TN3125
Information and Computation

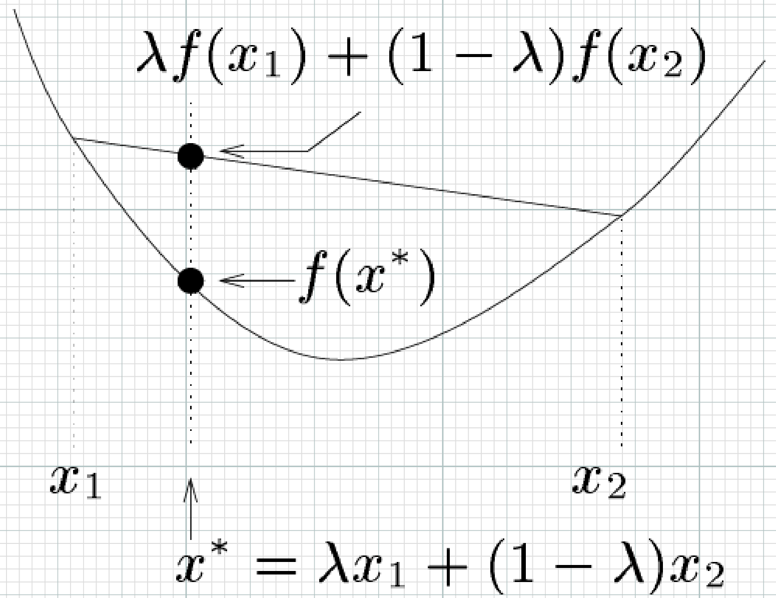
Lecture 1

3 – Properties of entropy

Concave functions

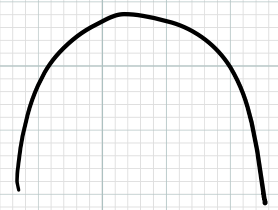
- **Definition.** A function f is concave if in the interval (a, b) if for all x_1, x_2 in the interval and $\lambda \in [0,1]$:

$$f(\lambda x_1 + (1-\lambda)x_2) \geq \lambda f(x_1) + (1-\lambda)f(x_2)$$

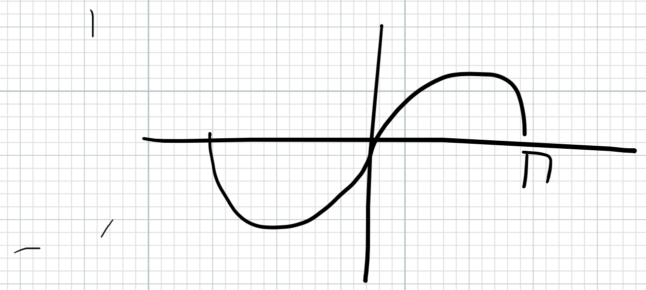


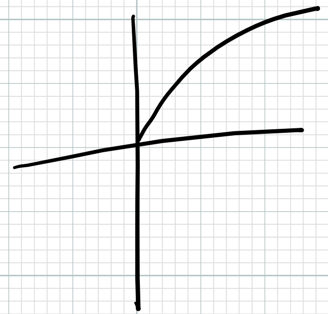
- **Lemma.** A function f is concave in the interval (a, b) if it is twice differentiable and $f'' \leq 0$ in the interval

$-x^2$

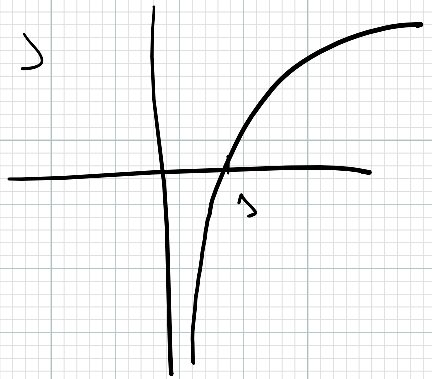


$\sin(x), \; x \in (0, \pi)$





log



Exercise

- Show that $f(x) = -x \log x$ is concave in the interval $(0,1)$

$$f'(x) = -\log x - 1$$

$$f''(x) = -\frac{1}{x}$$

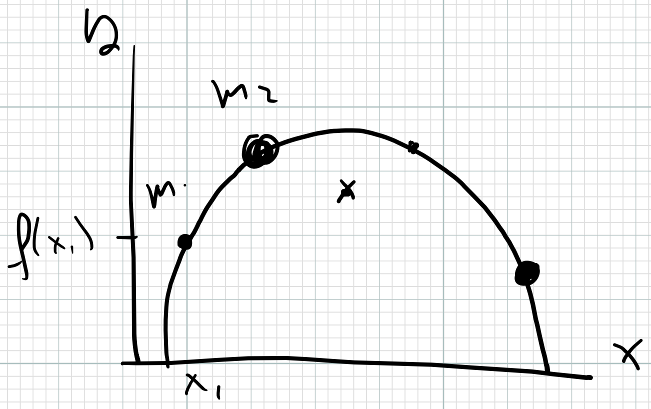
Jensen’s theorem

- Let f be a concave function in the interval (a, b). Then for any set of points $x_1, \dots, x_n$ in the interval and for $p_1, \dots, p_n$ non-negative and adding up to one:

$$f\left(\sum_{i=1}^{n} p_i x_i\right) \geq \sum_{i=1}^{n} p_i f(x_i)$$

Exercise

- Show that the center of masses of a set of masses placed over a concave curve, that is at locations $(x_i, f(x_i))$, lies below the curve.



$$c_x = \frac{\sum_i x_i \cdot m_i}{\sum_i m_i} = \sum_i x_i \frac{m_i}{M}$$

$$c_y = \sum_i f(x_i) \cdot \frac{m_i}{M}$$

$$f(c_x) = f\left(\sum_i x_i \frac{m_i}{M}\right) \geq \sum_i \frac{m_i}{M} f(x_i) = c_y$$

Exercise

Find the entropy of the uniform distribution on n elements

$$H(x) = -\sum_{i=1}^{n} \frac{1}{n} \log \frac{1}{n} = -\log \frac{1}{n} = \log n$$

Exercise

- Show that the distribution that maximizes entropy for any alphabet size is the uniform distribution. (Hint: use Jensen’s inequality and the concavity of the log function).

Exercise (concavity of entropy)

- Let X_1, X_2 be two independent random variables modelling two different but otherwise indistinguishable bent coins.
- Let Y be a binary random variable that models a process where we throw X_1 with probability t and X_2 with probability $1 - t$.
- Show that

$$H(Y) \geq tH(X_1) + (1-t)H(X_2)$$

$f(x) = -x \log x$ is concave

$$y_i = t x_i^1 + (1-t) x_i^2$$

$$H(y) = \sum_i -y_i \log y_i \geq t \sum_i -x_i^1 \log x_i^1 + (1-t) \sum_i -x_i^2 \log x_i^2$$

$$= t H(x_1) + (1-t) H(x_2)$$

Exercise (telescopic property)

- Let X_1, X_2 be two independent random variables modelling two different perfectly distinguishable bent coins.
- Let Y be a binary random variable that models a process where we throw X_1 with probability t and X_2 with probability $1-t$.
- Show that

$$H(Y) = H(t, 1-t) + tH(X_1) + (1-t)H(X_2)$$