Homework 2 Simple Solution

TA: Heng-Chien Liou*

1. (Maximum entropy) [22]

In the lecture we show that for a random variable taking values in a *finite* alphabet $\mathcal{X}$, the maximum entropy is $\log|\mathcal{X}|$, and the maximizing distribution is $\mathrm{Unif}(\mathcal{X})$. In this problem, we consider a random variable X taking values in the countable alphabet $\mathbb{N}$ (the set of positive integers) with mean $\mu < \infty$. Please use the steps below to prove that its entropy is upper bounded as follows:

$$\mathrm{H}(X) \le \mu_X \mathsf{h_b}(\mu_X^{-1}).$$

Notation: The PMF of X is denoted by $\mathsf{P}_X(i) \equiv p_i$, $i \in \mathbb{N}$, and $\mathsf{E}[X] \equiv \mu_X = \sum_{i=1}^{\infty} ip_i$.

a) Use the concavity of $\log(\cdot)$ and Jensen's inequality to show that for any non-negative sequence $\{q_i\}_{i=1}^{\infty}$ with $\sum_{i=1}^{\infty} q_i = 1$,
$$\mathrm{H}(X) \le -\sum_{i=1}^{\infty} p_i \log q_i. \qquad [8]$$

b) Find a non-negative sequence $\{q_i\}_{i=1}^{\infty}$ and α, β such that $\sum_{i=1}^{\infty} q_i = 1$, $\sum_{i=1}^{\infty} iq_i = \mu_X$, and
$$-\log q_i = \alpha i + \beta \quad \forall\, i \in \mathbb{N}. \qquad [8]$$

c) Complete the proof by plugging in the $\{q_i\}_{i=1}^{\infty}$ found in Part b) into the upper bound of $\mathrm{H}(X)$ found in Part a). Show that this upper bound on $\mathrm{H}(X)$ is attainable with an appropriate choice of $\{p_i\}_{i=1}^{\infty}$ and hence it is the maximum entropy of random variables taking values in positive integers with expected value being μ_X. [6]

Solution:

a) $\mathrm{H}(X) + \sum\limits_{i=1}^{\infty} p_i \log q_i = \sum\limits_{i=1}^{\infty} p_i \log \frac{q_i}{p_i}$. By Jensen, $\sum\limits_{i=1}^{\infty} p_i \log \frac{q_i}{p_i} \le \log \sum\limits_{i=1}^{\infty} p_i \frac{q_i}{p_i} = 0$.

b) $$\begin{cases} 1 = \sum\limits_{i=1}^{\infty} q_i = \frac{2^{-(\alpha+\beta)}}{1-2^{-\alpha}} \\ \mu_X = \sum\limits_{i=1}^{\infty} iq_i = \frac{2^{-(\alpha+\beta)}}{(1-2^{-\alpha})^2} \end{cases} \implies \begin{cases} \alpha = \log \frac{\mu_X}{\mu_X - 1} \\ \beta = \log(\mu_X - 1) \end{cases}$$

c) $\mathsf{H_b}(\frac{1}{\mu_X}) = \frac{1}{\mu_X} \ln \mu_X + \frac{\mu_X - 1}{\mu_X} \ln \frac{\mu_X}{\mu_X - 1}$

*with contributions by Wen-Shao Ho

$\mathrm{H}(X) \le -\sum\limits_{i=1}^{\infty} p_i \log q_i = \sum\limits_{i=1}^{\infty} p_i[i \ln \frac{\mu_X}{\mu_X - 1} + \log(\mu_X - 1)] = \mu_X \mathsf{H}_\mathsf{b}(\frac{1}{\mu_X})$

Equality holds iff $p_i = q_i \ \forall i$.

Grading Policy:

a) Correct arithmetic [4] and application of Jensen's inequality on the right variable [4].

b) Specify α and β [8].

c) Arithmetic [3] and selection of $\{p_i\}_{i=1}^{\infty}$ [3].

2. (Entropy of a random variable with a infinitely countable support) [14]

In the lecture, we define the entropy for a random variable with a finite alphabet $\mathcal{X}$ (in fact a finite support supp_{P_X} suffices). For a random variable X that has an infinitely *countable* support, sometimes $\mathrm{H}(X)$ is finite and sometimes $\mathrm{H}(X)$ becomes infinite. In this problem we look at an example.

a) Consider an infinite series

$$\sum_{n=2}^{\infty} \frac{1}{n\,(\log n)^{\alpha}}$$

where $\alpha \ge 0$. Use the *integral test for convergence* to show that the series converges if and only if $\alpha > 1$. [7]

b) Let s_α denote the above series if the series converges. Let us define a random variable $X_\alpha \in \{2, 3, \ldots\}$ with PMF

$$\mathsf{P}_{X_\alpha}(n) = \frac{1}{s_\alpha n\,(\log n)^{\alpha}}.$$

Show that $\mathrm{H}(X_\alpha)$ exists if $\alpha > 2$ and it diverges to ∞ if $1 < \alpha \le 2$. [7]

Solution:

a) $\int_2^{\infty} \frac{dx}{x(\log x)^{\alpha}} = \begin{cases} \frac{ln2}{(\alpha-1)} & (\alpha > 1) \\ \infty & (0 \le \alpha \le 1) \end{cases}.$

By the integral test for convergence, the series converges if and only if $\alpha > 1$.

Many students derive wrong integral when $\alpha = 1$.

b) Let $s_\alpha := \sum\limits_{n=2}^{\infty} \frac{1}{n(\log n)^{\alpha}}$, for $\alpha > 1$.

$$\begin{aligned}\mathrm{H}(X_n) &= \sum_{n=2}^{\infty} \mathsf{P}_{X_\alpha}(n) \log \frac{1}{\mathsf{P}_{X_\alpha}(n)} \\ &= \log s_\alpha + \frac{1}{s_\alpha} \sum_{n=2}^{\infty} \frac{1}{n(\log n)^{\alpha-1}} + \frac{1}{s_\alpha} \sum_{n=2}^{\infty} \frac{\alpha \log(\log n)}{n(\log n)^{\alpha}}\end{aligned}$$

$$\begin{cases} \leq \infty & (\alpha > 2) \\ = \infty & (1 < \alpha \leq 2) \end{cases}$$

Grading Policy:

a) Correct statement of the integral test [2] and correct evaluation of the integral [5].

b) Correct evaluation of the relevant integrals [3] and correct argument of the series's convergence and divergence [4].

3. (Mixture of random processes) [14]

In this problem we look at different ways to generate mixtures of random processes, and the entropy rate of the mixture of random processes. Consider two stationary random processes $\{X_0[i] \mid i \in \mathbb{N}\}$ and $\{X_1[i] \mid i \in \mathbb{N}\}$ taking values in disjoint alphabets $\mathcal{X}_0$ and $\mathcal{X}_1$ respectively. The two processes are independent from each other, that is, $\{X_0[i]\} \perp\!\!\!\perp \{X_1[i]\}$, and they have entropy rates $\mathcal{H}_0$ and $\mathcal{H}_1$ respectively. Let $\{\Theta_i \mid i \in \mathbb{N}\}$ be a **stationary** Bernoulli random process, independent of everything else.

a) Let $\Theta_i = \Theta$ for all $i \in \mathbb{N}$, where $\Theta \sim \text{Ber}(q)$. Is the random process $\{X_{\Theta_i}[i]\}$ stationary? What is its entropy rate? [6]

b) Let $\{\Theta_i\}$ be Markov with a probability transition matrix

$$\mathsf{P}_{\Theta_2|\Theta_1} = \begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix}, \text{ for } \alpha, \beta \in (0,1).$$

Suppose that both $\{X_0[i]\}$ and $\{X_1[i]\}$ are i.i.d. processes in this problem. Is the random process $\{X_{\Theta_i}[i]\}$ stationary? What is its entropy rate? [8]

Solution:

a) Since $\{X_0[i] | i \in \mathbf{N}\}$ and $\{X_1[i] | i \in \mathbf{N}\}$ are stationary processes, $\forall n.l \in \mathbf{N}$,

$$\begin{aligned}
&\mathsf{P}_{X_{\Theta_1}[1],\ldots,X_{\Theta_n}[n]} \\
=&(1-q)\mathsf{P}_{X_0[1],\ldots,X_0[n]} + q\mathsf{P}_{X_1[1],\ldots,X_1[n]} \\
=&(1-q)\mathsf{P}_{X_0[l+1],\ldots,X_0[n]} + q\mathsf{P}_{X_1[1],\ldots,X_1[l+n]} \\
=&\mathsf{P}_{X_{\Theta_{l+1}}[l+1],\ldots,X_{\Theta_{l+n}}[l+n]}
\end{aligned}$$

So $\{X_{\Theta_i}[i]\}$ is stationary.
Let $Y_i = X_{\Theta_i}[i]$.

$$\begin{aligned}
\mathcal{H}(X_{\Theta_i}[i]) &= \lim_{n\to\infty} \mathrm{H}\left(Y_n \middle| Y^{n-1}\right) \\
&= \lim_{n\to\infty} \mathrm{H}\left(Y_n, \Theta_n \middle| Y^{n-1}, \Theta^{n-1}\right) \quad (\mathcal{X}_0 \cap \mathcal{X}_1 = \varnothing)
\end{aligned}$$

$$
\begin{aligned}
&= \lim_{n\to\infty} \mathrm{H}\left(Y_n \middle| Y^{n-1}, \Theta^n\right) + \mathrm{H}\left(\Theta_n \middle| Y^{n-1}, \Theta^{n-1}\right) \quad (\textit{chain rule}) \\
&= \lim_{n\to\infty} \mathrm{H}(Y_n|\Theta_n) + \mathrm{H}(\Theta_n|\Theta_{n-1}) \\
&= \lim_{n\to\infty} \mathrm{H}(Y_n|\Theta) + \mathrm{H}(\Theta|\Theta) \\
&= q\mathcal{H}_1 + (1-q)\mathcal{H}_0 + 0
\end{aligned}
$$

b) $\{X_{\Theta_i}[i]\}$ is stationary. Let $Y_i = X_{\Theta_i}[i]$. $\mathcal{H}(\{X_{\Theta_i}[i]\}) = \lim_{n\to\infty} \mathrm{H}(Y_n|Y^{n-1})$.

$$
\begin{aligned}
&\mathrm{H}\left(Y_n \middle| Y^{n-1}\right) \\
&= \mathrm{H}\left(Y_n, \Theta_n \middle| Y^{n-1}\right) \quad (\mathcal{X}_0 \cap \mathcal{X}_1 = \varnothing) \\
&= \mathrm{H}\left(Y_n \middle| \Theta_n, Y^{n-1}\right) + \mathrm{H}\left(\Theta_n \middle| Y^{n-1}\right) \quad \textit{chain rule} \\
&= \mathsf{Pr}\{\Theta_n = 1\}\mathrm{H}\left(X_1[n] \middle| X_1^{n-1}\right) + \mathsf{Pr}\{\Theta_n = 0\}\mathrm{H}\left(X_0[n] \middle| X_0^{n-1}\right) + \mathrm{H}\left(\Theta_n \middle| Y^{n-1}\right) \\
&= \mathsf{Pr}\{\Theta_n = 1\}\mathrm{H}\left(X_1[n] \middle| X_1^{n-1}\right) + \mathsf{Pr}\{\Theta_n = 0\}\mathrm{H}\left(X_0[n] \middle| X_0^{n-1}\right) + \mathrm{H}\left(\Theta_n \middle| \Theta^{n-1}\right) \quad (\mathcal{X}_0 \cap \mathcal{X}_1 = \varnothing) \\
&= \frac{\alpha}{\alpha+\beta}\mathcal{H}_1 + \frac{\beta}{\alpha+\beta}\mathcal{H}_0 + \mathrm{H}(\Theta_2|\Theta_1) \\
&= \frac{\alpha}{\alpha+\beta}(\mathcal{H}_1 + \mathsf{H_b}(\beta)) + \frac{\beta}{\alpha+\beta}(\mathcal{H}_0 + \mathsf{H_b}(\alpha))
\end{aligned}
$$

Grading Policy:

a) Stationary argument [2], chain rule and entropy calculation [3], asympototic [1].

b) Stationary argument [2], chain rule and entropy calculation [6].