EE 599: Random Signals & Systems
Project 2, Fall 2008:

Project #2 is due on Thursday, December 4, 2008, in class. You have 3 weeks to work on this project. You may turn the project report in early. Late projects are accepted for one week past the due date (until 2 PM Thursday, December 11, 2008) at a penalty of 20% off.

I encourage you to complete the programming portion of this project as soon as possible.

1 Introduction

Project #2 is to be done individually, although student interaction is encouraged. The project report must be done individually, and may be handed in as hard copy or emailed in electronic form to the instructor.

You may use either Matlab or C/C++ for the programming portions of this project.

1.1 Project Reports

Show all your work! Simply giving the numeric answer to a question is not sufficient to receive full credit.

If you use Matlab to solve any of the project questions, please include your Matlab code as well as any requested data plots in your report. Similarly, if you use C/C++, please include your C/C++ code. Also include all input used and output obtained from your code in answer to any project questions.

If you are unable to solve a project problem, or your code does not work, please include whatever work you have done in your project report so that I can give you partial credit for the work you did do.

2 Channel Characterization of Binary Symmetric Channel and Confidence Intervals

Assume that you are sending binary data across a binary symmetric channel (BSC) with crossover probability (probability of a bit flip, or \( n_i = 1 \)) \( \rho \). However, you don’t know what \( \rho \) is.

Since \( n_i \) is a Bernoulli random variable with probability of success \( p(n_i = 1) = \rho \), you decide that you can determine what \( \rho \) is by sending \( L \) bits \( x_i = 0 \) across the BSC, and counting how many of the received bits \( r_i = 1 \) (or by summing all the received bits) and dividing by \( L \). You are calculating the sample mean \( M_L(r) \) in this operation; since all \( x_i = 0 \), this is equivalent to calculating the sample mean \( M_L(n) \).

Note: The Bernoulli random variable \( n \) has PMF \( p(n) \), where \( p(n = 0) = 1 - \rho \) and \( p(n = 1) = \rho \). Each sample \( n_i \) is i.i.d., with the PMF of \( n \). Thus we can either view each \( n_i \) as a separate Bernoulli random variable, or each \( n_i \) as one sample from the distribution \( p(n) \).
Exercise:
1. What is the expected value of the BSC noise $n$, $E[n]$?
2. What is the variance of $n$, $\text{Var}(n)$?
3. Show that the sample mean $M_k(n)$ converges to the expected value $E[n]$.
4. Generate 100 samples of BSC noise $n_i$ with $p(n_i = 1) = \rho = 0.3$, $\forall i = 1, \ldots, 100$. Use the function you wrote in Project 1 to generate a sequence of Bernoulli random variables. Include the command(s) you used here, but NOT your values for $n$, in your report.
5. Write a function to calculate the sample mean $M_k(n)$ for a sequence $n$ of length $k$. Include this function in your report.
6. Also write a function that calculates the 99% confidence intervals, and the 95% confidence intervals.
   You may include this function inside your sample mean function. Be sure to include this function in your report also.
7. Calculate the sample mean $M_k(n)$ using your samples $n$ for $k = 1$ to 100. The vector used to calculate the sample mean should include the first sample of $n$ to sample $k$. Thus only the final sample mean $M_{100}(n)$ will include sample 100. Include the command(s) used here in your report, but NOT the values of $M_k(n)$.
8. Plot the sample mean $M_k(n)$ versus $k$, as well as the 99% and 95% confidence intervals. Be sure to use a legend and label each curve.
9. At what $k$ are your 95% confidence intervals within $0.3 \pm 0.03$? At what $k$ are your 99% confidence intervals within $0.3 \pm 0.03$?
10. At what $k$ are your 95% confidence intervals within $0.3 \pm 0.01$? At what $k$ are your 99% confidence intervals within $0.3 \pm 0.01$?
11. Using the central limit theorem to represent $n$ as a Gaussian random variable, how many samples are required to guarantee that a confidence interval estimate of length $2c = 0.06$ has confidence coefficient $1 - \alpha \geq 0.95$? How many for the same confidence interval estimate but with confidence coefficient $1 - \alpha \geq 0.99$? How do your results match with the required number of samples?
12. Using the central limit theorem to represent $n$ as a Gaussian random variable, how many samples are required to guarantee that a confidence interval estimate of length $2c = 0.02$ has confidence coefficient $1 - \alpha \geq 0.95$? How many for the same confidence interval estimate but with confidence coefficient $1 - \alpha \geq 0.99$? How do your results match with the required number of samples?
13. Could you accurately estimate the crossover probability $\rho$ using the sample mean $M_k(n)$ in 100 samples? What was your error $e = M_{100}(n) - E[n]$ at the 100th value for the sample mean $M_k(n)$?
3 Central Limit Theorem and Gaussian Noise

The central limit theorem states that the CDF of the sum of a sequence of i.i.d. random variables approximates the CDF of a Gaussian random variable more closely with increasing sequence length, reaching equality as $n \to \infty$. The central limit theorem justifies the use of a Gaussian approximation to represent the CDF of $W = \sum_{i=1}^{L} X_i$, where $X_i$ are all i.i.d. (discrete or continuous) random variables with mean $\mu_x$ and variance $\sigma_x^2$, as

$$F_W(w) \approx \Phi \left( \frac{w - n\mu_x}{\sqrt{n\sigma_x^2}} \right)$$

(1)

We are going to use this approximation for the case of summing several i.i.d. Gaussian sources of noise together and determine the relative accuracy of this method.

Exercise:

1. Using your Gaussian noise function from Project 1, generate a Gaussian-distributed sequence $n$ of length $L = 1000$, with zero mean and variance $\sigma_n^2 = 0.001$.

2. Sum your sequence $n$ together to generate $W$, such that

$$W = \sum_{i=1}^{1000} n_i.$$ 

3. Experimentally determine $F_W(-1) = P(W \leq -1)$, using your sequence $W$ found above.

4. Using the central limit theorem approximation, calculate $F_W(-1)$, with length $L = 1000$, $\mu_n = 0$ and $\sigma_n^2 = 0.001$.

5. How does your experimental value for $F_W(-1)$ compare to the central limit theorem approximation?

4 Central Limit Theorem and BSC Noise

Now we are going to look at how well the central limit theorem approximates a discrete random variable CDF, by examining the number of errors generated by a BSC. As in Section 2, at each transmission time $i$, the BSC generates binary noise $n_i$ with probability $p(n_i = 1) = \rho$; $n_i$ is a Bernoulli random variable. The number of errors $E$ found by summing all $n_i$ as $E = \sum_{i=1}^{L} n_i$ is a binomial random variable.

1. As in Section 2, generate $L = 1000$ samples of BSC noise $n_i$ with $p(n_i = 1) = \rho = 0.3$, $\forall i = 1, \ldots, 1000$.

2. Sum your sequence $n$ together to generate $E$, such that $E = \sum_{i=1}^{1000} n_i$.

3. Experimentally determine $P(300 \leq E \leq 350)$, using your sequence $E$ found above.
4. Using the central limit theorem approximation, calculate $P(300 \leq E \leq 350)$.

5. How does your experimental value for $P(300 \leq E \leq 350)$ compare to the central limit theorem approximation?

6. Now use the De Moivre-Laplace formula to calculate $P(300 \leq E \leq 350)$. Is this approximation any better than the central limit theorem approximation? Why or why not?