CS3236: Tutorial 5
(Continuous Channels)

1. **[Differential Entropies]**

   (a) Assume that a continuous-valued random variable Z has a probability density that is 0 except in the interval $[-a, a]$. Show that the differential entropy $h(Z)$ is upper bounded by $1 + \log_2 a$, with equality if and only if Z is uniformly distributed between $[-a, a]$.

   (Hint: Use a similar trick based on KL divergence to the proof of $H(X) \le \log_2 |\mathcal{X}|$ in the discrete setting.)

   (b) Suppose that X is a Laplace random variable, i.e., $f_X(x) = \frac{1}{2b}\exp(-|x|/b)$ for $x \in \mathbb{R}$. Find the differential entropy $h(X)$.

   (Hint: This distribution satisfies $\mathbb{E}[|X|] = b$.)

2. **[Entropy vs. Differential Entropy]** Let X and Y be discrete real-valued random variables with joint probability mass function P_{XY}, and let U and V be continuous real-valued random variables with joint probability density function f_{UV}. Recall that $H(\cdot)$ denotes the entropy for discrete random variables, and $h(\cdot)$ denotes the differential entropy for continuous random variables.

   For each of the following, either explain why the given statement is <u>always true</u>, or explain why it is <u>sometimes false</u>. In your answers, you may make use of any statement proved in the lectures, unless it is the exact statement in the question.

   (Hint: If done well, each of these can be answered correctly in 1 or 2 sentences.)

   (a) $I(X;Y) \le H(X)$
   (b) $I(U;V) \le h(U)$
   (c) $H(X) = H(cX)$ for any fixed constant $c > 0$
   (d) $h(U) = h(cU)$ for any fixed constant $c > 0$
   (e) $H(X) \le \frac{1}{2}\log_2\left(2\pi e\,\mathbb{E}[X^2]\right)$
   (f) $h(U) \le \frac{1}{2}\log_2\left(2\pi e\,\mathbb{E}[U^2]\right)$

3. **[Typical Set in the Continuous Setting]**

   Let $X_1, \dots, X_n$ be i.i.d. continuous random variables with density function f_X. Similarly to the discrete setting, we can define the *typical set*:

$$\mathcal{T}_n(\epsilon) = \left\{ \mathbf{x} \in \mathbb{R}^n \,:\, 2^{-n(h(X)+\epsilon)} \le f_{\mathbf{X}}(\mathbf{x}) \le 2^{-n(h(X)-\epsilon)} \right\},$$

   where $f_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^n f_X(x_i)$. Prove the following properties:

   (a) (Equivalent definition) We have $\mathbf{x} \in \mathcal{T}_n(\epsilon)$ if and only if

$$h(X) - \epsilon \le \frac{1}{n}\sum_{i=1}^n \log_2 \frac{1}{f_X(x_i)} \le h(X) + \epsilon.$$

(b) (High probability) $\mathbb{P}[\mathbf{X} \in \mathcal{T}_n(\epsilon)] \to 1$ as $n \to \infty$.

(c) (Volume upper bound) $\mathrm{Vol}(\mathcal{T}_n(\epsilon)) \le 2^{n(h(X)+\epsilon)}$.

(d) (Volume lower bound) $\mathrm{Vol}(\mathcal{T}_n(\epsilon)) \ge (1 - o(1))2^{n(h(X)-\epsilon)}$, where $o(1)$ represents a term that vanishes as $n \to \infty$.

Here the volume of a set $\mathcal{A}$ is defined as $\mathrm{Vol}(\mathcal{A}) = \int_{\mathbf{x} \in \mathcal{A}} d\mathbf{x}$.

4. **[Fading Channel]**

Consider a model where we have additive *and* multiplicative noise:

$$Y = XV + Z,$$

where Z and V are both "noise" random variables. The presence of V is often referred to as *fading*, with applications in wireless communication.

Suppose that X and V are independent. Argue that knowledge of V increases the channel capacity by showing that $I(X; Y|V) \ge I(X; Y)$.

5. **[Discrete-Input Continuous-Output Channel]**

Consider the additive channel $Y = X + Z$ with $Z \sim \mathrm{Uniform}[0, a]$ for some $a > 1$, and where X can only take values in $\{0, 1\}$ (discrete inputs). The channel capacity formula $C = \max_{P_X} I(X; Y)$ still holds in this case, where $I(X; Y) = H(X) - H(X|Y)$ (in terms of regular entropy) and also $I(X; Y) = h(Y) - h(Y|X)$ (in terms of differential entropy).

(Hint: The assumption $a > 1$ is very important. When can we say that $H(X|Y = y) = 0$?)

(a) Let $X = 1$ with probability p, and therefore $X = 0$ with probability $1 - p$. Calculate $H(X)$ and $H(X|Y)$, and deduce an expression for $I(X; Y)$ *(Note: As an optional extra, you could try doing this also via $h(Y)$ and $h(Y|X)$)*

(b) Maximize your answer from part (a) over p to deduce the channel capacity.

6. **[Parallel Gaussian Channel]**

Consider a channel with two inputs (X_1, X_2) and two outputs (Y_1, Y_2), where:

- $Y_1 = X_1 + Z_1$ with $Z_1 \sim N(0, 1)$;
- $Y_2 = X_2 + Z_2$ with $Z_2 \sim N(0, 10)$, and Z_2 is independent of Z_1.

Since the noises are independent, we can think of this as transmitting information separately over two "sub-channels": One from X_1 to Z_1, and the other from $X_1 \to Z_2$. However, we only have an *overall* power constraint of P (i.e., we have some freedom in how to allocate powers $P_1 \ge 0$ and $P_2 \ge 0$ to the two sub-channels, but these must satisfy $P_1 + P_2 \le P$).

Prof. Smith states the following: "The second channel is a lot noisier, so there is no point in using it. Setting $P_1 = P$ and $P_2 = 0$ is the best we can ever hope to do, regardless of the value of P.". Is he correct?

7. **(Advanced) [Two-Look Gaussian Channel]**

Consider the Gaussian channel with two correlated looks at X; specifically, we have $Y = (Y_1, Y_2)$, where

$$Y_1 = X + Z_1$$
$$Y_2 = X + Z_2,$$

where (Z_1, Z_2) are jointly Gaussian with mean zero, equal variance $\mathbb{E}[Z_1^2] = \mathbb{E}[Z_2^2] = N$, and correlation $\mathbb{E}[Z_1 Z_2] = \sigma^2 \rho$ for some correlation coefficient $\rho \in [-1, 1]$.

(a) Show that the channel capacity is $\frac{1}{2}\log_2\left(1+\frac{2P}{\sigma^2(1+\rho)}\right)$.

(Hint: You may use that the capacity-achieving input distribution is $X \sim N(0,P)$, even if you don't prove it. Also note that we gave a formula for the differential entropy of a multivariate Gaussian in the lecture – you may use the fact that in the bivariate case with covariance matrix $\begin{bmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{bmatrix}$*, it simplifies to*

$$h(Z_1, Z_2) = \frac{1}{2}\log_2((2\pi e)^2\sigma^4(1-\rho^2))$$

without having to prove this. The identity $1-\rho^2 = (1-\rho)(1+\rho)$ is also useful.)

(b) Specialize the answer in (a) to $\rho = -1$, $\rho = 0$, and $\rho = 1$, and try to interpret the capacities obtained.

8. **[Infinite Capacity]**

Consider the additive channel $Y = X + Z$ with the usual power constraint $\mathbb{E}[X^2] \le P$, but with a not-so-usual noise distribution:

$$Z = \begin{cases} Z_{\text{Gaussian}} & \text{with probability } 0.9 \\ 0 & \text{with probability } 0.1, \end{cases}$$

where $Z_{\text{Gaussian}} \sim N(0,\sigma^2)$. By studying $\max_{P_X : \mathbb{E}[X^2]\le P} I(X;Y)$, it can be shown that the channel capacity is infinite. Give an alternative proof of this fact by describing a simple coding scheme to transmit infinitely many bits.

(Hint: (i) A single real number can carry infinitely many bits, since its binary expansion goes on indefinitely such as 0.0110101011110101010...*; (ii) If U is a Gaussian random variable with non-zero variance, then $\mathbb{P}[U = u] = 0$ for any u (whereas $\mathbb{P}[u_1 \le U \le u_2] > 0$ provided that $u_2 > u_1$))*

Hints

1. Hints given in the question.

2. It is useful to note that entropy is non-negative but differential entropy may be negative. Also revise the properties of differential entropy proved in the lecture. For further help, the answers are given as follows, and you are left to work out why: T/F/T/F/F/T

3. Use the analysis for the discrete notion of typicality (see the lectures) as a template.

4. Use the independence assumption in the question, and the fact that conditioning reduces entropy.

5. Argue that $H(X|Y = y) > 0$ only for $y \in [1, a]$, and show that for any such y, $\mathbb{P}[X = 1|Y = y] = p$.

6. The overall rate is a sum of the two sub-channel rates, each of which can be characterized using the Gaussian capacity formula. Try plugging in some numbers (both with $P_2 = 0$ and $P_2 > 0$) when the overall power P is large (e.g., $P = 100$ is sufficient).

7. Hints given in the question.

8. Follow the hint, and transmit the same value on every channel use. Argue that if (at the channel output) we see the exact same symbol more than once, we know it must be exactly what was transmitted, and we can then read off the infinitely many bits.