ASSIGNMENT 5

Define the *differential entropy* $h(X)$ of a continuous random variable X with density $f(x)$ as

$$h(X) = -\int_{-\infty}^{\infty} f(x)\log f(x)dx,$$

if the integral exists. The conditional differential entropy $h(X|Y)$ is defined analogously.

Exercise 1. Calculate the differential entropy for the following distributions:

a. Uniform distribution on $[0, a]$, $a > 0$.

b. Gaussian distribution $\mathcal{N}(0, \sigma^2)$.

Is $h(X)$ always non-negative? Provide a proof or a counterexample.

Exercise 2. (Scaling and translation) For c a constant, how are $h(cX)$ and $h(X + c)$ related to $h(X)$?

Exercise 3. (Relation to discrete entropy) Consider a random variable X with density $f(x)$. Divide the range of X into consecutive segments of length Δ. Assume that the density is continuous within the segments. By the mean value theorem, there exists a value x_i within each segment i such that

$$f(x_i)\Delta = \int_{i\Delta}^{(i+1)\Delta} f(x)dx.$$

Consider the quantized random variable X^Δ, defined by $X^\Delta = x_i$ if $i\Delta \leq X < (i + 1)\Delta$.

a. Calculate the (discrete) entropy $H(X^\Delta)$.

b. Conclude that under suitable conditions[1], as $\Delta \to 0$,

$$H(X^\Delta) + \log\Delta \to h(X).$$

c. Interpret the result as: *the entropy of an n-bit quantization of a continuous random variable X is approximately* $h(X) + n$ by considering $X \sim$ Unif $[0, 1]$ and $X \sim \mathcal{N}(0, 1)$.

Exercise 4. (KL divergence) Define the KL divergence between two densities f and g as

$$D(f||g) = \int f(x)\log\frac{f(x)}{g(x)}dx.$$

a. Using Jensen's inequality, prove that $D(f||g)$ is always non-negative.

b. Show that for a random variable $X \sim f$ with variance σ^2,

$$h(X) \leq \frac{1}{2}\log 2\pi e\sigma^2$$

with equality if and only if X is a Gaussian random variable with variance σ^2.
Hint – Calculate the KL divergence between f and the Gaussian density.

[1] If $f(x)\log f(x)$ is Riemann integrable

Exercise 5. (Mutual information) Define the mutual information between continuous random variables X and Y with joint distribution $f_{XY}(x,y)$ and marginals $f_X(x)$ and $f_Y(y)$ as

$$I(X;Y) = D(f_{XY}||f_X f_Y).$$

a. Show that $I(X;Y) = h(Y) - h(Y|X)$.

b. Consider independent random variables X and Z with $Z \sim \mathcal{N}(0,N)$ and $\mathbb{E}[X^2] \leq P$. Let $Y = X + Z$. Show that

$$C \triangleq \max_{f(x):\mathbb{E}X^2 \leq P} I(X;Y) = \frac{1}{2}\log\left(1+\frac{P}{N}\right). \tag{1}$$

Hint – Prove the inequality (without the $\max$) first and exhibit an example distribution of X (Gaussian?) for which the inequality becomes an equality.

Exercise 6. (AEP for continuous random variables) Define the volume of a set $A \subset \mathbb{R}^n$ as

$$Vol(A) = \int_A dx_1 dx_2 \cdots dx_n.$$

For $\epsilon > 0$ and any n, define the *typical set* $A_\epsilon^{(n)}$ with respect to $f(x)$ as follows:

$$A_\epsilon^{(n)} = \left\{(x_1,\ldots,x_n) : \left|-\frac{1}{n}\log f(x_1,\ldots,x_n) - h(X)\right| \leq \epsilon\right\},$$

where $f(x_1,\ldots,x_n) = \prod_{i=1}^n f(x_i)$.

a. Prove the following for a typical set.

1. $\mathbb{P}(A_\epsilon^{(n)}) > 1-\epsilon$ for n sufficiently large.
2. $Vol(A_\epsilon^{(n)}) \leq 2^{n(h(X)+\epsilon)}$.
3. $Vol(A_\epsilon^{(n)}) \geq (1-\epsilon)2^{n(h(X)-\epsilon)}$ for n sufficiently large.

b. Do the arguments above extend to joint distributions? Define the typical set $A_\epsilon^{(n)}$ with respect to $f_{XY}(x,y)$ (with marginals f_X and f_Y) as

$$\begin{aligned} A_\epsilon^{(n)} = \Bigg\{(x^n,y^n) : &\left|-\frac{1}{n}\log f_X(x^n) - h(X)\right| \leq \epsilon, \ \left|-\frac{1}{n}\log f_Y(y^n) - h(Y)\right| \leq \epsilon, \\ &\left|-\frac{1}{n}\log f_{XY}(x^n,y^n) - h(X,Y)\right| \leq \epsilon\Bigg\}. \end{aligned}$$

Prove the following: If $(\overline{X}^n, \overline{Y}^n) \sim f_X(x^n)f_Y(y^n)$, then

$$\mathbb{P}(\overline{X}^n, \overline{Y}^n) \in A_\epsilon^{(n)}) \leq 2^{-n(I(X;Y)-3\epsilon)}.$$

c. If X_i are drawn i.i.d. from a distribution f such that $\mathbb{E}X_i^2 \leq P - \epsilon$ where $P - \epsilon > 0$, argue that the probability of the event

$$E_0 = \left\{\frac{1}{n}\sum_{i=1}^n X_i^2 > P\right\}$$

goes to 0 as $n \to \infty$.

Exercise 7. (Achievability for Gaussian channels) Consider a time-discrete channel with output Y_i at time i, where Y_i is the sum of the input X_i and noise Z_i independent of X_i with $Z_i \sim i.i.d.\ \mathcal{N}(0, N)$. If there is a *power constraint*, namely, for any codeword $(x_1, x_2, \ldots, x_n)$ transmitted over the channel, we require that

$$\frac{1}{n}\sum_{i=1}^{n} x_i^2 \leq P.$$

Following the arguments in the proof of achievability in the discrete channel coding theorem (and the previous exercise), show that the maximum rate of communication over this channel, $R > C - \epsilon$ for every $\epsilon > 0$ where C is as defined in (1).