PHYS 7895 Fall 2013
Introduction to Quantum Information Theory
Homework 1

Due Monday 30 September 2013, by 5pm in Nicholson 447
(You are allowed to work with others as long as you write down who your collaborators are. Any late assignments will be penalized in the amount of 25% per day late.)

This assignment has a first part and a second part.

First part: Exercises in **arXiv:1106.1445**:

2.2.1, 3.2.1, 3.3.2, 3.3.3, 3.3.4, 3.3.5, 3.3.10, 3.4.5

Second part: The following exercises:

1. Concentration inequalities:

   (a) Prove the Markov inequality. That is, for a random variable X whose realizations are non-negative, prove that
$$\Pr\{X \geq \varepsilon\} \leq \frac{\mathbb{E}\{X\}}{\varepsilon}.$$

   (b) Prove the Chebyshev inequality. That is, for any random variable with finite second moment, show that the following inequality holds:
$$\Pr\{|X - \mathbb{E}\{X\}| \geq \varepsilon\} \leq \frac{\mathrm{Var}\{X\}}{\varepsilon^2},$$
where $\mathrm{Var}\{X\} = \mathbb{E}\{|X - \mathbb{E}\{X\}|^2\}$.

   (c) Prove the following law of large numbers. For a large number of pairwise independent and identically distributed random variables $X_1, \ldots, X_n$, (such that $\mathbb{E}\{X_i\} = \mu$ and $\mathbb{E}\{|X_i - \mu|^2\} = \sigma^2$ for all $i \in \{1, \ldots, n\}$) the probability that the sample mean deviates from the true mean has a power law decay:
$$\Pr\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| \geq \varepsilon\right\} \leq \frac{\sigma^2}{\varepsilon^2 n}.$$

   (d) Prove the first part of the Chernoff-Hoeffding bound. That is, for a large number of bounded independent and identically distributed random variables $X_1, \ldots, X_n$, show that the probability that the sample mean deviates from the true mean by an additive constant (one-sided) decays exponentially with the number of samples taken:
$$\Pr\left\{\frac{1}{n}\sum_{i=1}^{n} X_i - \mu \geq \varepsilon\right\} \leq \inf_{t>0} \frac{[\mathbb{E}_X\{\exp\{tX\}\}]^n}{\exp\{t(\mu+\varepsilon)\}^n}.$$
The idea to finish it off from here is to choose t small enough so that we have
$$[\mathbb{E}_X\{\exp\{tX\}\}]/\exp\{t(\mu+\varepsilon)\} < 1$$
(implying an exponential decay with n). Bonus points for taking it from here to get the exponential decay.

2. Kebei asked a question about relating typicality and convergence towards the typical set to the central limit theorem. Zhihao asked about having the typicality tolerance δ decrease as n becomes larger (rather than taking it to be a fixed constant). We will address both of these questions with this exercise.

   Before you were born, the mathematicians (probabilists) were hard at work, trying to obtain ever finer characterizations of the convergence rate in the central limit theorem. The initial best characterizations are due to Berry and Esseen, who proved the following theorem. Let $Z_1, \ldots, Z_n$ be a sequence of i.i.d. random variables (assume finite cardinality for simplicity and each with mean μ and variance σ^2). Then the deviation of the tail of the sample mean of the normalized $Z_1, \ldots, Z_n$ from the tail of a normalized Gaussian falls off as the inverse square-root of the number of samples:

$$\left|\Pr\left\{\frac{\sum_{i=1}^{n} Z_i - \mu}{\sqrt{n\sigma^2}} \geq \delta\right\} - Q(\delta)\right| \leq \frac{C\xi}{\sigma^3\sqrt{n}},$$

   where ξ is the third central moment of each Z_i, C is a fixed positive constant that seems to keep improving, and $Q(x)$ is the tail of a standard Gaussian:

$$Q(x) \equiv \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp\{-|u|^2/2\} \, du.$$

   Let us define the set $\mathfrak{T}^{X^n}$ of "one-sided" variance-typical sequences for a distribution $p_X(x)$ to be as follows:

$$\mathfrak{T}^{X^n} \equiv \left\{x^n : -\frac{1}{n}\log(p_{X^n}(x^n)) - H(X) \leq \sqrt{\frac{V(X)}{n}} Q^{-1}(\varepsilon)\right\},$$

   where $V(X) = \text{Var}_X\{-\log p_X(X)\}$, ε is a fixed positive constant, and Q^{-1} is the inverse of the Q function. (The above definition differs from the one we defined in class just by focusing on one side of the typicality tolerance and by choosing δ to decrease with n, i.e., $\delta = \sqrt{V(X)/n}\, Q^{-1}(\varepsilon)$.)

   (a) Find an upper bound on the size of the set $\mathfrak{T}^{X^n}$.

   (b) Using the Berry-Esseen theorem, find an upper bound on the probability that a random sequence X^n falls outside the set $\mathfrak{T}^{X^n}$.

   (c) Put these two facts together and describe a data compression scheme (Shannon-style). That is, for a specified error probability ε' (and such that n is of the order $(1/\varepsilon')^2$), to how many bits can this scheme allow for us to compress a random length-n sequence?[1]

3. In class, we proved the achievability part of Shannon's channel capacity theorem by employing random coding, conditional typicality decoding, and analysis of the expectation of the error probability. We used conditional typicality so that it would be easier

[1] The Berry-Esseen theorem has been put to great use recently in many problems in both classical and quantum information theory to account for finite-size effects that get "washed away" in the limit of many instances of a resource.

for us to understand the transition from the classical theorem to the quantum theorem regarding classical data transmission over quantum channels.

You will now prove Shannon's channel capacity theorem using a different approach. Define the "simultaneous typical set" for two joint random variables $(X, Y) \sim p_{X,Y}(x, y)$ to be as follows:

$$T_\delta^{X^nY^n} \equiv \{(x^n, y^n) : \text{(A) and (B) and (C)}\}$$

where

$$\left| -\frac{1}{n} \log p_{X^n,Y^n}(x^n, y^n) - H(XY) \right| \leq \delta, \tag{(A)}$$

$$\left| -\frac{1}{n} \log p_{X^n}(x^n) - H(X) \right| \leq \delta, \tag{(B)}$$

$$\left| -\frac{1}{n} \log p_{Y^n}(y^n) - H(Y) \right| \leq \delta. \tag{(C)}$$

(a) Prove that a randomly chosen pair of sequences (X^n, Y^n) (chosen according to $p_{X^n,Y^n}(x^n, y^n)$) has a high probability of ending up in the simultaneous typical set. That is, prove that for all $\varepsilon, \delta > 0$ and sufficiently large n we have that

$$\Pr\{(X^n, Y^n) \in T_\delta^{X^nY^n}\} \geq 1 - \varepsilon.$$

(b) Prove that a randomly chosen pair of sequences $(\widetilde{X}^n, Y^n)$ (chosen according to the product of the marginals $p_{X^n}(x^n)p_{Y^n}(y^n)$ has a probability no larger than $2^{-n[I(X;Y)-3\delta]}$ of ending up in the simultaneous typical set:

$$\Pr\left\{(\widetilde{X}^n, Y^n) \in T_\delta^{X^nY^n}\right\} \leq 2^{-n[I(X;Y)-3\delta]}.$$

(c) Analyze the expectation of the error probability of a random code under the following decoding algorithm. After receiving the output sequence y^n from the channel, Bob checks whether it is simultaneously typical with the first codeword, i.e., whether $(x^n(1), y^n) \in T_\delta^{X^nY^n}$. If it is, he decodes as the first message. If not, he tests with the second codeword (i.e., if $(x^n(2), y^n) \in T_\delta^{X^nY^n}$) and proceeds doing so until the answer is "yes." What do you find is an achievable rate with this decoding algorithm?

4. We did not prove it yet, but the converse part of Shannon's channel capacity theorem is that the following formula is an upper bound on the capacity of any classical channel:

$$\max_{p_X(x)} I(X;Y),$$

where the mutual information is maximized over all input random variables X with distribution $p_X(x)$ (recall that the distribution of the channel is $p_{Y|X}(y|x)$).

(a) Can the resource of a random bit string shared between the sender and receiver help to increase capacity? Why or why not?

(b) Let the private capacity with unlimited secret key be the maximum rate at which one party can send data to another such that the data can be reliably decoded and the data is private from a potential eavesdropper. A secret key is a uniformly random bit string shared between sender and receiver and unknown to any eavesdropper. (We can say that the data is private from an eavesdropper if the joint distribution of the data and the eavesdropper is close or identical to one that factors as the product of a uniform distribution on the data and some other distribution for the eavesdropper. In the latter ideal case, the chance of guessing the message is then exponentially small in the number of bits of the message.) Does Shannon's formula above serve as an upper bound on this capacity? Why or why not?

(c) If it does serve as an upper bound, is there a way to achieve it? To answer this, consider if we add a uniformly random bit to another bit. What is the distribution of the second bit if we marginalize over the uniformly random bit?

5. Kaushik asked an interesting question in class: "If a sequence y^n is in the conditionally typical set, then shouldn't it be in the unconditionally typical set?" This is not necessarily the case when using the notions of typicality that we discussed in class, but it can be so with a different notion of typicality (strong typicality as discussed in the textbook arXiv:1106.1445). Prove that this is so. In more detail, suppose that a sequence x^n is strongly typical and a sequence y^n is strongly conditionally typical with x^n (the distributions to consider here are $p_X(x)$ and $p_{Y|X}(y|x)$). Then prove that y^n is in the strongly unconditionally typical set for $p_Y(y)$. (You may need to adjust the typicality tolerances δ as needed.)