MATH 191 FINAL EXAM SOLUTIONS

Instructions: This is a take-home exam. Please work on all problems by yourself and do not consult anyone. The exam is due at 11am on Wednesday, Dec. 10th and should be given directly to the instructor.

Problem 1. ("Guess-my-number") Someone thinks of a number x between 1 and N (you can take $N = 16$ if you wish).
(a) You may specify a subset S of $\{1, \ldots, N\}$ and ask "Is $x \in S$?" How many questions do you need to determine the value of x? (You vary the set S from one question to another, of course). Prove that your answer is optimal.
(b) Imagine that the person answering your question lies 30% of the time. Each time you ask, the probability that the person lies this time does not depend on whether he lied or told the truth before. Is there some number M so that after having asked M questions, you are guaranteed to know the correct value of x with probability of at least 99%? (c) Is there a lower bound for M in part b? (Hint: another way to explain the set-up in b is that the questions are answered truthfully but then the answer is changed with a 30% probability when given to you).

Let us denote by A_S the (truthful) answer to the question "is $x \in S$?"

For part (a), we note that the answers to our questions give rise to a binary string, and the fact that we can learn the value of x means that different sets of answers must correspond to different x. Since there are $N-1$ different x, there should be at least N different binary strings of answers. If these strings have length $\leq P$ there are $\leq 2^P$ such strings; thus $P \geq \log_2(N-1)$. If $P \geq \log_2(N-1)$ is an integer, every integer in $\{0, \ldots, N-1\}$ can be uniquely represented as a binary number with $\leq P$ digits. Take $S_j = \{\text{numbers in } \{0, \ldots, N-1\} \text{ whose } j\text{-th digit is } 1\}$. Then the set of answers $A_{S_P}, \ldots, A_{S_1}$ precisely records the binary digits of x and so knowing these answers gives us x.

Let us model part (b) using communication theory. We are given a random vairiable X with $N-1$ possible values (the value of X is our x). Let us say that we are asking questions involving sets $S_1, \ldots, S_P$ and the truthful answers are $A_1, \ldots, A_p$. We can think of these answers as a (perfect) communication channel converting X to another variable Y with values binary strings with P digits (=truthful answers). Finally, someone corrupts Y by changing its digits, each with a probability of 30%. This is the same as transmitting the digits of Y through a binary symmetric channel with a 30% corruption probability. The output of this channel, Z, is the binary string which is presented to us as the "answers" to our question.

Now, it is clear that we can find the correct value of x by taking our questions from part (a) and repeating each many times and then using a majority vote decoder to decide the "true" answer. Indeed, we care about determining P binary values (here P is the smallest integer $\geq \log_2 N - 1$) with a total probability of error $< 1\%$. Thus it is enough to determine each answer with a probability of error $< (1/P)\%$, since the probability that at least one of P answers is given to us wrongly is at most P times the probability that an individual answer is given to us wrongly. Now, if we ask each answer to be repeated L times, the probability that hte majority of the answer are wrong will go to zero with L. Thus we determine the value of x with a probability of error at most 1% taking $M = LP$ large enough.

The lower bound for M can be deduced from the strong converse to Shannon's theorem for binary symmetric channels. Indeed, suppose that we can get the value of x after asking M questions with probability of error at most 1%. Assuming it takes 1 second to get an answer to a question, we are able to transmit the value of x ($= P$ bits) in M seconds through our channel successfully with this low probability of error. But the channel capacity is that of a binary symmetric channels operating in parallel, i.e., $C = 1 - H(30\%, 60\%)$. This is not possible to do reliably unless $C > P/M$, i.e., $M > P/C$. Thus $M > \log_2(N-1)/[1 + 0.3 \log 0.3 + 0.7 \log 0.7] > 8.4 \log_2(N-1)$. So if $N = 16$ we need at least 34 questions.

Problem 2. Give a statement of the fundamental theorem of information theory.

see p. 66 of the book.

Problem 3. Let f be a function from a set S with N elements $\{s_1, \ldots, s_N\}$ to a set T with M elements $\{t_1, \ldots, t_M\}$. If X is a random variable valued in S, then $f(X)$ has the obvious meaning: it is a random variable valued in T so that $\mathbb{P}(f(X) = t) = \mathbb{P}(X \in f^{-1}(\{t\}))$ (here $f^{-1}(\{t\}) = \{s : f(s) = t\}$).

Assume that for any t, all values of X so that $f(X) = t$ are equally likely. In other words, we assume that for any $s, s' \in S$, if $f(s) = f(s')$ then $\mathbb{P}(X = s) = \mathbb{P}(X = s')$.

Let $f'(s) = 1/[\#\{f^{-1}(\{f(s)\})\}]$. Thus $f'(s)$ is the reciprocal of the number of pre-images of $f(s)$; $f'(s) = 1/k$ means that f has the same value at s as some other $k-1$ points in S, i.e., f is "k to 1". Note that f' is a kind of "derivative" of f in that it counts the factor by which the space X is "shrunk" at s by the transformation f; this is similar to how a Jacobian can be interpreted.
(a) Is there a Chain rule for this "derivative"? In other words, is it true that $(f \circ g)' = (f' \circ g)g'$?
(b) Show that the uncertainty function H satisfies $H(f(X)) = H(X) - \sum_s \mathbb{P}(X = s)\ \log f'(s)$.

If r has a pre-images under f and *each of them* has s pre-images under g, then r will have rs pre-images under $f \circ g$. But this may not be true in general. For example, assume that $g(0) = 0$, $g(1) = g(-1) = 1$, and $f(0) = f(1) = 0$. Then $f \circ g = 0$ and so $(f \circ g)' = 3$ at every point $\{-1, 0, 1\}$. On the other hand, f is two-to-one and so $f' = 2$ everywhere. But $g' = 1$ at 0 and $g' = 2$ at ± 1. Thus $f'(g(s))g'(s)$ is 2 at $s = 0$ and 4 at $s = \pm 1$.

Note that if $Y = f(X)$, then $H(f(X)) = H(Y) = H(X, Y) - H(X|Y)$. Now, $H(X, Y) = -\sum_{\{x,y\}} p(x, y) \log p(x, y)$ if we denote by $p(x, y)$ the probability that $X = x$ and $Y = y$. But $p(x, y) = 0$ unless $y = f(x)$ and $p(x, y) = p(x)$ if $y = f(x)$. Thus the sum becomes $-\sum_x p(x) \log p(x) = H(X)$, so $H(X, Y) = H(X)$. On the other hand, $H(X|Y) = -\sum_y p(y) \sum_x p(x|y) \log p(x|y)$. But $p(x|y) = 0$ unless $y = f(x)$, i.e., $x \in f^{-1}(\{y\})$, in which case it is exactly $1/\#f^{-1}(\{y\})$. Thus

$$\begin{aligned} H(X|Y) &= -\sum_y p(y) \sum_{x \in f^{-1}(\{y\})} \frac{1}{\#f^{-1}(\{y\})} \log\left(\#f^{-1}(\{y\})\right) \\ &= -\sum_y p(y) \log\left(\#f^{-1}(\{y\})\right). \end{aligned}$$

Now, for any g

$$\sum_y g(y) = \sum_x \frac{1}{\#f^{-1}(\{y\})} g(f(y)),$$

since a single value of y arises from exactly $f^{-1}(\{y\})$ values of x. Moreover, $p(y) = \#f^{-1}(\{y\})p(x)$ if $y = p(x)$. From this, we conclude that

$$H(X|Y) = \sum_x p(x) \log \frac{1}{\#f^{-1}(f(x))} = \sum_x p(x) \log f'(x).$$

Thus $H(f(X) = H(X) - \sum_x p(x) \log f'(x)$.

Problem 4. Let X be a continuous random variable valued in $\mathbb{R}$. For our purposes, let us assume that this means that we are given a function $\phi : \mathbb{R} \to \mathbb{R}$ so that $\phi(t) \geq 0$ at all t, ϕ is continuous and the integral $\int_{-\infty}^{\infty} \phi(t)dt$ converges and equal 1. In this case we define X to be the random variable whose probability having its value in an interval $[a, b]$ is $\int_a^b \phi(t)dt$.

Define the uncertainty of X by $H(X) = -\int_{-\infty}^{\infty} \phi(t) \log \phi(t)\ dt$, where $x \log x$ is defined to be zero when $x = 0$. We'll assume that the integral defining $H(X)$ converges (this is a condition on ϕ).

Let f be a smooth 1-1 function, and let $Y = f(X)$. (a) Find an expression for a function ψ so that the probability that Y lies in an interval $[a, b]$ is given by $\int_a^b \psi(t)dt$ (hint: this is the same as asking that $f(X)$ lies in the interval $[a, b]$; express this quantity as an integral of ϕ and change variables).
(b) Show that $H(Y) = H(X) - + \int \phi(t) \log f'(t)dt$. Compare this with problem 3b.

The probability that Y lies in $[a,b]$ is the probability that X lies in $[f^{-1}(a), f^{-1}(b)]$, i.e., $\int_{f^{-1}(a)}^{f^{-1}(b)} \phi(t)dt$. Let $s = f(t)$. Then changing variables gives us that the probability that Y lies in $[a,b]$ is

$$\int_a^b \phi(f^{-1}(s)) \frac{1}{f'(f^{-1}(s))} ds.$$

Thus $\theta = \phi(f^{-1}(s))/f'(f^{-1}(s))$.

We now compute the entropy of Y:

$$\begin{aligned} H(Y) &= -\int \phi(f^{-1}(s)) \frac{1}{f'(f^{-1}(s))} \log \left[\phi(f^{-1}(s)) \frac{1}{f'(f^{-1}(s))} \right] ds \\ &= -\int \phi(t) \log \left[\phi(t) \frac{1}{f'(t)} \right] dt, \qquad \text{where we substituted } s = f(t) \\ &= -\int \phi(t) \log \phi(t) dt + \int \phi(t) \log f'(t) dt = H(X) + \int \phi(t) \log f'(t) dt \end{aligned}$$

Problem 5. Let C_1 be a binary symmetric channel, and C_2 be another binary symmetric channel. Let α_1 be the probability that C_1 transmits a digit incorrectly, and let α_2 be the probability that C_2 transmits a digit incorrectly. The output of C_1 is connected to the input of C_2 and the resulting channel is called C. Compute the channel capacity of C.

Clearly, C is a binary symmetric channel in which the probability of an incorrect transmission is $\alpha_1(1-\alpha_2) + \alpha_2(1-\alpha_1)$ (because C will transmit incorrectly iff exactly one of C_1 and C_2 transmits incorrectly). So the capacity is $1 - H(\alpha_1(1-\alpha_2) + \alpha_2(1-\alpha_1))$.