Signals, noise, and information

Fall 2008

Problem Set 5

Note: I would like to give you a problem that uses the ideas of estimation theory in the analysis real data, but it is taking a little longer than expected to get a good data set in the right form. In the meantime, try these ...

Problem 1: `MATLAB` provides a command `rand` which generates random numbers uniformly in the interval from 0 to 1. There is a trick which allows you to transform these numbers, generating samples from a different probability distribution. As an example, consider this short program to generate 10^5 samples of the random variable x:

```
z = rand(100000,1);
x = -log(1-x);
```

Estimate the probability distribution of `x` from the samples that you generate in this way, and show that it has an exponential form, $P(x) = e^{-x}$. Can you show analytically this should be true? Notice that `x` is guaranteed to be positive.

Problem 2: Now that you know how to generate samples from an exponential distribution, imagine that the data `x` are measured with an instrument whose readout `y` is just `x` with added, Gaussian noise. Again `MATLAB` has an easy way to make Gaussian random variables, so you can write

```
y = x + (0.2)*randn(100000,1);
```

and this will generate data in which the noise has standard deviation 0.2. Place the data `y` into bins (say, with a width of 0.1, but you should play with this a bit). Then, in each bin, find the average value of the underlying variable `x`. Use this construction to verify explicitly what we discussed in class, that the best estimator is not a linear function of the data even though the input/output relation of the measuring device is linear. How much better is this optimal estimator than simply guess that $x = y$? Can you go further and measure the distribution of `x` in each bin, and see that the most likely value has the thresholding behavior that we found analytically?