Problem Set #2
Due: Thursday, Oct. 20, 2016

1. *Random Time Index:* Let X^n be an i.i.d. sequence distributed according to $\prod P_X$, let Y^n be an arbitrary random sequence distributed according to P_{Y^n} (not necessarily i.i.d.), and let Z^n be correlated with Y^n according to a product distribution $P_{Z^n|Y^n}(z^n|y^n) = \prod_{i=1}^n P_{Z|Y}(z_i|y_i)$ (i.e. a memoryless channel).

   Let a random time index $T \sim \text{Unif}\{1, ..., n\}$ be independent of (X^n, Y^n, Z^n).

   (a) Show that X_T is independent of T.
   (b) Show that $X_T \sim P_X$.
   (c) Give an example of Y^n such that $I(Y_T; T) > 0$.
   (d) Show that $Y_T \sim \mathbb{E}[\pi(y|Y^n)]$.
   (e) Show that $T - Y_T - Z_T$ forms a Markov chain.
   (f) Show that $P_{Z_T|Y_T} = P_{Z|Y}$.
   (g) Show that $\mathbb{E}[g(Y_T)] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n g(Y_i)\right]$.

2. *Capacity with Input Cost Constraint:* Sometimes we wish to impose a cost constraint on a transmitter in a communication setting. For example, the capacity of the additive Gaussian noise channel would be infinite if the transmission signal had no constraint. But a practical constraint is the power of the transmission. This falls within the general framework of imposing an average cost constraint on each codeword:

$$\frac{1}{n}\sum_{t=1}^{n} c(x_t(m)) \leq C \quad \forall m, \tag{1}$$

   where $c(\cdot)$ is a cost function (e.g. $c(x) = x^2$ if the constraint is an average power constraint).

   The capacity of a memoryless channel $P_{Y|X}$ with an input cost constraint is

$$\text{Capacity} = \max_{P_X \,:\, \mathbb{E}[c(X)] \leq C} I(X; Y). \tag{2}$$

   The achievability proof for this claim uses the same construction and analysis as the channel capacity proof without a cost constraint but with P_X chosen such that $\mathbb{E}[c(X)] < C$. The additional step of the proof is to argue that with high probability most of the codewords will satisfy the average cost constraint due to the law of large numbers and the random construction of the codebook. The Markov inequality can be used to bound the fraction of the codewords in the codebook that exceed the power constraint. You then prune the codebook to not use these codewords. Because it's a small fraction, it affects the communication rate by only a negligible amount. This same technique, referred to as *expurgation*, can be used to throw out the small fraction of codewords that might have high error probability, so that the communication system has small error not just on average over randomly chosen messages but even for every individual message that the system can attempt to transmit.

   Prove the converse for (2).

3. *Source Channel Separation Theorem:* We've proven two of the most fundamental results in information theory. Channel capacity gives the highest rate of digital communication through a noisy channel, and the rate-distortion theorem gives the optimal trade-off between signal quality and digital compression rate.

   Information theory has been a driver of the digital age that we are now in. One of the best justifications for the use of digital media comes from the source-channel separation theorem, which says that there is no penalty for dividing the overall problem into these two parts. The real objective of communication technology is to transmit a signal (stochastic process) through a communication medium (noisy channel) so that it arrives with the best quality possible. Fortunately, there is no penalty to first converting the signal into a digital object (i.e. bits) and then transmitting the digital message using a communication device that needs no understanding of the meaning of the bits. This allows engineering and technology to be modular. The compression algorithm and the network adapter do not need any knowledge of each other in order to function efficiently, and there is no loss of efficiency for using this digital interface (or, at least, the loss is negligible in the most basic setting).

   Define the problem as follows. Let S^n be an i.i.d. signal distributed according to P_S. Consider a memoryless channel governed by $P_{Y|X}$. A modulator takes S^n and produces a transmission signal X^n. A demodulator receives the transmission as Y^n and reconstructs the signal as $\hat{S}^n$. The quality of the reconstruction is measured by a distortion function $d(s, \hat{s})$. Let D^* be the minimum average distortion that can be achieved. In other words,

   $$D^* = \inf \left\{ \mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n} d(S_t, \hat{S}_t)\right] \right\}, \tag{3}$$

   where the infimum is over all blocklengths n and modems (modulators and demodulators).

   Prove that

   $$D^* = \min \left\{ D \ : \ \begin{array}{c} \exists P_{\hat{S}|S}, P_X \text{ such that} \\ I(S;\hat{S}) \leq I(X;Y), \\ \mathbb{E}\left[d(S,\hat{S})\right] \leq D. \end{array} \right\}. \tag{4}$$

   One direction of the proof (achievability or converse) will be very easy. For the easy direction, there is no need to be overly pedantic with the ϵ's and δ's or to reprove things we've already done in class. By the way, the right side of (4) can be rewritten very concisely as a sequence of four symbols.

4. *Property 7.c*:* Let $\epsilon' > \epsilon$, and let $(x^n, y^n) \in \mathcal{T}_\epsilon^{(n)}$. Let $Z^n \sim \prod_{t=1}^{n} P_{Z_t|Y_t=y_t}$. Prove that

   $$\mathbb{P}\left[(x^n, y^n, Z^n) \in \mathcal{T}_\epsilon^{(n)}\right] \leq 2^{-n(I(X;Z|Y)-4\epsilon)}, \tag{5}$$

   $$\mathbb{P}\left[(x^n, y^n, Z^n) \in \mathcal{T}_{\epsilon'}^{(n)}\right] \geq (1-\epsilon')2^{-n(I(X;Z|Y)+4\epsilon')} \text{ for } n \text{ large enough.} \tag{6}$$

   In fact, we can get 2ϵ by tightening Property 7.b as well, but no need to do that here.