**CS 559: Homework Set 1**
**Due: Oct. 8, 11:59pm**

**Collaboration Policy.** Homeworks will be done individually: each student must hand in their own answers. Use of partial or entire solutions obtained from others or online is strictly prohibited.

**Late Policy.** No late submissions will be allowed without consent from the instructor. If urgent or unusual circumstances prohibit you from submitting a homework assignment in time, please e-mail me explaining the situation.

**Submission Format.** Electronic submission on Canvas is mandatory.

**Problem 1. (30 points)** You have $\$B$ in your bank account. You are asked if you would like to participate in a bet in which, if you win, your bank account will become $\$W$. However, if you lose, your bank account will contain only $\$L$. You win the bet with probability $p_w$. How large should $p_w$ be to accept the bet? (I assume that you would rather have a higher expected value of the amount of money in your account.)

**Problem 2. (30 points)** Assume you have a green and a black wallet. The green wallet contains 6 pennies and 4 dimes. The black wallet contains 8 pennies and 2 dimes. Which wallet were you more likely to have picked if you pulled a dime followed by two pennies from it? Based on past experience, you use the green wallet 4 times more often than the black one.

What is the probability that the optimal answer you gave in the previous question was wrong?

**Note:** Answering the second part without attempting the first part will receive 0 credit.

**Problem 3. (40 points)**

**Part 1.** In Matlab, or the programming language of your choice, do the following:

- Generate $N$ observations from a normal distribution: $data = randn(N, 1)$; This will generate $N$ 1-D samples with mean equal to 0 and variance equal to 1.
- Estimate the mean and variance of the data for $N = 10,\ 100,\ 1000$ etc.
- Modify the code so that the generated data have mean and variance equal to user-specified parameters $mean$ and $var$.

**Submit** a function that accepts $mean$, $var$ and $N$ and generates data as in the last step above. No explanations required.

**Part 2.** Using the function above, generate two datasets: one with $N_1 = 2000$, $\mu_1 = 1$, $\sigma_1^2 = 4$ and one with $N_2 = 1000$, $\mu_2 = 4$, $\sigma_2^2 = 9$. Derive theoretically the mean and variance of the combined dataset that includes all 3000 samples. Verify experimentally that your theoretical estimates are correct.

**Submit** the derivations for the mean and variance of the combined dataset. Make sure it is written with $mean_i$, $var_i$ and $N_i$ as parameters. No code required.