INFORMATION THEORY EXERCISE

(1) Consider the problem of minimizing $E[(Y-X)^2]$ when X has pdf

$$p(x) = \frac{1}{2\sqrt{2\pi}}\left(e^{-\frac{1}{2}x^2} + e^{-\frac{1}{2}(x-1)^2}\right)$$

and Y is constrained to have a discrete marginal distribution, concentrated on just two points (which are, however, subject to choice).

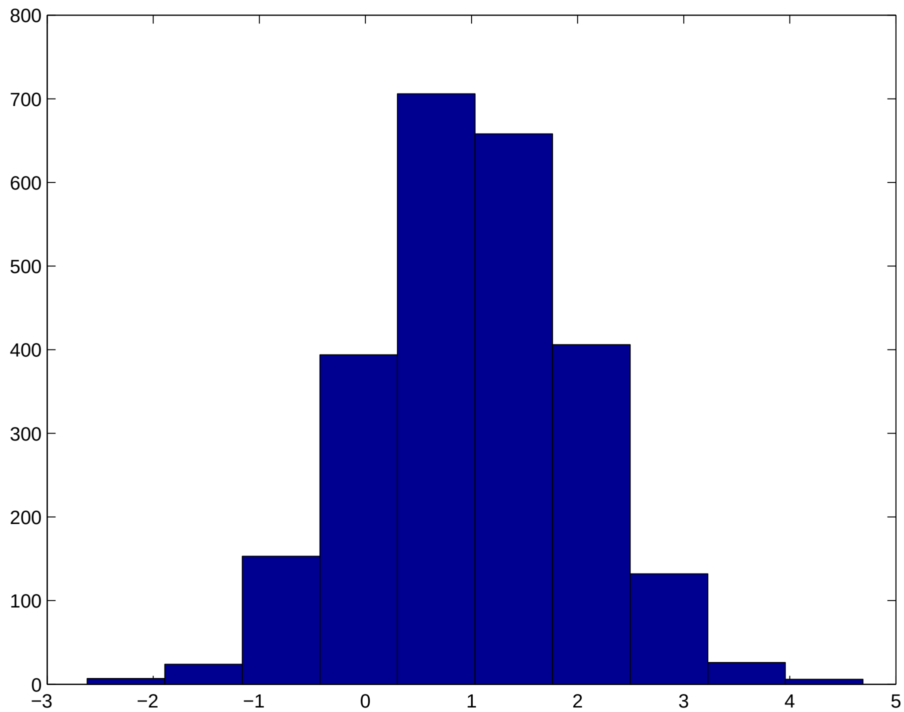
(a) Find the optimal joint distribution and the conditional distributions for $Y \mid X$ and $X \mid Y$.

(b) This problem can be thought of as equivalent to coding X into a form where it can be sent through an error-free one-bit per time unit channel (like a telegraph key). How close does your solution come to achieving the optimal 1-bit rate?

[Hint: This $p(x)$ is the convolution of a normal with a certain discrete distribution. This should allow you to exploit the optimality result from the notes.]

The hint in this problem is a bum steer. If Y is zero or 1 with equal probability, and if $X \mid Y \sim N(Y,1)$, we can calculate the rate of information transmission. Since the conditional distribution is normal with the same variance for either value of Y, the expected entropy after observation is $\log_2 \sigma + .5 + .5\log_2(2\pi) = 1.8257$. The entropy of the given $p(x)$ is, by my numerical integration, 2.2076. The bit rate is just the difference between these two numbers, or about .28 bits per time unit. The optimality result cited in the hint shows that if our capacity is constrained to .28 bits per time unit, then the joint distribution of Y and X that makes $X \mid Y\ N(Y,1)$ is optimal. But the problem said you were given a channel with a higher capacity.

If we can actually send a 1 or a 0 without error each period, we have a channel with capacity 1 bit per time unit, so we should be able to do much better than this. And in fact if we map $X > .5$ into $Y = E[X \mid X > .5]$ and $X \leq .5$ into $Y = E[X \mid X \leq .5]$, we get notably better mean square error. It is not too hard to calculate directly that $E[X \mid X > .5] = 1.325$, so that if we choose Y this way we get a RMSE of .89 in place of the 1.0 that we know arises from the "ideal" .28 bit per period channel. In fact, suppose we send only one bit every three periods and set a corresponding value of Y that stays constant for three periods. With $Y = E[X_i \mid X_1 + X_2 + X_3 > 1.5]$ when the mean of the three X's exceeds .5 and $Y = E[X_i \mid X_1 + X_2 + X_3 < 1.5]$ otherwise, we get an RMSE of .96 and a bit rate (obviously) of 1/3 bits per time unit. This is pretty close to the ideal .28 bit per time period setup. The figure below, a histogram for 5000 draws from $X_1 \mid X_1 + X_2 + X_3 > 1.5$, shows that the conditional distribution of $X_i \mid Y$ in this setup is not far from a $N(Y,1)$. If we go instead to using the mean of four successive X's to determine Y, the RMSE moves slightly above 1.



(2) Consider the problem of minimizing $E[(Y-X)^2]$ when X is a random variable distributed on the integers from 0 to 5 with binomial probabilities, i.e.

$$\frac{1}{32}, \frac{5}{32}, \frac{5}{16}, \frac{5}{16}, \frac{5}{32}, \frac{1}{32}.$$

We choose Y as a function of X subject only to the constraint that it can take on only 3 values, which we will call y_1, y_2, y_3. What these values are is subject to choice.

(a) What are the optimal y_i's and what is the optimal mapping from X to Y? As everyone concluded, one way or another, it is optimal simply to divide the 6 integers into three connected sets of equal size: $\{0,1\}$, $\{2,3\}$, and $\{4,5\}$. The optimal y values are of course the conditional means: $5/6$, $2\frac{1}{2}$, and $4\frac{1}{6}$.

(b) What is the channel capacity of a device (like the one considered in this problem) that can send, without error, one of 3 values each period? The maximal entropy distribution over any finite set is the uniform distribution over the points of the set. A channel that can send one of three values each period can completely resolve uncertainty in any distribution over three points, so its information transmission rate is maximized when the input distribution puts probability $\frac{1}{3}$ on each point, and it transmits $E[-\log_2 p] = \log_2 3 = 1.58$ bits per period.

(c) What is the rate of information transmission in the solution you have proposed as the optimal mapping in this problem? The entropy of the X distribution is 2.1982. The expected entropy of $X \mid Y$ is

$$2\frac{3}{16}\left(\frac{1}{6}\log_2 6 + \frac{5}{6}\log_2 \frac{6}{5}\right) + \frac{5}{8} \cdot 1 = .7044\,.$$

The information transmission per period is therefore 1.49 bits — only modestly short of the 1.58 bit capacity.

(d) If the problem were to choose the joint distribution of X and Y optimally without any constraint except that the rate of information flow had to be no greater than that of this 3-symbol channel, what would the optimal choice of the joint distribution be? (Of course here you are still taking the same distribution of X as given.) [This part of the problem may be really hard. You can give up after trying for a reasonable time and writing down any insights you have on the subject.] The optimality result cited in the first problem hint applies here, but does not really give much guidance, since the support of X is very much bounded. We expect the density of $X \mid Y$ to start looking Gaussian when we have things coded optimally. But there's no way to apply this insight that I can see. Since no one who understood the question attempted this part, I'm not presenting a numerical analysis here. Possible approaches include transmitting linear combinations of successive value of X or using block codes to achieve near-optimal transmission rates with low error.

[Hint: You will probably want to use the fact that it is optimal to choose $y_i = E[X \mid Y = y_i]$. You will probably need to use numerical methods. A fairly efficient numerical optimization routine, `csminwel.m`, will be on the course web site. If you can't get convergence of an optimization, shrewd guesses about the optimal joint distribution of X and Y are ok as answers. Answers will be graded according to how close they come to the best answer submitted. Finally, you might start with the simpler problem where X has a three-point distribution on $\{0, 1, 2\}$ with probabilities $\{.25, .5, .25\}$ and y has to have a two-point distribution. That might even be solvable analytically.