 1$$

3. Prove that the limit for $\alpha \rightarrow 1$ is the Shannon entropy.

4. Repeat point 1 with different values of α (smaller and bigger than 1)

5. Comment the results

Exercise 1.C

- Prove that this averaging operation always increases the entropy

$$\{p, p, p_3\} \qquad p = \frac{p_1 + p_2}{2}$$

Hint: use the log inequality

$X \quad p_1 \; p_2 \; p_3 \quad H(x)$

$\leq$

$X' \quad p \; p \; p_3 \quad H(x')$

Exercise 2 - Application of the principle of maximum entropy

Exercise 2

1. Invent an exercise where you have a random variable X with alphabet Ω_X where each outcome has a given “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. Repeat with a mean value equal to the arithmetic average
4. Repeat with other values of the mean value
5. 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])
```