EE 231A
Information Theory
Instructor: Rick Wesel

Handout #14, Problem Set 6
Tuesday May 5, 2020
Due: Tuesday, May 12, 2020

90 pts
Reading: Chapters 8 & 9 of *Elements of Information Theory*

Lecture 11: Gaussian Channel Capacity

1. (12 pts) *A truncated Gaussian.*

   (a) Prove that the Normal density $\phi(x)$ maximizes differential entropy for a fixed second moment σ^2.
   $$\phi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2} \tag{1}$$

   (b) Use this fact to show that a Gaussian input maximizes the mutual information $I(X;Y)$ on the memoryless Gaussian channel $Y = X + Z$ where Z is a Gaussian independent of X and Y.

   (c) Prove that a truncated Normal density $\tau(x)$ maximizes differential entropy for a fixed second moment under a peak limitation constraint. For example, suppose that the peak limitation constraint is that $x \in (-1, 1)$. The truncated Normal density $\tau(x)$ that maximizes the entropy for a fixed second moment σ^2 is described as follows:
   $$\tau(x) = \begin{cases} \frac{\frac{1}{\sqrt{2\pi\gamma^2}} e^{-x^2/2\gamma^2}}{K} & \text{if } x \in (-1,1) \\ 0 & \text{otherwise} \end{cases}, \tag{2}$$
   where
   $$K = \int_{-1}^{1} \frac{1}{\sqrt{2\pi\gamma^2}} e^{-x^2/2\gamma^2} dx\,, \tag{3}$$
   and γ^2 is chosen so that
   $$\frac{1}{K} \int_{-1}^{1} x^2 \frac{1}{\sqrt{2\pi\gamma^2}} e^{-x^2/2\gamma^2} dx = \sigma^2\,. \tag{4}$$

   (d) Now consider a channel that is characterized by the peak limitation $x \in (-1, 1)$ on the transmit signal and additive noise that is a truncated Normal with distribution $\tau(x)$ as described above. Is the mutual information of this channel obviously maximized by a truncated Gaussian input or not? Please explain the issues.

2. (10 pts) *Shaping Gain.*

   Consider the additive white Gaussian noise channel with power constrained input X and noise $Z \sim \mathcal{N}(0, N)$. For this channel the optimal input distribution is a Gaussian $X_g \sim \mathcal{N}(0, P)$. For practical reasons, a suboptimal X_s that is not a Gaussian but does have $EX_s^2 = P$ is often used.

   The maximum shaping gain is that performance improvement that could theoretically be obtained by using X_g rather than X_s. i.e.

   $$\text{maximum shaping gain} = I(X_g; Y_g) - I(X_s; Y_s)\,, \tag{5}$$

   where $Y_g = X_g + Z$ and $Y_s = X_s + Z$.

   Show that the maximum shaping gain may be expressed succinctly as a relative entropy involving f_s (the p.d.f. for Y_s) and f_g (the p.d.f. for Y_g).

   Lecture 12A: Binary Input Gaussian Channel

3. (10 pts) BSC vs. Gaussian Channel

   (a) Use Matlab to plot the channel capacity of the Gaussian channel and of the BSC with transition probability

   $$p = Q\left(\sqrt{\frac{P}{N}}\right) \tag{6}$$

   that is produced by a one bit quantization of the channel input and output as described in the beginning of Chapter 9. Note that

   $$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2} dt\,. \tag{7}$$

   The Y-axis of your plot should be channel capacity and the X-axis should be signal to noise (SNR) in units of dB:

   $$\text{SNR} = 10 \log_{10}\left(\frac{P}{N}\right)\,. \tag{8}$$

   Plot the SNR range between -20 dB and 30 dB. A MATLAB function computing Q is provided on the course web page.

   (b) Now suppose that the input is quantized as before, but the output is not quantized before decoding. Thus, the channel is not a binary symmetric channel, but rather akin to the discrete-input continuous-output channel of problem 2 on the previous problem set, with Z being a Gaussian and X being $\pm\sqrt{P}$.

   This is the very common Gaussian channel with a binary pulse amplitude modulation (PAM). Prove that the Gaussian channel with binary PAM has a maximum mutual information curve that lies above the capacity curve of the BSC channel of the previous part, but has the same asymptote.

(c) For completeness, plot the upper and lower bounds on the mutual information of the AWGN channel when the input is BPSK with a uniform distribution. This plot, and the related plot for 4-PSK input which is obtained by scaling this plot by 2, provide practical limits to many commonly used transmission systems. For this you should use the Matlab routines bincapup.m and bincaplo.m (which both use the subroutine Q.m). These routines were derived and implemented by Mark Shane.

Lecture 12B: Capacity of a Bandlimited Gaussian Channel

4. (12 pts) *Shannon, Sensors, A/D convertors* Fall 98 oral

(a) Write down the capacity of the discrete-time additive white Gaussian noise channel where $Z \sim \mathcal{N}(0, N)$ and transmit energy is constrained to P joules per symbol. Noise energy N is also in units of joules. X, Y, and Z are real-valued. What are the units of capacity?

(b) Write down the capacity of the continuous-time bandlimited AWGN channel with bandwidth W Hz, i.e. signal constrained to $(-W, W)$. The additive noise Z is a white Gaussian process with power spectral density $N_0/2$ watts/Hz. Transmit power is constrained to P watts. Again, what are the units of capacity?

(c) Find the capacity of the infinite bandwidth AWGN channel with noise psd and transmitter power constraint as in (b).

(d) The channel between a remote sensor and a base station is modeled by a discrete time AWGN channel with $N = 1$ joule. Give a lower bound on the amount of transmitter energy required per transmitted bit.

(e) Actually, the remote sensor has a continuous time AWGN channel to the base station, and must select a bandwidth. What is the correct choice of bandwidth in order to minimize energy per bit? How does the efficiency compare with the discrete time AWGN channel? What are the practical limitations to achieving this ultimate efficiency?

(f) An A/D converter is designed to have 1 bit of resolution at the highest possible sampling rate on a continuous AWGN channel where the noise psd is $N_0/2$ with $N_0 = 10^{-9}$ watts/Hz and the received signal power is 3 watts. This A/D converter may employ extensive processing and require extensive latency. Furthermore, the transmitted waveform is carefully designed to allow the A/D converter the best possible performance. What is the highest sampling rate that is theoretically possible while still truly achieving one bit of resolution? You may assume that the A/D is preceded by an ideal lowpass filter with cutoff frequency W at half the sampling rate.

Lecture 12C: Capacity of Parallel Gaussian Channels

5. (8 pts) *Parallel channels and waterfilling.* Consider a pair of parallel Gaussian channels, i.e.,

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} + \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}, \tag{9}$$

where

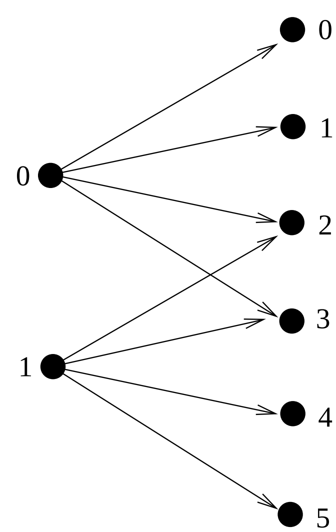
$$\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}\right), \tag{10}$$

and there is a power constraint $E(X_1^2 + X_2^2) \leq 2P$. Assume that $\sigma_1^2 > \sigma_2^2$. At what value of the transmit power constraint $2P$ does the channel stop behaving like a single channel with noise variance σ_2^2, and begin behaving like a pair of channels?

Lecture 12D: Sufficient Statistics

6. (8 pts) *Split Ends.*

Consider the channel shown below where all transition probabilities are 1/4:



(a) As usual, let X be the input and Y be the output of this channel. Define Z as the following function of Y:

$$z = \begin{cases} a & \text{if } y \in \{0, 1\}, \\ b & \text{if } y \in \{2, 3\}, \\ c & \text{if } y \in \{4, 5\}. \end{cases} \tag{11}$$

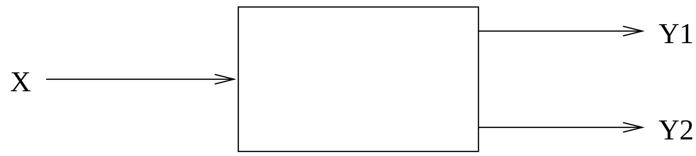
Find the capacity of the channel between X and Y.
Hint: Z is a sufficient statistic of Y for X.

(b) Prove the hint. That is, prove that Z is a sufficient statistic of Y for X.

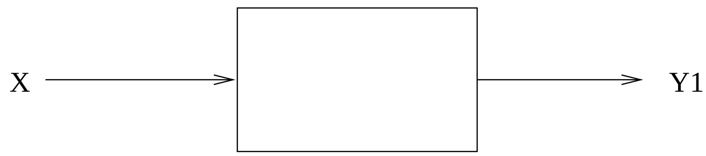
7. (10 pts) *A channel with two independent looks at Y.* Let Y_1 and Y_2 be conditionally independent and conditionally identically distributed given X.

   (a) Show $I(X; Y_1, Y_2) = 2I(X; Y_1) - I(Y_1; Y_2)$.

   (b) Conclude that the capacity of the channel



   is less than or equal to twice the capacity of the channel

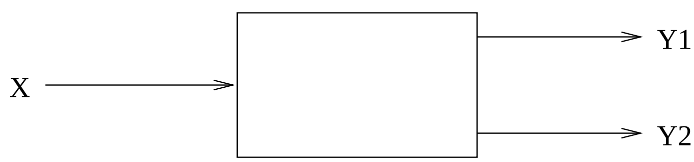


8. (10 pts) *Mutual information for correlated normals.* Find the mutual information $I(X; Y)$, where

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N_2 \left(0, \begin{bmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{bmatrix} \right).$$

   Evaluate $I(X; Y)$ for $\rho = 1, \rho = 0$, and $\rho = -1$, and comment.

9. (10 pts) *The two-look Gaussian channel.*



   Consider the ordinary Shannon Gaussian channel with two correlated looks at X, i.e., $Y = (Y_1, Y_2)$, where

$$Y_1 = X + Z_1 \qquad (12)$$

$$Y_2 = X + Z_2 \qquad (13)$$

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

$$K = \begin{bmatrix} N & N\rho \\ N\rho & N \end{bmatrix}. \qquad (14)$$

   Find the capacity C for

   (a) $\rho = 1$

   (b) $\rho = 0$

   (c) $\rho =$ -1