Andrea Gadotti 28/04/2015

Exercise 18. We must prove

$$\varphi\left(\sum_{i=1}^{\infty} x_i p(x_i)\right) \leq \sum_{i=1}^{\infty} \varphi(x_i)p(x_i).$$

First we deal with the finite form.
Claim: For all $N \in \mathbb{N}$,

$$\varphi\left(\frac{\sum_{i=1}^{N} x_i p(x_i)}{\sum_{i=1}^{N} p(x_i)}\right) \leq \frac{\sum_{i=1}^{N} \varphi(x_i)p(x_i)}{\sum_{i=1}^{N} p(x_i)}. \tag{$*$}$$

Proof: First recall that "φ convex" means

$$\forall x_1, x_2 \in \mathbb{R}, \forall p_1, p_2 \in [0,1] \text{ s.t. } p_1 + p_2 = 1 : \quad \varphi(p_1 x_1 + p_2 x_2) \leq p_1\varphi(x_1) + p_2\varphi(x_2).$$

We prove $(*)$ by induction on N. Define

$$q(x_i) := \frac{p(x_i)}{\sum_{i=1}^{N} p(x_i)}.$$

If we prove

$$\varphi\left(\sum_{i=1}^{N} x_i q(x_i)\right) \leq \sum_{i=1}^{N} \varphi(x_i)q(x_i) \tag{$**$}$$

we are done. Now observe that $q(x_i) \in [0,1]$ and $\sum_{i=1}^{N} q(x_i) = 1$. So:
If $N = 2$, then $(**)$ is precisely the definition of "φ convex".
If $N > 2$, this can be easily proven by "grouping" $q(x_1), ..., q(x_N)$ and $q(x_N)$, and then applying the inductive hypothesis. ■

So $(*)$ holds for all $N \in \mathbb{N}$. Taking both sides to the limit we get

$$\lim_{N\to\infty} \varphi\left(\frac{\sum_{i=1}^{N} x_i p(x_i)}{\sum_{i=1}^{N} p(x_i)}\right) \leq \lim_{N\to\infty} \frac{\sum_{i=1}^{N} \varphi(x_i)p(x_i)}{\sum_{i=1}^{N} p(x_i)} = \frac{\sum_{i=1}^{\infty} \varphi(x_i)p(x_i)}{\sum_{i=1}^{\infty} p(x_i)} = \sum_{i=1}^{\infty} \varphi(x_i)p(x_i).$$

Recall that any convex function is continuous. So we can take the limit inside:

$$\varphi\left(\lim_{N\to\infty} \frac{\sum_{i=1}^{N} x_i p(x_i)}{\sum_{i=1}^{N} p(x_i)}\right) \leq \sum_{i=1}^{\infty} \varphi(x_i)p(x_i),$$

whereby

$$\varphi\left(\sum_{i=1}^{\infty} x_i p(x_i)\right) \leq \sum_{i=1}^{\infty} \varphi(x_i)p(x_i),$$

as wanted.

Exercise 19. Let $\mathcal{X} \subseteq \mathbb{R}$ be the range of X. Consider $Y := f(X)$. For all $y \in f[\mathcal{X}]$ (**note:** $f^{-1}(y)$ is, a priori, <u>a subset</u> of $\mathcal{X}$, not an element), define $p_Y(y) := p_X(f^{-1}(y))$. By definition of entropy we have

$$H(f(X)) = -\sum_{y \in f[\mathcal{X}]} p_Y(y) \log_2 p_Y(y).$$

Now observe that if f is injective, then $\forall y \in f[\mathcal{X}]\ \exists! x \in \mathcal{X} : \ f^{-1}(y) = x$. So $p_Y(y) = p_X(f^{-1}(y)) = p_X(x)$. This means

$$H(f(X)) = -\sum_{y \in f[\mathcal{X}]} p_Y(y) \log_2 p_Y(y) = -\sum_{x \in \mathcal{X}} p_X(x) \log_2 p_X(x) = H(X).$$

So, if f is injective, then $H(f(X)) = H(X)$.
Now suppose that f is not injective, i.e. there exist at least two distinct elements $x_1, x_2 \in \mathcal{X}$ such that $f(x_1) = f(x_2) =: y$. Observe that $p_Y(y) = p_X(x_1) + p_X(x_2)$. Thus

$$\begin{aligned}
&- p_Y(y) \log_2 p_Y(y) = -(p_X(x_1) + p_X(x_2)) \log_2(p_X(x_1) + p_X(x_2)) \\
&\qquad = -p_X(x_1) \log_2(p_X(x_1) + p_X(x_2)) - p_X(x_2) \log_2(p_X(x_1) + p_X(x_2)) =: h(y),
\end{aligned}$$

while

$$-\sum_{x \in \{x_1, x_2\}} p_X(x) \log_2 p_X(x) = -p_X(x_1) \log_2 p_X(x_1) - p_X(x_2) \log_2(p_X(x_2)) =: h(x_1, x_2).$$

Since $p_X(x_1)$ and $p_X(x_2)$ are positive and $\log_2$ is a monotone increasing function, we have that $\log_2(p_X(x_1) + p_X(x_2)) \geq \log_2 p_X(x_1)$ and $\log_2(p_X(x_1) + p_X(x_2)) \geq \log_2 p_X(x_2)$. Therefore $h(y) \leq h(x_1, x_2)$.
Since the addends of $H(f(X))$ are made like $h(y)$, and the ones of $H(X)$ are made like $h(x_1, x_2)$, repeating the same argument for all $y \in f(\mathcal{X})$ we get that $H(f(X)) \leq H(X)$, as wanted.

Given any random variable Z, it's immediate to see that every addend in $H(Z)$ is ≥ 0. Hence, if $H(Z) = 0$ then necessarily every addend is 0. So $p_Z(z) \log_2 p_Z(z) = 0$ for all z. Since we can suppose w.l.o.g. $p_Z(z) \neq 0$, we get $\log_2 p_Z(z) = 1$, i.e. $p_Z(z) = 1$.
So, $H(f(X)) = 0$ implies $p_Y(y) = 1$ for some $y \in f[\mathcal{X}]$, i.e. $p_X(f^{-1}(y)) = 1$, i.e. $f[x] = y$ for almost every $x \in \mathcal{X}$. That is, f is almost surely constant.

Exercise 20. Define $Z := -\log_2 \sqrt[n]{\prod_{i=1}^n \mathbb{P}[X_i]}$. For all $i \in \mathbb{N}$, define $Y_i(\omega) := \log_2 \mathbb{P}[X_i = X_i(\omega)]$. Observe that

$$\begin{aligned} Z(\omega) &= -\log_2 \sqrt[n]{\prod_{i=1}^n \mathbb{P}[X_i(\omega)]} \\ &= -\frac{1}{n}\sum_{i=1}^n \log_2 \mathbb{P}[X_i(\omega)] \\ &= -\frac{1}{n}\sum_{i=1}^n \log_2 \mathbb{P}[X_i = X_i(\omega)] \\ &= -\frac{1}{n}\sum_{i=1}^n Y_i(\omega). \end{aligned}$$

Since the X_i's are iid, it's immediate to check that also the Y_i's are iid. Furthermore, write the range of X_1 as $\mathcal{X}_1 = \{x_1, x_2, ..., x_k, ...\}$. We trivially have

$$\mathbb{E}[Y_1] = \sum_{k=1}^{\infty} \mathbb{P}[X_1 = x_k] \log_2 \mathbb{P}[X_1 = x_k] = -H(X_1).$$

Therefore the statement follows directly by the Strong Law of Large Numbers.

Exercise 21. The (a) part has already been proven in the last paragraph of Exercise 19.
For the (b) part, for all $i = 1, ..., n$ define $p_i := \mathbb{P}[X = x_i]$. Now observe that by Exercise 22, taking $q_i = \frac{1}{n}$ for all $i = 1, ..., n$, we get

$$H(X) = -\sum_{i=1}^n p_i \log_2 p_i \leq -\sum_{i=1}^n p_i \log_2 q_i = -\sum_{i=1}^n p_i \log_2 \frac{1}{n} = \log_2 n \sum_{i=1}^n p_i = \log_2 n,$$

and the equality holds iff $p_i = q_i = \frac{1}{n}$ for all $i = 1, ..., n$. So we are done.

Exercise 22. Since

$$\log_2 a = \frac{\ln a}{\ln 2}$$

it is sufficient to prove the statement using the natural logarithm. Note that the natural logarithm satisfies $\ln x \leq x - 1$ for all $x > 0$ *with equality if and only if* $x = 1$.
Of course we can suppose w.l.o.g. $p_i > 0$ for all i. So

$$-\sum_{i=1}^n p_i \ln \frac{q_i}{p_i} \geq -\sum_{i=1}^n p_i \left(\frac{q_i}{p_i} - 1\right) = -\sum_{i=1}^n q_i + \sum_{i=1}^n p_i = -\sum_{i=1}^n q_i + 1 \geq 0. \qquad (*)$$

Therefore

$$-\sum_{i=1}^{n} p_i \ln q_i \geq -\sum_{i=1}^{n} p_i \ln p_i,$$

as wanted.
As for the equality, of course it holds if $p_i = q_i$ for all i. On the other hand, suppose

$$-\sum_{i=1}^{n} p_i \ln q_i = -\sum_{i=1}^{n} p_i \ln p_i.$$

Then $-\sum_{i=1}^{n} p_i \ln \frac{q_i}{p_i} = 0$, and thus all the inequalities in $(*)$ are equalities, i.e.

$$0 = \sum_{i=1}^{n} p_i \left(\frac{q_i}{p_i} - 1\right) - \sum_{i=1}^{n} p_i \ln \frac{q_i}{p_i} = \sum_{i=1}^{n} p_i \underbrace{\left(\frac{q_i}{p_i} - 1 - \ln \frac{q_i}{p_i}\right)}_{\geq 0 \text{ since } \ln x \leq x-1}.$$

So we have a sum of positive numbers which equals 0. Thus each term is zero, and since $p_i \neq 0$ for all i we get

$$\frac{q_i}{p_i} - 1 = \ln \frac{q_i}{p_i}$$

for all i. So $\frac{q_i}{p_i} = 1$ for all i, by the sentence written in italic at the top.