Official solutions to Exercise 17 from Discrete Stochastics and Informamtion Theory

April 19, 2018

Let X be a discrete random variable with values $\{x_1, \ldots, x_n\}$ and the distribution $\mathbb{P}[X = x_i] = p_i$ where $\sum_{i=1}^n p_i = 1$.

(a) In part (a) we have to show that $H(X) \geq 0$ and $H(X) = 0$ iff X is almost surely constant.

By the definition of the entropy it holds, that $H(X) = \sum_{i=1}^n - \log_2(p_i) \cdot p_i$.

If $\exists i \in \{1, \ldots, n\} : p_i \in]0, 1[$, then $\ln(p_i) < 0$ and $p_i > 0$. As all summands have the same sign, $H(X) > 0$ in this case.
Otherwise, if $\forall i \in \{1, \ldots, n\} : p_i \in \{0, 1\}$, (i.e $\exists i \in \{1, \ldots, n\} : p_i = 1$) the product $p_i \cdot \log_2(p_i)$ is zero. Hence their sum is zero and so is $H(X)$.

(b) For part (b) we have to show that the maximum of the mapping $p \mapsto H(p)$ is reached iff X is uniformely distributed.

Solution using Lagrange Optimization The mapping $p \mapsto H(p) = \sum_{i=1}^n - \ln(p_i) \cdot p_i$ is definded on the following set:

$$\Delta_n = \left\{ (p_1, \ldots, p_n) \in \mathbb{R}^n \mid \forall i \in \{1, \ldots, n\} : p_1 \geq 0 \wedge \sum_{i=1}^n p_i = 1 \right\}$$

On $\mathrm{Int}(\Delta_n) = \Delta_n \setminus \partial\Delta_n$[1], the entropy H is $\mathcal{C}^\infty$ smooth. By Lagrange optimization, we get a neccessary condition for extreme values, i.e that the first partial derivertives of the function

$$L(p) = \sum_{i=1}^n - \log_2(pi) \cdot p_i + \lambda \left(\sum_{i=1}^n p_i - 1 \right)$$

[1]The considered topology on Δ is the topology induced by the eqality constraint $\sum_{i=1}^n p_i = 1$, i.e. the topology in the affine subspace $\{(p_1, \ldots p_n) | \sum_{i=1}^n p_i = 1\}$. Alternatively, Δ_n can be considered as a $n-1$-dimensional smooth submanifold of $\mathbb{R}^n$ with boundary. Local coordinate charts are given by $\varphi_i : \Delta \to \mathbb{R}^{n-1}$ with $\varphi_i(p_1, \ldots p_n) = (p_1, \ldots, p_{i-1}, p_{i+1}, \ldots, p_n)$ for $i \in \{1, \ldots, n\}$

vanish. So let's compute these derivertives.

$$\frac{\partial L}{\partial p_i} = -\log_2(p_i) - \frac{1}{\ln(2)} + \lambda = 0$$

Equivalently,

$$-\log_2(p_i) = \frac{1}{\ln(2)} - \lambda$$

So $-\log_2(p_i)$ does not depend on i. As $-\log_2$ is bijective it follows that p_i is constant for all i. Hence, if H attains a maximum on $\mathrm{Int}(\Delta_n)$, this can only happen if H is uniformely distributed. It remains to show, that this point is in fact a maximum (and not a minimum or a saddle point). For this, we show:

$$H\left(\frac{1}{n}, \ldots, \frac{1}{n}\right) = \log_2(n) > H(p_1, \ldots p_n) \;\forall\, (p_1, \ldots p_n) \in \partial\Delta_n \tag{1}$$

The equlity of (1) is easy to see by plugging in into the definition of H. We show the inequality in (1) by iduction.

For $n = 1$, there is nothing to show. For $n = 2$, $\partial\Delta_2 = \{(0,1),(1,0)\}$ and $\mathrm{Int}(\Delta_2) = \{(p, 1-p) | p \in]0,1[\}$. In part (a), we have seen that H takes its unique minima on the boundary of Δ_2.

For the induction step, we consider the following induction hypothesis:

$$\forall m \leq n : H(p) \leq \log_2(m) \leq \log_2(n) \;\forall p \in \Delta_m \tag{2}$$

It is enough to show that $H(p) < \log_2(n+1)$ for $p \in \partial\Delta_{n+1}$. So let $p = (p_1, \ldots, p_{n+1}) \in \partial\Delta_{n+1}$. Because p lies on the boundary of Δ_{n+1}, there exists a $j \in \{1, \ldots, n+1\}$ such that $p_j = 0$. Consider $\tilde{p} = (p_1, \ldots, p_{j-1}, p_{j+1}, \ldots, p_{n+1}) \in \Delta_n$. Then, by the definition of the entropy, $H(\tilde{p}) = H(p)$. By induction hypothesis we get $H(\tilde{p}) \leq \log_2(n) < \log_2(n+1)$, which completes the induction proof.

Solution using Jensen Inequality Observe that the function $f :]0,1] \to \mathbb{R}$, $f(x) = -x \cdot \log_2(x)$ is strictly concave. This is because the second derivertive is strictly negative:

$$f''(x) = (-x \cdot \log_2(x))'' = -\left(x \cdot \frac{1}{x\ln(2)} + \log_2(x)\right)' = -\left(\frac{1}{\ln(2)} + \log_2(x)\right)' = -\frac{1}{x\ln(2)} < 0$$

The continuos continuation of f to $[0,1]$ is still strictly concave. By Jensen, inequality, we have for a $p = (p_1, \ldots, p_n) \in \Delta_n$:

$$\begin{aligned} H(p) &= \sum_{i=1}^{n} -p_i \log_2(p_i) = \sum_{i=1}^{n} f(p_i) = n \sum_{i=1}^{n} \frac{1}{n} f(p_i) \\ &\geq nf\left(\sum_{i=1}^{n} \frac{1}{n} p_i\right) = nf\left(\frac{1}{n}\right) = n \cdot \left(-\frac{1}{n}\log_2\left(\frac{1}{n}\right)\right) = \log_2(n) \end{aligned}$$

For strict concave function, equality holds in Jenen's inequality iff $f(p_1) = \ldots = f(p_n)$, which is only given if X is uniformely distributed.