Problem sheet 2, Information Theory, HT 2021

Designed for the second tutorial class

Question 1 We are given a fair coin, and want to generate a random variable X from i.i.d. sampling from tossing the coin, such that X follows the distribution

$$\mathbb{P}(X=1)=p,\ \mathbb{P}(X=0)=1-p$$

with any given constant $p \in (0,1)$.

Suppose $Z_1, Z_2, \cdots$ are the results of independent tossing of the coin, i.e., $\{Z_i\}$ is an i.i.d. sequence of random variable with the distribution $\mathbb{P}(Z=0)=\mathbb{P}(Z=1)=\frac{1}{2}$. Denote $U = \sum_{i=1}^{+\infty} Z_i 2^{-i}$, and define

$$X = \begin{cases} 1 & \text{if } U < p \\ 0 & \text{otherwise} \end{cases}.$$

(a) Show that U follows a uniform distribution over $[0,1)$, and hence show that $\mathbb{P}(X=1) = p,\ \mathbb{P}(X=0)=1-p$.

(b) Denote I as the minimal number of n such that we we can tell $U < p$ based on $Z_1, \cdots Z_n$. Calculate $\mathbb{E}[I]$ and show that $\mathbb{E}[I] \leq 2$.

Question 2 For any $q \in [0,1]$ and $n \in \mathbb{N}$ such that nq is an integer, show that

$$\frac{2^{nH(q)}}{n+1} \leq \begin{pmatrix} n \\ nq \end{pmatrix} \leq 2^{nH(q)}.$$

`Hint: Consider the i.i.d. Bernoulli sequence` $X_1, X_2, \cdots, X_n$ `with` $\mathbb{P}(X=1)=q,\ \mathbb{P}(X=0)=1-q$`).`

Question 3 Let X_1 be a random variable valued in $\mathcal{X}_1 = \{1,2,\cdots,m\}$ and X_2 be a random variable valued in $\mathcal{X}_2 = \{m+1,\cdots,n\}$ for integers $n > m$. Let θ be a random variable with $\mathbb{P}(\theta=1)=\alpha,\ \mathbb{P}(\theta=2)=1-\alpha$ for some $\alpha \in [0,1]$. Define a new random variable

$$X = X_\theta.$$

Furthermore, suppose θ, X_1, X_2 are independent to each other.

(a) Express $H(X)$ in terms of $H(X_1), H(X_2)$ and $H(\theta)$.

(b) Show that $2^{H(X)} \leq 2^{H(X_1)} + 2^{H(X_2)}$. Can the equality hold in this inequality?

Question 4 The *differential entropy* of a $\mathbb{R}^n$-valued random variable X with density function $f(\cdot)$ is defined as

$$h(X) := -\int_{\mathbb{R}^n} f(\mathbf{x}) \log(f(\mathbf{x})) d\mathbf{x}$$

with the convention $0 * \log(0)) = 0$.

(a) Calculate $h(X)$ for the following cases with $n = 1$.

(1) X is uniformly distributed on an interval $[a, b] \subset \mathbb{R}$;

(2) X is a standard normal distribution;

(3) X is exponential distributed with parameter $\lambda > 0$.

(b) For general n-dimensional case, if $\mathbb{E}[X] = 0$, and $\mathrm{Var}(X) = K$, (K is the variance-covariance matrix). Show that

$$h(X) \leq n \log(\sqrt{2\pi e}) + \log(\sqrt{|K|})$$

with the equality hold iff X is multivariable normal.

`Hint: you can firstly prove the continuous version of Gibbs' inequality: For any two density functions` $f(\cdot)$ `and` $g(\cdot)$`,`

$$-\int f(\mathbf{x}) \log(f(\mathbf{x})) d\mathbf{x} \leq -\int f(\mathbf{x}) \log(g(\mathbf{x})) d\mathbf{x}.$$

`Also, you can try to prove (or use it without proof) the following property of variance-covariance matrix: If` $X = (X_1, \cdots, X_n)^\top$ `has variance-covariance matrix` $\mathrm{Var}(X) = K$`, then`

$$\mathbb{E}[X^\top K^{-1} X] = n.$$

Question 5 *(Strong AEP in Proposition 2.10)* Let X be a random variable with pmf p over the image space $\mathcal{X}$ with finite elements $k = |\mathcal{X}|$, $\vec{X} = (X_1, \cdots, X_n)$, we label elements in $\mathcal{X}$ by a non-decreasing order of $p(x)$, such that $p_i = \mathbb{P}(X = x_i)$ is non-decreasing in i. By this labelling, we can easily rank the probability $\mathbb{P}(\vec{X} = \vec{x})$ for all $\vec{c} \in \mathcal{X}^n$, and explicitly construct the smallest set $\mathcal{S}_n^\varepsilon$ by greedily including the element in $\mathcal{X}^n$ with highest probabilities one-by-one until we have $\mathbb{P}(\vec{X} \in \mathcal{S}_n^\varepsilon) \geq 1 - \varepsilon$.

Show that for any $\varepsilon > 0$, we have

$$(1 - 2\varepsilon) 2^{n(H(X)-\varepsilon)} \leq |\mathcal{S}_n^\varepsilon| \leq 2^{n(H(X)+\varepsilon)}.$$

`Hint: For any` $\varepsilon_1 \in [0, 1), \varepsilon_2 \in [0, 1)$ `and events` A, B `with` $\mathbb{P}(A) \geq 1 - \varepsilon_1, \mathbb{P}(B) \geq 1 - \varepsilon_2$`, show that` $\mathbb{P}(A \cap B) \geq 1 - \varepsilon_1 - \varepsilon_2$`. Use this inequality to estimate` $\mathbb{P}(\mathcal{S}_n^\varepsilon \cap \mathcal{T}_n^\varepsilon)$`.`

Question 6 (Optional, revision/outlook on Markov chain) A Markov chain is a sequence of discrete random variables $(X_n)_{n\geq 1}$ such that for all $x_1, \cdots, x_{n+1}$ valued in $\mathcal{X}$,

$$\mathbb{P}(X_{n+1} = x_{n+1}|X_n = x_n, \cdots, X_1 = x_1) = \mathbb{P}(X_{n+1} = x_{n+1}|X_n = x_n).$$

The Markov chain is called homogenous if $p_n(x, y) := \mathbb{P}(X_{n+1} = y|X_n = x)$ does not depend on n (for every $(x, y) \in \mathcal{X}^2$). In this case we call $(p(x, y))_{(x,y)\in\mathcal{X}^2}$ the transition matrix of (X_n).

(a) Repeat rolling a fair die independently, and denote by $\{Z_n\}$ the resulted numbers. Which of the following are Markov chains? For those that are, give the transition matrix.

(1) $X_n = \max_{i\leq n} Z_i$, which is the largest roll up to the n^{th} roll;

(2) X_n is the number of sixes in the first n rolls;

(3) X_n is the number of rolls since the most recent six;

(4) X_n is the time until the next six.

(b) Let $(X_n)_{n\geq 1}$ be a Markov chain. Which of the following are Markov chains?

(1) $(X_{m+n})_{n\geq 1}$ for a fixed integer $m > 0$;

(2) $(X_{2n})_{n\geq 1}$;

(3) $(Y_n)_{n\geq 1}$ with $Y_n := (X_n, X_{n+1})$.