EEE 551 Information Theory (Spring 2022)

Homework 4 Solutions

---

1. Problem 8.1 from Cover-Thomas: *Differential entropy.* Evaluate the differential entropy $h(X) = -\int f \log f$ for the following:
   (a) The exponential density, $f(x) = \lambda e^{-\lambda x},\ x \geq 0$.
   (b) The Laplace density, $f(x) = \frac{1}{2}\lambda e^{-\lambda|x|}$.
   (c) The sum of X_1 and X_2, where X_1 and X_2 are independent normal random variables with means μ_i and variances $\sigma_i^2, i = 1, 2$.

   Solution:

   (a) For the exponential density:
   $$h(X) = -\mathbb{E}[\log f(X)] = -\mathbb{E}[\log(\lambda e^{-\lambda X})] = -\log\lambda + \lambda\log(e)\mathbb{E}(X) = -\log\lambda + \lambda\log(e)\frac{1}{\lambda} = \log\frac{e}{\lambda}.$$

   (b) For the Laplace density:
   $$\begin{aligned} h(X) &= -\int_{-\infty}^{\infty} \frac{1}{2}\lambda e^{-\lambda|x|} \log\left(\frac{1}{2}\lambda e^{-\lambda|x|}\right) dx \\ &= -\int_{\infty}^{0} \frac{1}{2}\lambda e^{\lambda x} \log\left(\frac{1}{2}\lambda e^{\lambda x}\right) dx - \int_{0}^{\infty} \frac{1}{2}\lambda e^{-\lambda x} \log\left(\frac{1}{2}\lambda e^{-\lambda x}\right) dx \\ &= -\int_{0}^{\infty} \lambda e^{-\lambda x} \log\left(\frac{1}{2}\lambda e^{-\lambda x}\right) dx \\ &= \int_{0}^{\infty} \lambda e^{-\lambda x} dx - \int_{0}^{\infty} \lambda e^{-\lambda x} \log\left(\lambda e^{-\lambda x}\right) dx \\ &= 1 - \int_{0}^{\infty} \lambda e^{-\lambda x} \log\left(\lambda e^{-\lambda x}\right) dx. \end{aligned}$$

   Note that the second term above is precisely the differential entropy of the exponential density. Thus for the Laplace density $h(X) = 1 + \log\frac{e}{\lambda}$.

   (c) $X_1 + X_2$ is itself normal with mean $\mu_1 + \mu_2$ and variance $\sigma_1^2 + \sigma_2^2$. Thus
   $$h(X_1 + X_2) = \frac{1}{2}\log 2\pi e(\sigma_1^2 + \sigma_2^2).$$

2. *Differential entropy.* Consider the continuous variables X, Y with joint PDF given by
   $$f(x, y) = \begin{cases} 2, & 0 < x < 1,\ 0 < y < 1,\ x + y < 1, \\ 0, & \text{otherwise.} \end{cases}$$

   Find $h(X)$, $h(Y)$, $h(X, Y)$, $h(X|Y)$, $h(Y|X)$, and $I(X;Y)$.

   Solution: To evaluate $h(X)$, we first need to find the PDF of X: for $0 < x < 1$,
   $$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f(x, y)\, dy \\ &= \int_{0}^{1-x} 2\, dy \\ &= 2(1 - x). \end{aligned}$$

Thus,

$$f_X(x) = \begin{cases} 2(1-x), & 0 < x < 1 \\ 0, & \text{otherwise.} \end{cases}$$

The differential entropy is now

$$\begin{aligned} h(X) &= \int_{-\infty}^{\infty} -f_X(x) \log f_X(x)\, dx \\ &= \int_0^1 -2(1-x) \log(2(1-x))\, dx \\ &= \int_0^1 -2(1-x) \left[1 + \log(1-x)\right]\, dx \\ &= -1 - 2\int_0^1 (1-x)\log(1-x)\, dx \\ &= -1 - 2\int_0^1 z \log(z)\, dz \end{aligned}$$

where we have rewritten the integral by the change of variables $z = 1-x$. To evaluate this integral with respect to z, we use integration by parts with the following identifications: $u = \log(z)$, $dv = z\, dz$, so $du = \frac{\log e}{z} dz$, $v = \frac{z^2}{2}$. Thus

$$\begin{aligned} \int_0^1 z\log(z)\, dz = \int_0^1 u\, dv = uv\Big|_0^1 - \int_0^1 v\, du = \frac{z^2}{2}\log z\Big|_0^1 - \int_0^1 \frac{z^2}{2}\frac{\log e}{z}\, dz \\ = -\frac{\log e}{2}\int_0^1 z\, dz = -\frac{\log e}{2}\frac{z^2}{2}\Big|_0^1 = -\frac{\log e}{4}. \end{aligned}$$

Thus

$$h(X) = -1 - 2\left(-\frac{\log e}{4}\right) = -1 + \frac{\log e}{2} \approx -0.2787.$$

Because the distribution is symmetric between X and Y, we also have

$$h(Y) = -1 + \frac{\log e}{2} \approx -0.2787.$$

For the joint differential entropy, we have

$$h(X,Y) = \mathbb{E}[-\log f(X,Y)] = \mathbb{E}[-\log 2] = -1.$$

The remaining quantities can be calculated from what we have already determined:

$$\begin{aligned} h(X|Y) &= h(X,Y) - h(Y) = -1 - \left(-1 + \frac{\log e}{2}\right) = -\frac{\log e}{2} \approx -0.7213, \\ h(Y|X) &= h(X,Y) - h(X) = -1 - \left(-1 + \frac{\log e}{2}\right) = -\frac{\log e}{2} \approx -0.7213, \\ I(X;Y) &= h(X) - h(X|Y) = \left(-1 + \frac{\log e}{2}\right) - \left(-\frac{\log e}{2}\right) = -1 + \log e \approx 0.4427. \end{aligned}$$

3. *Gaussian typical set.* Let $A_\epsilon^{(n)}$ be the typical set for the zero-mean Gaussian distribution $\mathcal{N}(0, \sigma^2)$. Find constants a and b such that $x^n \in A_\epsilon^{(n)}$ if and only if

$$a \leq \sum_{i=1}^n x_i^2 \leq b.$$

Solution: A sequence $x^n \in \mathbb{R}^n$ is in $A_\epsilon^{(n)}$ if

$$\left| -\frac{1}{n} \log f(x^n) - h(X) \right| \leq \epsilon.$$

The differential entropy is $h(X) = \frac{1}{2}\log(2\pi e \sigma^2)$. The joint PDF is

$$f(x^n) = \prod_{i=1}^n f(x_i) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x_i^2/(2\sigma^2)} = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\sum_{i=1}^n x_i^2/(2\sigma^2)}.$$

Thus

$$\begin{aligned} -\frac{1}{n}\log f(x^n) &= -\frac{1}{n}\log\left[\frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\sum_{i=1}^n x_i^2/(2\sigma^2)}\right] \\ &= \frac{1}{n}\left[\frac{n}{2}\log(2\pi\sigma^2) + \frac{\log e}{2\sigma^2}\sum_{i=1}^n x_i^2\right] \\ &= \frac{1}{2}\log(2\pi\sigma^2) + \frac{\log e}{2\sigma^2 n}\sum_{i=1}^n x_i^2 \end{aligned}$$

so

$$-\frac{1}{n} f(x^n) - h(X) = \frac{1}{2}\log(2\pi\sigma^2) + \frac{\log e}{2\sigma^2 n}\sum_{i=1}^n x_i^2 - \frac{1}{2}\log(2\pi e\sigma^2) = \frac{\log e}{2\sigma^2 n}\sum_{i=1}^n x_i^2 - \frac{1}{2}\log e.$$

Therefore $x^n \in A_\epsilon^{(n)}$ if

$$-\epsilon \leq \frac{\log e}{2\sigma^2 n}\sum_{i=1}^n x_i^2 - \frac{1}{2}\log e \leq \epsilon.$$

This condition is equivalent to

$$\frac{1}{2}\log e - \epsilon \leq \frac{\log e}{2\sigma^2 n}\sum_{i=1}^n x_i^2 \leq \frac{1}{2}\log e + \epsilon$$

or

$$\sigma^2 n - \frac{2\sigma^2\epsilon n}{\log e} \leq \sum_{i=1}^n \leq \sigma^2 n + \frac{2\sigma^2 \epsilon n}{\log e}.$$

Therefore

$$a = \sigma^2 n - \frac{2\sigma^2\epsilon n}{\log e}, \qquad b = \sigma^2 n + \frac{2\sigma^2\epsilon n}{\log e}.$$

4. *Binary-input Gaussian noise channel.* Consider a channel where $X \in \{-\sqrt{P}, \sqrt{P}\}$, and $Y = X + Z$ where $Z \sim \mathcal{N}(0, N)$. This can be considered a model for binary phase-shift keying (BPSK) with power P used for a Gaussian noise channel. Assume that X is equally likely to be $\sqrt{P}$ or $-\sqrt{P}$.

   (a) Find a closed-form expression for $h(Y|X)$.

   (b) Find a formula for $h(Y)$. You do not need to evaluate the integral.

   (c) Use some numerical integration software (such as Matlab's function `integral`) to calculate $I(X;Y) = h(Y) - h(Y|X)$ for the following parameters: $P = 2$, $N = 3$. Compare your answer to the capacity of the standard Gaussian channel with the same power and noise variance.

   Solution:

   (a) We may write

$$h(Y|X) = h(X+Z|X) = h(Z|X) = h(Z) = \frac{1}{2}\log(2\pi e N).$$

(b) The PDF of Y can be written

$$\begin{aligned}
f(y) &= \frac{1}{2} f_{Y|X}(y|\sqrt{P}) + \frac{1}{2} f_{Y|X}(y|-\sqrt{P}) \\
&= \frac{1}{2\sqrt{2\pi N}} e^{-(y-\sqrt{P})^2/(2N)} + \frac{1}{2\sqrt{2\pi N}} e^{-(y+\sqrt{P})^2/(2N)} \\
&= \frac{1}{2\sqrt{2\pi N}} \left(e^{-\frac{(y-\sqrt{P})^2}{2N}} + e^{-\frac{(y+\sqrt{P})^2}{2N}} \right).
\end{aligned}$$

Thus the differential entropy is

$$\begin{aligned}
h(Y) &= \int_{-\infty}^{\infty} -f(y) \log f(y)\, dy \\
&= \int_{-\infty}^{\infty} -\frac{1}{2\sqrt{2\pi N}} \left(e^{-\frac{(y-\sqrt{P})^2}{2N}} + e^{-\frac{(y+\sqrt{P})^2}{2N}} \right) \log \left[\frac{1}{2\sqrt{2\pi N}} \left(e^{-\frac{(y-\sqrt{P})^2}{2N}} + e^{-\frac{(y+\sqrt{P})^2}{2N}} \right) \right] dy.
\end{aligned}$$

(c) The following Matlab code computes $I(X;Y)$:

```
P = 2;
N = 3;
hY = integral(@(y)-1/2/sqrt(2*pi*N)*(exp(-(y-sqrt(P)).^2/(2*N))+...
    exp(-(y+sqrt(P)).^2/(2*N))).*...
    log2(1/2/sqrt(2*pi*N)*(exp(-(y-sqrt(P)).^2/(2*N))+...
    exp(-(y+sqrt(P)).^2/(2*N)))),-20,20);
I = hY-1/2*log2(2*pi*exp(1)*N);
```

Note that the integral is only computed from -20 to 20. Even though the integral should in principle be computed from $-\infty$ to ∞, Matlab has some numerical issues dealing with the full line; limiting it to a finite interval avoids the numerical issues, while barely changing the integral value.

This code produces the result $I(X;Y) = 0.3638$. The Gaussian channel capacity is $C = \frac{1}{2}\log(1 + \frac{P}{N}) = 0.3685$. This is almost the same, but slightly larger.

5. *Mutual information with mixed random variable.* Let $X \sim \text{Bern}(1/2)$. Define a channel from X to Y as follows. Given $X = x$, Y is a mixed random variable with probability $1-p$ of being equal to x, and with probability p of being drawn from a continuous uniform distribution on the interval $(0,1)$. Find $I(X;Y)$.

Solution: The general definition of mutual information is

$$I(X;Y) = \sup_{\mathcal{P},\mathcal{Q}} I([X]_{\mathcal{P}}; [Y]_{\mathcal{Q}})$$

where $\mathcal{P}$ and $\mathcal{Q}$ are partitions, and $[X]_{\mathcal{P}}, [Y]_{\mathcal{Q}}$ are quantized variables. Since X is already discrete, we do not need a quantization, so in fact

$$I(X;Y) = \sup_{\mathcal{Q}} I(X; [Y]_{\mathcal{Q}}).$$

Since Y only takes values in the closed interval $[0,1]$, we may assume $\mathcal{Q}$ is a partition of $[0,1]$. Consider a partition $\mathcal{Q}$ of the form

$$(\{0\}, Q_2, \ldots, Q_{K-1}, \{1\})$$

where $Q_1, \ldots, Q_{K-1}$ are a partition of the open interval $(0,1)$. To determine $I(X;[Y]_{\mathcal{Q}})$, we need to find the conditional PMF of $[Y]_{\mathcal{Q}}$ given X. We have

$$\begin{aligned}
&\Pr\{Y=1|X=1\} = 1-p, \quad \Pr\{Y=0|X=0\} = 0, \\
&\Pr\{Y=1|X=0\} = 0, \quad \Pr\{Y=0|X=0\} = 1-p.
\end{aligned}$$

Moreover, for any $i \in \{2, \ldots, K\}$, and any $x \in \{0, 1\}$

$$\Pr\{Y \in Q_i | X = x\} = p \Pr\{U \in Q_i\}$$

where U is a uniform random variable on $(0, 1)$. Define for convenience $q_i = \Pr\{U \in Q_i\}$. We have

$$H([Y]_{\mathcal{Q}}|X) = H(1-p, p\, q_2, \ldots, p\, q_{K_1}) = H(p) + pH(q_2, \ldots, q_{K-1}).$$

Moreover

$$H([Y]_{\mathcal{Q}}) = H(\tfrac{1-p}{2}, \tfrac{1-p}{2}, p\, q_2, \ldots, p\, q_{K_1}) = H(p) + 1 - p + pH(q_2, \ldots, q_{K-1}).$$

Thus

$$I(X; [Y]_{\mathcal{Q}}) = H([Y]_{\mathcal{Q}}) - H([Y]_{\mathcal{Q}}|X) = 1 - p.$$

That is, the mutual information for the quantized Y actually does not depend on the specifics of the sets $Q_2, \ldots, Q_{K-1}$. This implies that

$$I(X; Y) = 1 - p.$$

6. Problem 9.2 from Cover-Thomas: *Two-look Gaussian channel.* Consider the ordinary Gaussian channel with two correlated looks at X, that is, $Y = (Y_1, Y_2)$, where

$$\begin{aligned} Y_1 &= X + Z_1 \\ Y_2 &= X + Z_2 \end{aligned}$$

with a power constraint P on X, and $(Z_1, Z_2) \sim \mathcal{N}_2(0, K)$, where

$$K = \begin{bmatrix} N & N\rho \\ N\rho & N \end{bmatrix}.$$

Find the capacity C for

(a) $\rho = 1$
(b) $\rho = 0$
(c) $\rho = -1$

Solution:

(a) If $\rho = 1$, then $Z_1 = Z_2$. Thus the second look provides no additional information, so the capacity is just that of the ordinary Gaussian channel, *i.e.* $C = \frac{1}{2}\log(1 + \frac{P}{N})$.

(b) If $\rho = 0$, then Z_1 and Z_2 are independent. We have

$$\begin{aligned} I(X; Y) &= h(Y_1, Y_2) - h(Y_1, Y_2|X) = h(Y_1, Y_2) - h(Z_1, Z_2) \\ &= h(Y_1, Y_2) - h(Z_1) - h(Z_2) = h(Y_1, Y_2) - \frac{1}{2}\log(2\pi e N)^2. \end{aligned}$$

To calculate $h(Y_1, Y_2)$, let $\sigma_x^2 = \mathrm{Var}(X)$ (certainly $\sigma_x^2 \le P$, but equality may not be optimal). Now $\mathrm{Var}(Y_i) = \mathrm{Var}(X) + \mathrm{Var}(Z_i) = \sigma_x^2 + N$, and that $\mathbb{E}(Y_1 Y_2) = \sigma_x^2$. Thus the covariance matrix of (Y_1, Y_2) is

$$\begin{bmatrix} \sigma_x^2 + N & \sigma_x^2 \\ \sigma_x^2 & \sigma_x^2 + N \end{bmatrix}.$$

The differential entropy of (Y_1, Y_2) is upper bounded by the Gaussian with the same covariance; that is

$$h(Y_1, Y_2) \le \frac{1}{2}\log(2\pi e)^2[(\sigma_x^2 + N)^2 - \sigma_x^4] = \frac{1}{2}\log(2\pi e)^2[N^2 + 2\sigma_x^2 N] \le \frac{1}{2}\log(2\pi e)^2[N^2 + 2PN]$$

with equality if $X \sim \mathcal{N}(0, P)$. Thus

$$C = \frac{1}{2}\log(2\pi e)^2[N^2 + 2PN] - \log(2\pi e N)^2 = \frac{1}{2}\log\left(1 + \frac{2P}{N}\right).$$

(c) If $\rho = -1$, then $Z_2 = -Z_1$. In particular, $Y_1 = X + Z_1$ and $Y_2 = X - Z_1$, so $\frac{1}{2}(Y_1 + Y_2) = X$. That is, X can be recovered perfectly from $Y = (Y_1, Y_2)$, so $C = \infty$.

7. Let X be a nonnegative continuous random variable with pdf $g(x)$ and mean μ. Let Y be a random variable with exponential density and mean μ (i.e. $f(y) = \frac{1}{\mu}e^{-y/\mu}$, $y \geq 0$). Show that $h(X) \leq h(Y)$. (Hint: Evaluate the relative entropy $D(g\|f)$, where f is the pdf of Y.)

   Solution: Since Y has exponential density and mean μ, $f(y) = \frac{1}{\mu}e^{-y/\mu}$. Thus

$$\begin{aligned}
0 \leq D(g\|f) &= \int_0^\infty g(x) \log \frac{g(x)}{f(x)} dx \\
&= \int_0^\infty g(x) \log g(x) - \int_0^\infty g(x) \log f(x) dx \\
&= -h(X) - \int_0^\infty g(x) \log \frac{1}{\mu} e^{-x/\mu} dx \\
&= -h(X) - \int_0^\infty g(x) \left[-\log \mu - \frac{x}{\mu} \log e \right] dx \\
&= -h(X) - \left[-\log \mu - \frac{\mathbb{E}(X)}{\mu} \log e \right] \\
&= -h(X) + \log \mu + \frac{\mu}{\mu} \log e \\
&= -h(X) + \log e\mu.
\end{aligned}$$

   Hence $h(X) \leq \log e\mu$. Since $h(Y) = \log e\mu$ (see problem 8.1 from Cover-Thomas), $h(X) \leq h(Y)$.

8. Problem 9.4 from Cover-Thomas: *Exponential noise channels.* $Y_i = X_i + Z_i$, where Z_i is i.i.d. exponentially distributed noise with mean μ. Assume that we have a mean constraint on the signal (i.e. $EX_i \leq \lambda$). Show that the capacity of such a channel is $C = \log(1 + \frac{\lambda}{\mu})$. (**Note**: This problem is misstated in the book. There is an additional constraint that the input to the channel is nonnegative, i.e. $X_i \geq 0$.)

   Solution: We have

$$I(X;Y) = h(Y) - h(Y|X) = h(Y) - h(Z) = h(Y) - \log e\mu.$$

   We need to find the distribution on X that maximizes $h(Y)$. We know that $\mathbb{E}(Y) = \mathbb{E}(X) + \mathbb{E}(Z) \leq \lambda + \mu$. Thus, by problem 5, we have that $h(Y) \leq \log e(\lambda + \mu)$, with equality if Y is exponential. To prove that this bound is achievable, we must show that there is a distribution on X so that $X + Z$ is exponential with mean $\lambda + \mu$. Note that an exponential X will not work, since the sum of two exponentials is not exponential. The easiest way to find the appropriate distribution on X is through moment generating functions. If Z and Y are both exponential with means μ and $\lambda + \mu$ respectively, then

$$\begin{aligned}
\mathbb{E}[e^{tZ}] &= \frac{1}{1 - \mu t} \\
\mathbb{E}[e^{tY}] &= \frac{1}{1 - (\lambda + \mu)t}.
\end{aligned}$$

   Since X and Z are independent, $\mathbb{E}[e^{tY}] = \mathbb{E}[e^{t(X+Z)}] = \mathbb{E}[e^{tX}]\mathbb{E}[e^{tZ}]$. Thus

$$\mathbb{E}[e^{tX}] = \frac{\mathbb{E}[e^{tY}]}{\mathbb{E}[e^{tZ}]} = \frac{1 - \mu t}{1 - (\lambda + \mu)t} = \frac{\mu}{\lambda + \mu} + \frac{\frac{\lambda}{\lambda + \mu}}{1 - (\lambda + \mu)t}.$$

   The last expression is the moment generating function of the following distribution: Letting X' be an exponential random variable with mean $\lambda + \mu$,

$$X = \begin{cases} 0 & \text{w.p. } \frac{\mu}{\lambda+\mu} \\ X' & \text{w.p. } \frac{\lambda}{\lambda+\mu}. \end{cases}$$

We see that $\mathbb{E}(X) = \mu$, as required, and the moment generating function is given by $\mathbb{E}[e^{tX}]$ above. Thus this distribution achieves the optimal distribution on Y, so we have

$$C = \log(\lambda + \mu)e - \log e\mu = \log\left(1 + \frac{\lambda}{\mu}\right).$$