Problem Set 5 Solution

Solutions courtesy of Joy A. Thomas, with editing by Frank R. Kschischang.

8.1 *Differential entropy.*

(a) Exponential density.

$$\begin{aligned} h(X) &= -\int_0^\infty \lambda e^{-\lambda x}(\ln(\lambda) - \lambda x)\,\mathrm{d}x \\ &= -\ln(\lambda) + 1 \text{ nats.} \\ &= \log_2\left(\frac{e}{\lambda}\right) \text{ bits.} \end{aligned}$$

(b) Laplace density.

$$\begin{aligned} h(X) &= -\int_{-\infty}^\infty \frac{1}{2}\lambda e^{-\lambda|x|}(\ln(1/2) + \ln(\lambda) - \lambda|x|]\,\mathrm{d}x \\ &= -\ln\left(\frac{1}{2}\right) - \ln(\lambda) + 1 \\ &= \ln\left(\frac{2e}{\lambda}\right) \text{ nats.} \\ &= \log_2\left(\frac{2e}{\lambda}\right) \text{ bits.} \end{aligned}$$

(c) Sum of two normal distributions. The sum of two normal random variables is also normal, i.e., $X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$, and we have

$$h(X_1 + X_2) = \frac{1}{2}\log_2\left(2\pi e(\sigma_1^2 + \sigma_2^2)\right) \text{ bits.}$$

8.7 *Differential entropy bound on discrete entropy.* As in the hint, we have $Y = X' + U$, and therefore the distribution of Y has the shape of a histogram (that is, $f_Y(y) = p_i$ for $i \le y < i+1$). It is clear that $H(X') = H(X)$, since discrete entropy depends only on the probabilities and not on the values of the outcomes. Now

$$\begin{aligned} H(X') &= -\sum_{i=1}^\infty p_i \log_2 p_i \\ &= -\sum_{i=1}^\infty \left(\int_i^{i+1} f_Y(y)dy\right)\log_2\left(\int_i^{i+1} f_Y(y)dy\right) \\ &= -\sum_{i=1}^\infty \int_i^{i+1} f_Y(y)\log_2 f_Y(y)dy \\ &= -\int_1^\infty f_Y(y)\log_2 f_Y(y)dy \\ &= h(Y), \end{aligned}$$

since $f_Y(y) = p_i$ for $i \leq y < i+1$.

Hence we have the following chain of inequalities:

$$\begin{aligned}
H(X) &= H(X') \\
&= h(Y) \\
&\leq \frac{1}{2}\log_2(2\pi e \mathrm{VAR}(Y)) \\
&= \frac{1}{2}\log_2(2\pi e(\mathrm{VAR}(X') + \mathrm{VAR}(U)) \\
&= \frac{1}{2}\log_2\left(2\pi e\left(\sum_{i=1}^{\infty} p_i i^2 - \left(\sum_{i=1}^{\infty} p_i i\right)^2 + \frac{1}{12}\right)\right).
\end{aligned}$$

Since entropy is invariant with respect to permutation of $p_1, p_2, \ldots$, we can also obtain a bound by a permutation of the p_i's. We conjecture that a good bound on the variance will be achieved when the high probabilities are close together, i.e., by the assignment $\ldots, p_5, p_3, p_1, p_2, p_4, \ldots$ for $p_1 \geq p_2 \geq \ldots$.

8.8 *Channel with uniformly distributed noise.* We have

$$\begin{aligned}
C &= \max_{p(x)} I(X;Y) \\
&= \max_{p(x)} h(Y) - h(Y|X) \\
&= \max_{p(x)} h(Y) - h(X+Z|X) \\
&= \max_{p(x)} h(Y) - h(Z) \\
&= \max_{p(x)} h(Y) - \log_2 2,
\end{aligned}$$

where in the last line we have used the fact that the differential entropy of a random variable that is distributed uniformly between α and $\alpha + a$ is $\log_2 a$ bits.

Furthermore, we see that the output of the channel, Y, is limited to values in the range $[-3, 3]$. From a result on distributions with maximum entropy (specifically, see Chapter 12, Example 12.2.4), we see that $h(Y)$ will be maximized if we select $p(x)$ such that the distribution of Y is uniform in the range $[-3, 3]$.

Now, for

$$p(x=0) = p(x=2) = p(x=-2) = \frac{1}{3},$$

it is easy to see that the distribution of Y is uniform in the range $[-3, 3]$, and from our previous discussion, $h(Y) = \log_2 6$.

Therefore, we have

$$C = \log_2 6 - \log_2 2 = \log_2 3.$$

Note that one could have arrived at this result using calculus (without explicit knowledge that the uniform distribution maximizes entropy for a variable with bounded range), but it would have involved lengthy (and tedious) calculations.

8.10 *Shape of the typical set.* Since the X_i are i.i.d., $f(x_1, \ldots, x_n) = c^n e^{-(x_1^4+x_2^4+\cdots+x_n^4)}$. From the definition of the typical set, $(x_1, \ldots, x_n)$ is typical if and only if

$$2^{-n(h+\epsilon)} \leq c^n e^{-(x_1^4+x_2^4+\cdots+x_n^4)} \leq 2^{-n(h-\epsilon)},$$

which is equivalent to

$$-n(h+\epsilon)\ln 2 \leq n \ln c - (x_1^4 + x_2^4 + \cdots + x_n^4) \leq -n(h-\epsilon)\ln 2.$$

Finally, re-arranging the preceding condition, we have

$$A_\epsilon^{(n)} = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \colon n(\ln C + (h-\epsilon)\ln 2) \leq x_1^4 + x_2^4 + \cdots + x_n^4 \leq n(\ln C + (h+\epsilon)\ln 2).\}$$

This is reminiscient of the fact that the typical set of a Gaussian distribution is a thin spherical shell; in this case, the sphere has been replaced by a shape of the form $x_1^4 + x_2^4 + \cdots + x_n^4 = r$ (the surface of an $L4$ ball) but the 'thin shell' property persists.

9.3 *Output power constraint.* Since Z has zero mean and is independent of X, we have

$$E(Y^2) = E(X^2) + E(Z^2) = E(X^2) + \sigma^2,$$

and by the output power constraint, we have

$$E(X^2) \leq P - \sigma^2.$$

In the following, we assume $P > \sigma^2$, since otherwise the problem is uninteresting, as the output power constraint would be violated by the noise alone. Now, for a maximum expected output power P, the entropy of Y is maximized when $Y \sim \mathcal{N}(0, P)$, which is achieved when $X \sim \mathcal{N}(0, P - \sigma^2)$. Therefore, the channel is equivalent to one with an input power constraint $E(X^2) \leq P - \sigma^2$, and it follows that the capacity is

$$C = \frac{1}{2}\log_2\left(1 + \frac{P - \sigma^2}{\sigma^2}\right) = \frac{1}{2}\log_2\left(\frac{P}{\sigma^2}\right).$$

9.4 *Exponential noise channel.* Since

$$E(X) \leq \lambda,$$

it follows that

$$E(Y) \leq \mu + \lambda.$$

Furthermore, from the non-negativity constraints on X and Z, we also have $Y \geq 0$.

Since the noise is independent of the input, we have

$$\begin{aligned} I(X;Y) &= h(Y) - h(Y \mid X) \\ &= h(Y) - h(X + Z \mid X) \\ &= h(Y) - h(Z). \end{aligned}$$

Now, since Y is a non-negative random variable with $E(Y)) \leq \mu + \lambda$, it follows from Example 12.2.4 (Cover & Thomas) that $h(Y)$ is maximized when Y is exponentially distributed. Furthermore, from Problem 8.1, we know that the differential entropy of an exponential distribution is a monotonically increasing function of the mean. Therefore,

$$h(Y) \leq \log_2(e(\mu + \lambda)),$$

and it follows that

$$C \leq \log_2(e(\mu + \lambda)) - \log_2(e\mu)) = \log_2\left(1 + \frac{\lambda}{\mu}\right).$$

To confirm that $C = \log_2(1 + \frac{\lambda}{\mu})$, we need to show that X can be chosen such that $X + Z$ has an exponential distribution with mean $\mu + \lambda$. We will do so by explicitly deriving the required distribution.

The characteristic function of an exponential random varible Z with mean ρ is

$$\Phi_Z(\omega) = \frac{\frac{1}{\rho}}{\frac{1}{\rho} - j\omega}.$$

Since

$$\Phi_Y(\omega) = \Phi_X(\omega)\Phi_Z(\omega),$$

we have that the required $\Phi_X(\omega)$ is given by

$$\begin{aligned}
\Phi_X(\omega) &= \frac{1 - j\mu\omega}{1 - j(\mu + \lambda)\omega} \\
&= \frac{1}{1 - j(\mu + \lambda)\omega} - \left(\frac{\mu}{\mu + \lambda}\right)\left(\frac{1}{1 - j(\mu + \lambda)\omega}\right) + \frac{\mu}{\mu + \lambda}.
\end{aligned}$$

Finally, taking the inverse transform, we have

$$f_X(x) = \left(\frac{\lambda}{\mu + \lambda}\right)\left(\frac{e^{-\frac{x}{\mu+\lambda}}}{\mu + \lambda}\right) + \frac{\mu}{\mu + \lambda}\delta(x), \text{ for } x \geq 0.$$

Observe that this solution specifies that we should transmit zero with probability $\mu/(\mu + \lambda)$, and otherwise transmit an exponentially distributed nonzero value with as large a mean value as possible. (In a sense we are "saving up" for the possibility of a "louder" transmission whenever we send zero.)

9.5 *Fading channel.* We have

$$\begin{aligned}
I(X; Y \mid V) &= H(X \mid V) - H(X \mid Y, V) \\
&= H(X) - H(X \mid Y, V) \\
&\geq H(X) - H(X \mid Y) \\
&= I(X; Y),
\end{aligned}$$

where the second equality follows since X and V are independent, and the inequality follows since conditioning reduces entropy. Therefore, as is intuitively reasonable, knowledge of the fading factor improves capacity.

9.6 *Parallel channels and water-filling.* By the result of Section 10.4, it follows that we will put all the signal power into the channel with less noise until the total power of noise + signal in that channel equals the noise power in the other channel. After that, we will split any additional power evenly between the two channels.

Thus the combined channel begins to behave like a pair of parallel channels when the signal power is equal to the difference of the two noise powers, i.e., when $2P = \sigma_1^2 - \sigma_2^2$.

9.7 *Multipath Gaussian channel.*

(a) The channel output is

$$Y = 2X + Z_1 + Z_2 = 2X + Z,$$

where $Z = Z_1 + Z_2$ is the sum of two Gaussian random variables, and is thus itself Gaussian, and

$$E(Z^2) = E(Z_1^2) + 2(EZ_1Z_2) + E(Z_2^2) = 2\sigma^2(1+\rho),$$

therefore $Z \sim \mathcal{N}(0, 2\sigma^2(1+\rho))$. Similarly, since Y is itself the sum of two Gaussian random variables ($2X$ and Z), we have $Y \sim \mathcal{N}(0, 4P + 2\sigma^2(1+\rho))$. Therefore,

$$\begin{aligned} C &= \frac{1}{2}\log_2\left(\frac{4P + 2\sigma^2(1+\rho)}{2\sigma^2(1+\rho)}\right) \\ &= \frac{1}{2}\log_2\left(1 + \frac{2P}{\sigma^2(1+\rho)}\right) \end{aligned}$$

(b) When $\rho = 0$, we have $Z \sim \mathcal{N}(0, 2\sigma^2)$, and thus the noise power is doubled (relative to the case of a single noise source), while the source power is (independent of ρ) increased by a factor of 4, since two copies of X are coherently added at the receiver. Therefore,

$$C = \frac{1}{2}\log_2\left(1 + \frac{2P}{\sigma^2}\right).$$

When $\rho = 1$, we have $Z \sim \mathcal{N}(0, 4\sigma^2)$, and thus the noise power is multiplied by a factor of 4 (relative to the case of a single noise source), since $Z_1 = Z_2$ and they are added coherently at the receiver. Therefore,

$$C = \frac{1}{2}\log_2\left(1 + \frac{P}{\sigma^2}\right).$$

When $\rho = -1$, the noise is perfectly canceled, since $Z_1 = -Z_2$. Therefore, for any non-zero P, the capacity is infinitely large.

9.9 *Vector Gaussian channel.* For the vector Gaussian channel, we have

$$C = \max_{K_X} \frac{1}{2}\log_2 \frac{|K_X + K_Z|}{|K_Z|},$$

where the maximization is over all K_X satisfying $\mathrm{tr[K_X] \leq P}$. Now, for the given noise covariance matrix, we have $|K_Z| = 0$. Furthermore, from the water-filling argument in the case of parallel channels with coloured noise, the optimal K_X satisfies $|K_X + K_Z| > 0$ for any $P > 0$. Therefore, the capacity of this channel is infinitely large. Why is this so? It turns out that the given Z is of the form $Z = (Z_1, Z_2, Z_1 + Z_2)$, where $Z_1 \sim \mathcal{N}(0, 1)$ and $Z_2 \sim \mathcal{N}(0, 1)$ are independent, and the value of Z_3 is specified exactly by knowledge of Z_1 and Z_2. Therefore, the optimal coding scheme is to set $X_1 = X_2 = 0$, from which the receiver detects the values of Z_1 and Z_2, then uses these to cancel the noise in Y_3, thus providing a noise-free channel from X_3 to Y_3, which clearly permits an infinite rate of error-free communication for any $P > 0$.