1.5. Shannon entropy and von Neumann entropy

$\boxed{\text{Definition 1.5.1}}$

Any quantity X capable of a <u>discrete number</u> of <u>real values</u> x_i, $i \in I$ with probabilities

$$p_i^X := p(X = x_i) = p(x_i) \geq 0 \text{ such that } \sum_{i \in I} p_i^X = 1$$

is called a <u>discrete random variable</u> with probability $\Pi_X = \{p(x_i)\}_{i \in I}$

$\boxed{\text{Example 1.5.1.}}$ $\quad \#(I) := \mathrm{card}(I) = N$

$p_1^X = p(x_1) = 1 \Rightarrow X$ is deterministic (it always returns $X = x_1$)

$p_i^X = p(x_i) = \frac{1}{n} \quad \forall\, i = 1, 2, \ldots, N \Rightarrow X$ is completely random

Observation : the higher the probability $p_i = p(x_i)$, the more likely x_i is, the less informative is the aquisition of the knowledge that $X = x_i$.

Definition 1.5.2

Two random variables $(X, \{p(x_i)\}_{i\in I} = \Pi_X)$ and $(Y, \{p(y_j)\}_{j\in J} = \Pi_Y)$ are statistically independent iff the joint probabilities $p^{XY}_{ij} = p(X = x_i, Y = y_j) = p(x_i, y_j)$ factorize: $\boxed{p^{XY}_{ij} = p^X_i \, p^Y_j}$.

Namely, $\Pi_{XY} = \{p^{XY}_{ij}\}_{i\in I, j\in J} = \Pi_X \Pi_Y$

Observation : Π_X itself, as a function of X, is a random variable

Observation : any measure of the information content of a random event (one of the outcomes of X) should be a function $I : \Pi_X \longrightarrow \mathbb{R}$

such that

$$\boxed{I(p^X_i) \geq 0 \quad \forall i} \quad a)$$

$$\boxed{I(p^X_i = 1) = 0} \quad b)$$

$$\boxed{I(p^X_i p^Y_j) = I(p^X_i) + I(p^Y_j)} \quad c)$$

for statistically independent X, Y.

Remark : asking for sufficient smoothness of $I(u)$ on $[0,1]$, one gets from c), with $I'(u) = \frac{dI}{du}$,

$$\partial_{p^X_i} I(p^X_i p^Y_j) = p^Y_j \, I'(p^X_i p^Y_j) = I'(p^X_i)$$

$$\partial^2_{p^Y_j p^X_i} I(p^X_i p^Y_j) = I'(p^X_i p^Y_j) + p^X_i p^Y_j \, I''(p^X_i p^Y_j) = 0$$

$$I'(u) + u I''(u) = \frac{d}{du}\left(u I'(u)\right) = 0 \Longrightarrow u I'(u) = a \Longrightarrow \boxed{\begin{array}{l} I(u) = -\log u \\ a = -1 \end{array}}$$

Definition 1.5.3

$I(p_i^X) = -\log p_i^X$ is the information content of the event $X = x_i$ occurring with probability $p_i^X = p(x_i) = p(X = x_i)$.

Remark: each event occurring with probability p_i^X, the average information content of X is the expectation of the random variable $\{I(p_i^X)\}_{i \in I} = I(X)$ with respect to π_X.

Definition 1.5.4. (Shannon entropy)

The average information content about X, random variable with probability π_X, is called Shannon entropy and is given by

$$\boxed{H(\pi_X) := -\sum_{i \in I} p_i^X \log p_i^X = E[-\log \pi_X]} \qquad \left(H(\pi_X) = H(X)\right)$$

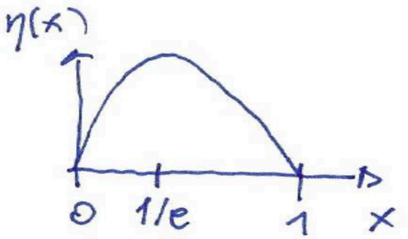
Remark: $H(X)$ measures the average information that one collects after knowing $X = x_i$ or else the average ignorance that one has about X before knowing $X = x_i$

$\boxed{\text{Example } 1.5.2.}$

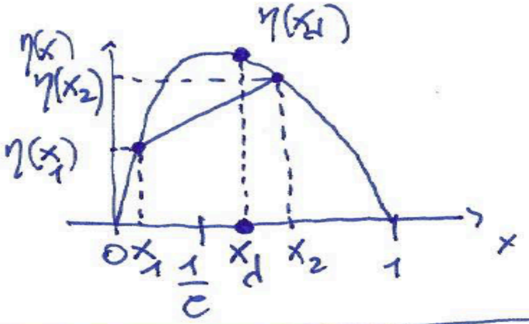
- If X is deterministic then $H(X) = 0$ by setting $\boxed{\eta(x) :\,= -x \log x}$ for $0 < x \leq 1$ and $\eta(0) = 0$
- If X is completely random and $\#(I) = N$, then $H(X) = -\sum_{i=1}^{N} \frac{1}{N} \log \frac{1}{N} = \log N$

Observation: $\boxed{\eta(x) \text{ is concave}}$

$$\eta'(x) = -(1 + \log x)$$
$$\eta''(x) = -\frac{1}{x}$$



$\boxed{\text{Concavity}}$



$x_d = d\,x_1 + (1-d)\,x_2$

$0 \leq d \leq 1$

- $\boxed{\eta(x_d) = \eta(d x_1 + (1-d) x_2) \geq d\,\eta(x_1) + (1-d)\,\eta(x_2)}$ $\quad \forall d \in [0,1]$, $\forall\, x_1, x_2 \in [0,1]$

- $\eta(x_2) - \eta(x_1) \leq \eta'(x_1)\,(x_2 - x_1) \qquad \forall\, x_1, x_2 \in [0,1]$

$$-x_2 \log x_2 + x_1 \log x_1 \leq (1 + \log x_1)(x_1 - x_2)$$

$\boxed{x_2 (\log x_2 - \log x_1) \geq x_2 - x_1}$ $\quad$ $\boxed{= \text{ holds iff } x_1 = x_2}$

$\boxed{\text{Exercise 1.5.1}}$ show that $H(X) \leq \log N$ if $\Pi_X = \{p_i^X\}_{i=1}^N$

Let $\Pi_U = \{\frac{1}{N}\}_{i=1}^N$: $H(X) - H(U) = -\sum_{i=1}^N p_i^X \log p_i^X - \log N = -\sum_{i=1}^N p_i^X \left(\log p_i^X - \log \frac{1}{N}\right)$

$$\leq -\sum_{i=1}^N \left(p_i^X - \frac{1}{N}\right) = 0$$

$\boxed{\text{Proposition 1.5.1.}}$ The Shannon entropy is

- Concave : if $\pi^{(1)}_X$ and $\pi^{(2)}_X$ are probabilities for a random variable X and $[0,1] \ni \alpha$, then

$$H\left(\alpha \pi^{(1)}_X + (1-\alpha)\pi^{(2)}_X\right) \geq \alpha H(\pi^{(1)}_X) + (1-\alpha) H(\pi^{(2)}_X)$$

- Sub-additive : $H(X,Y) \leq H(X) + H(Y)$

Proof: $\pi^{(1)}_X = \{p^X_{1i}\}_{i\in I}$, $\pi^{(2)}_X = \{p^X_{2i}\}_{i\in I}$

$$\pi_\alpha := \alpha \pi^{(1)}_X + (1-\alpha)\pi^{(2)}_X = \{\alpha p^X_{1i} + (1-\alpha) p^X_{2i}\}_{i\in I}$$

$$H(\pi_\alpha) = + \sum_{i\in I} \eta\left(\alpha p^X_{1i} + (1-\alpha) p^X_{2i}\right) \geq \alpha \sum_{i\in I} \eta(p^X_{1i}) + (1-\alpha) \sum_{i\in I} \eta(p^X_{2i})$$

$$= \alpha H(\pi^{(1)}_X) + (1-\alpha) H(\pi^{(2)}_X)$$

$\boxed{\text{Remark: Mixing probabilities increases the uncertainty}}$

•• $\Pi_{XY} = \{p^{XY}_{ij}\}_{\substack{i\in I\\ j\in J}}$: joint probability

$$\left.\begin{aligned} \Pi_X &= \{p^X_i\}_{i\in I} : \quad p^X_i = \sum_{j\in J} p^{XY}_{ij} \\ \Pi_Y &= \{p^Y_j\}_{j\in J} : \quad p^Y_j = \sum_{i\in I} p^{XY}_{ij} \end{aligned}\right\} : \text{Marginal Probabilities}$$

$$H(X) + H(Y) - H(X,Y) = \sum_{\substack{i\in I\\ j\in J}} p^{XY}_{ij} \log p^{XY}_{ij} - \sum_{i\in I} p^X_i \log p^X_i - \sum_{j\in J} p^Y_j \log p^Y_j$$

$$= \sum_{\substack{i\in I\\ j\in J}} p^{XY}_{ij} \left(\log p^{XY}_{ij} - \log\left(p^X_i\, p^Y_j\right)\right)$$

$$\geq \sum_{\substack{i\in I\\ j\in J}} \left(p^{XY}_{ij} - p^X_i\, p^Y_j\right) = 0$$

<u>Remark</u>: $H(X,Y) = H(X) + H(Y)$ iff $p^{XY}_{ij} = p^X_i\, p^Y_j \quad \forall\, i\in I, j\in J$

Definition 1.5.5

Given two random variables $X = \{x_i\}_{i\in I}$ and $Y = \{y_j\}_{j\in J}$ with joint probability $\Pi_{XY} = \{p^{XY}_{ij}\}_{\substack{i\in I\\ j\in J}}$ and marginal probabilities $\Pi_X = \{p^X_i = \sum_{j\in J} p^{XY}_{ij}\}_{i\in I}$ and $\Pi_Y = \{p^Y_j = \sum_{i\in I} p^{XY}_{ij}\}_{j\in J}$, the conditional probabilities of X given $Y = y_j$ and of Y given $X = x_i$ are given by

$$\boxed{\Pi_{X|Y=y_j} = \left\{ \frac{p^{XY}_{ij}}{p^Y_j} \right\}_{i\in I}} \quad \text{and} \quad \boxed{\Pi_{Y|X=x_i} = \left\{ \frac{p^{XY}_{ij}}{p^X_i} \right\}_{j\in J}}$$

Definition 1.5.6 The conditional entropies of X given Y and of Y given X are

$$\boxed{H(X|Y) = \sum_{j\in J} p^Y_j \, H(\Pi_{X|Y=y_j})} \quad \text{and} \quad \boxed{H(Y|X) = \sum_{i\in I} p^X_i \, H(\Pi_{Y|X=x_i})}$$

Consider $H(X|Y) = \sum_{j \in J} p_j^Y \, H(\pi_{X|Y=y_j})$ $\boxed{(\geq 0)}$

$$= -\sum_{j \in J} p_j^Y \sum_{i \in I} p_i^{X|Y=y_j} \log p_i^{X|Y=y_j}$$

$$= -\sum_{\substack{j \in J \\ i \in I}} p_j^Y \frac{p_{ij}^{XY}}{p_j^Y} \log \frac{p_{ij}^{XY}}{p_j^Y}$$

$$= -\sum_{\substack{j \in J \\ i \in I}} p_{ij}^{XY} \left(\log p_{ij}^{XY} - \log p_j^Y \right)$$

$$= H(X,Y) + \sum_{\substack{j \in J \\ i \in I}} p_{ij}^{XY} \log p_j^Y.$$

$$= H(X,Y) + \sum_{j \in J} p_j^Y \log p_j^Y = H(X,Y) - H(Y)$$

$\boxed{\text{Lemma 1.5.1.}}$ $H(X,Y) = H(Y) + H(X|Y) \geq H(Y)$

$\boxed{\text{Exercise 1.5.2}}$ Prove that $H(X,Y) = H(X) + H(Y|X) \geq H(X)$

Remark: if X and Y describe two physical systems S_1 and S_2 and (X,Y) describes the compound system S_1+S_2, then knowing S_1+S_2, $H(X,Y)=0$, means knowing both S_1, $H(X)=0$, and S_2, $H(Y)=0$.

$\boxed{\text{Example 1.5.3}}$ Given a macrostate M consisting of Ω_M possible microstates (configurations), the Boltzmann entropy of M is $\boxed{H_B(M) = \log \Omega_M}$

Given $X, \pi_X = \{p^X_i\}_{i \in I}$, let M be the collection of N outcomes of X. For $N \gg 1$ each x_i turns up $\boxed{N_i = N p^X_i}$ times and $\boxed{N = \sum_{i \in I} N_i}$.

Then $\boxed{\Omega_M = \dfrac{N!}{\prod_{i \in I}(N_i!)}}$. Using Stirling, $\boxed{\log N! \simeq N \log N - N}$

$$
\begin{aligned}
H_B(M) &\simeq N \log N - \sum_{i \in I} N_i \log N_i - N + \sum_{i \in I} N_i \\
&\simeq N \log N - \sum_{i \in I} N p^X_i (\log p^X_i + \log N) \\
&\simeq N\, H(X)
\end{aligned}
$$

Definition 1.5.7. von Neumann Entropy

Given a quantum system described by a density matrix ρ, the ignorance (before measurement of ρ) or knowledge (after measurement of ρ) about the system is measured by

$$\boxed{S(\rho) = -\operatorname{Tr} \rho \log \rho}$$

Observations

- If $\rho = \sum_{i \in I} r_i |r_i\rangle\langle r_i|$, $0 \le r_i \le 1$, $\sum_{i \in I} r_i = 1$, $\langle r_i | r_j \rangle = \delta_{ij}$ then

  $S(\rho) = -\sum_i r_i \log r_i = H(Sp(\rho))$, $Sp(\rho) = \{r_i\}_{i \in I} =: \Pi_\rho$

- $S(P_\psi) = 0$ $\forall$ $P_\psi = |\psi\rangle\langle\psi|$; if $\#(I) = N$ then $S(\rho) \le \log N = S(\frac{\mathbb{1}}{N})$

- $S(\rho)$ emerges as Boltzmann entropy of the macrostates consisting of the outcomes of N measurements of ρ.

Proposition 1.5.2 — Upper and lower bounds to $S(\rho)$

Let $\rho = \sum_{i\in I} d_i \rho_i \in \mathcal{S}(S)$:

$\forall i \in I : d_i \geq 0$; $\sum_{i\in I} d_i = 1$

$$\boxed{\sum_{i\in I} d_i S(\rho_i) \leq S(\rho) \leq \sum_{i\in I} d_i S(\rho_i) - \sum_i d_i \log d_i}$$

concavity

Proof

- $S(\rho) \geq \sum_{i\in I} d_i S(\rho_i)$: let $\rho|r_\alpha\rangle = r_\alpha |r_\alpha\rangle$, $r_\alpha \geq 0$, $\sum_\alpha r_\alpha = 1$, $\langle r_\alpha | r_\beta\rangle = \delta_{\alpha\beta}$

$\rho_i |r^i_\alpha\rangle = r^i_\alpha |r^i_\alpha\rangle$, $r^i_\alpha \geq 0$, $\sum_\alpha r^i_\alpha = 1$, $\langle r^i_\alpha | r^i_\beta\rangle = \delta_{\alpha\beta}$

$$S(\rho) = \sum_\alpha \eta(r_\alpha) = \sum_\alpha \eta\left(\langle r_\alpha | \rho | r_\alpha\rangle\right) = \sum_\alpha \eta\left(\sum_{i\in I} d_i \langle r_\alpha | \rho_i | r_\alpha\rangle\right)$$

$$\geq \sum_\alpha \sum_{i\in I} d_i \, \eta\left(\langle r_\alpha | \rho_i | r_\alpha\rangle\right) = \sum_\alpha \sum_{i\in I} d_i \, \eta\left(\sum_\beta r^i_\beta \, |\langle r^i_\beta | r_\alpha\rangle|^2\right)$$

$$\geq \sum_\alpha \sum_{i\in I} \sum_\beta |\langle r^i_\beta | r_\alpha\rangle|^2 d_i \, \eta(r^i_\beta) = \sum_{i\in I} \sum_\beta d_i \, \eta(r^i_\beta)$$

$$= \sum_{i\in I} d_i S(\rho_i)$$

$\boxed{\text{Since } \eta(x) = -x \log x \text{ is concave}}$

- $S(\rho) \leq \sum_{i\in I} d_i\, S(\rho_i) - \sum_{i \in I} d_i \log d_i$: it follows from the matrix-monotonicity of $f(x) = \log x$ (see next two lemmas) namely if $\boxed{A \geq B \geq 0}$ (positive operators) then

$\boxed{\log A \geq \log B}$.

Choose $A = \sum_{i\in I} d_i \rho_i$ and $B = d_j \rho_j$; then, $\boxed{\log\left(\sum_{i\in I} d_i \rho_i\right) \geq \log(d_j \rho_j)}$

and $\operatorname{Tr}\left(\left(\sum_{j\in I} d_j \rho_j\right) \log\left(\sum_{i\in I} d_i \rho_i\right)\right) \geq \operatorname{Tr}\left(\sum_{j \in I} d_j \rho_j \log(d_j \rho_j)\right)$

Then, $-S(\rho) \geq \sum_{j\in I} d_j \operatorname{Tr}(\rho_j \log \rho_j) + \sum_{j\in I} d_j \log d_j \operatorname{Tr}(\rho_j)$

Finally, $\boxed{S(\rho) \leq \sum_{j\in I} d_j\, S(\rho_j) - \sum_{j\in I} d_j \log d_j}$

Matrix-monotone functions

Definition 1.5.8 A function $f : \mathbb{R}^+ \to \mathbb{R}$ is called matrix monotone if $\boxed{A \geq B \Rightarrow f(A) \geq f(B)}$

Lemma 1.5.2 $A \geq B > 0 \Rightarrow \frac{1}{A} \leq \frac{1}{B}$

Proof: $A \geq B \iff A^{-1/2} B A^{-1/2} \leq 1 \iff \| A^{-1/2} B A^{-1/2} \| \leq 1$

$\iff \| B^{1/2} A^{-1/2} \| \leq 1 \iff \| A^{-1/2} B^{1/2} \| \leq 1$

$\iff \| B^{1/2} A^{-1} B^{1/2} \| \leq 1 \iff B^{1/2} A^{-1} B^{1/2} \leq 1 \iff \frac{1}{A} \leq \frac{1}{B}$

Lemma 1.5.3 $A \geq B > 0 \Rightarrow \log A \geq \log B$

Proof: $\log x = \int_0^{+\infty} dt \left(\frac{1}{1+t} - \frac{1}{x+t} \right)$

$$\log A - \log B = \int_0^{+\infty} dt \left(\frac{1}{t+B} - \frac{1}{t+A} \right) \geq 0 \Leftarrow t+A \geq t+B$$