2.4. Majorization and Lorenz curves

Recall from last time: if $p, q \in \mathbb{R}^m$ are probability vectors, (i.e. $p_i, q_i \geq 0$, $\sum p_i = \sum q_i = 1$), then

$$p > q \iff \sum_{i=1}^{m} p_i^k \geq \sum_{i=1}^{m} q_i^k \text{ for all } k = 1, \ldots, m,$$

where $p_i^k$ are the entries of $p$ in non-increasing order.

Furthermore, write $p \xrightarrow{\text{ noisy}} q$ if for every $\varepsilon > 0$ there exists $q_{\varepsilon}$ with $\|q - q_{\varepsilon}\| < \varepsilon$ and a noisy classical operation $D_\varepsilon$ with $D_\varepsilon(p) = q_{\varepsilon}$.

Lemma: Let $p, q \in \mathbb{R}^m$ be probability vectors of the same size. Then $p \xrightarrow{\text{ noisy}} q$ if and only if $p > q$.

Proof "$\Rightarrow$": Suppose $p \xrightarrow{\text{ noisy}} q$. Consider $D_\varepsilon$, then $D_\varepsilon$ maps prob. vectors to prob. vectors, and $D_\varepsilon(\frac{1}{m}, \ldots, \frac{1}{m}) = (\frac{1}{m}, \ldots, \frac{1}{m})$, $D_\varepsilon$ linear.

Problem 8, Exercise 2 $\Rightarrow$ $D_\varepsilon$ is a bistochastic matrix, and thus $p > q_{\varepsilon}$ for every $q_{\varepsilon} := D_\varepsilon(p)$. But the set of prob. vectors majorized by $p$ is topologically closed $\Rightarrow$ $p > q$. 
Suppose \( p > q \). Problem 6, Exercises 2 ⇒

\[ q \text{ can be obtained from } p \text{ as a mixture of permutations } \]

\[ q = \sum_{j=1}^{d} \lambda_j \Pi_j(p), \quad \lambda_j \geq 0, \quad \sum \lambda_j = 1, \text{ all } \Pi_j \text{ permutations of entries} \]

Consider only special case \( k = 2 \) (\( k \geq 3 \) analogous).

\[ q = \alpha \Pi_1(p) + (1 - \alpha) \Pi_2(p), \quad 0 \leq \alpha \leq 1. \]

Construct a noisy operation with \( p \rightarrow q_\varepsilon \):

\[ p = (p_1, \ldots, p_N) \]

\[ \pi_N \xi_N = \frac{1}{N} \sum_{i=1}^{N} (p_{\pi_1(i)}, \ldots, p_{\pi_N(i)}) \]

\[ \Pi[p \xi_N] = \frac{1}{N} \sum_{i=1}^{N_1} p_{\Pi_1(i)} \Pi_1(p) + \frac{1}{N_2} \sum_{i=1}^{N_2} p_{\Pi_2(i)} \Pi_2(p) \]

\[ (\Pi [p \xi_N])_{\text{out}} = \frac{1}{N} \left( N_1 p_{\Pi_1(1)} \Pi_1(p) + N_2 p_{\Pi_2(1)} \Pi_2(p) \right) + \frac{1}{N} \left( N_1 p_{\Pi_1(N)} \Pi_1(p) + N_2 p_{\Pi_2(N)} \Pi_2(p) \right) \]

By making \( N \to \infty \) and choosing \( N = N_1(N) \) appropriately, we can approximate \( \alpha \) to arbitrary accuracy.
Graphical tool: Low-rank curve: linear interpolation of
Prob. vector \( p \in \mathbb{R}^m \).

**Ex:** \( p = (0, \frac{3}{4}, \frac{1}{4}) \)

\[
L_p(x) = \sum_{i=1}^{k} p_i x^i (1-x)^{k-i}
\]

**Observation:** \( p \) and \( p \otimes y_n = p \otimes \left( \frac{1}{n}, \ldots, \frac{1}{n} \right) \) have identical
Low-rank curves

E.g. \( n = 2 \), \( p \otimes \left( \frac{1}{2}, \frac{1}{2} \right) = (0, 0, \frac{3}{8}, \frac{3}{8}, \frac{3}{8}, \frac{1}{8}) \)

**Lemma:** \( p \) using \( q \) if and only if \( L_p(x) \geq L_q(x) \) for all \( x \),
even if \( p \) and \( q \) do not have the same
dimensionalities

**Proof:** If \( n = n \) then \( L_p \geq L_q \iff p > q \) by definition.

Otherwise: \( p \geq q \iff p \otimes y_n \geq q \otimes y_n \)

\[
\iff L_{p \otimes y_n} \geq L_{q \otimes y_n} \iff L_p \geq L_q.
\]

All Low-rank curves are concave, and lie above the Low-rank
curve of \( y_n = \left[ \frac{1}{n}, \ldots, \frac{1}{n} \right] \).

\[
\begin{pmatrix}
\frac{1}{5} \\
0 \\
\frac{4}{5}
\end{pmatrix}
\geq
\begin{pmatrix}
\frac{1}{12} \\
\frac{11}{12}
\end{pmatrix}
\]

\( p \) is a more valuable
resource than \( q \).
2.5. Nonuniformity monotones

Let $P_n$ be the set of probability vectors in $\mathbb{R}^n$.

Def: A function $F : P_n \rightarrow \mathbb{R}$ is Schur-convex (on $P_n$) if $p > q \Rightarrow F(p) \geq F(q)$
(or Schur-concave if $p > q \Rightarrow F(p) \leq F(q)$).

Example: $H_\infty(p) := -\log \max_i p_i$

$p > q \Rightarrow p^*_1 > q^*_1 \Rightarrow \max_i p_i \geq \max_i q_i$

$\Rightarrow H_\infty(p) \leq H_\infty(q)$

$H_\infty$ is Schur-concave.

We will now prove the intuition that "noisy operations increase the entropy."

Lemma: For every convex (concave) function $f : [0,1] \rightarrow \mathbb{R}$, the function $F : P_n \rightarrow \mathbb{R}$ defined by

$$F(p) := \sum_{i=1}^{n} f(p_i)$$

is Schur-convex (Schur-concave).
If \( f \) is convex, then \( \int (1 - 2x) f(x) \leq f(y) \) for all \( y \in [0, 1] \).

If \( f \) is twice differentiable, the \( f'' > 0 \) ⇒ function is concave.

Proof: exercise.

Now consider Shannon entropy:

\[
H(p) := -\sum_{i=1}^{n} p_i \log p_i \quad (0 \log 0 := 0),
\]

\[
= \sum_{i=1}^{n} h(p_i), \quad h(p_i) = -p_i \log p_i.
\]

\( h \) concave ⇒ \( H \) Schur-concave.

\( p \neq q \Rightarrow H(p) \leq H(q) \quad \Rightarrow \text{Noisy operators } D : \mathbb{R}^n \rightarrow \mathbb{R}^n \text{ never decrease the entropy.} \)

Reminiscent of "Second Law".

For different dimensionalities: \( p \in \mathbb{R}^m, q \in \mathbb{R}^n \)

\[
p \leftrightarrow q \Leftrightarrow p \otimes y_m \leftrightarrow q \otimes y_n \Leftrightarrow p \otimes y_m > q \otimes y_n
\]

\[
\Rightarrow H(p \otimes y_m) \leq H(q \otimes y_n)
\]

\[
\Rightarrow H(p) + \log m \leq H(q) + \log n
\]

\[
\Rightarrow I(p) \geq I(q)
\]

For \( I(p) := \log d_p - H(p) \) "negentropy".

This is necessary, but is it sufficient? No.
Recall \( p = \left( \frac{1}{2}, \frac{1}{4}, \frac{1}{4}, 0 \right) \) and \( q = \left( \frac{4}{5}, \frac{1}{5} \right) \) (for now, \( \log \equiv \log_2 \)).

\[ I(p) = \frac{1}{2}, \quad I(q) \approx 0.28 \]

But \( p \not\leq q \).

More generally, \( \leq \) is a preorder, but \( I \) (like every other monad) introduces an order:

\[ p \leq q \iff I(p) \geq I(q), \quad \text{all states comparable.} \]

**Def:** A nonunifomity monad is a functor \( F : \mathbb{U} P_\mathbb{U} \rightarrow \mathbb{P} \) such that

\[ p \xrightarrow{\text{using}} q \implies F(p) \geq F(q). \]

(Example: negentropy \( I \).)

Because of (4), no single monad can fully characterize the possible state transitions.

26. Extracting work from states: distillable nonunifomity and nonuniformity of entropy.

How much work does it cost to create a state (as in Landauer's erasure) \( \mathcal{0} \), if we want to succeed with probability one? We will show:

\[ (\log d - H_\mathcal{0}(P))_{\mathcal{B}T} \ll \mathcal{2} \]
How much work can we extract from a state (in a Szilard engine)?

We show: \( (\log dp - H_0(p)) \leq kT \ln 2 \)

where

\[
H_\infty \leq H \leq H_0
\]

\( \begin{array}{ll}
\text{max entropy} & \text{Shannon entropy} \\
\text{min entropy} & \\
\end{array} \)

Fundamental irreversibility: have to invest more work than we can get out of the state later on.

But we will show: reversibility is recovered in the thermodynamic limit (work extraction "on average" for many copies).

**Def. (Rényi entropies):** Let \( p \in \mathbb{R}^n \) be any probability vector. For \( x > 0, x \neq 1 \), define

\[
H_x(p) = \frac{1}{1-x} \log \sum_{i=1}^n p_i^x
\]

By taking the limits, we also get

\[
H_0(p) = \log \text{rank}(p), \quad H_1(p) = -\sum_{i=1}^n p_i \log p_i; \quad \text{Shannon entropy}
\]

\[
H_\infty(p) = -\log \max_i p_i
\]

Analogously for quadruple states \( S \):

\[
H_x(S) = \frac{1}{1-x} \log \text{tr}(S^x)
\]

\[
H_0(S) = \log \text{rank}(S); \quad H_1(S) = S(S) \text{ von Neumann entropy}
\]

\[
H_\infty(S) = -\log \max_{i,j} (S_{ij})\text{ largest eigenvalue}
\]