1. This problem has some parts with pencil and paper, and some parts wtih Matlab. *Submit paper answers in class and post your Matlab solutions online (published results (in pdf or doc format)* ***and*** **.m files). Make sure your *.m files run and output results accordingly.*

   (a) Suppose that you are given a random variable U which is uniformly distributed on the interval between 0 and 1. Constract a function g such that $Y = g(X)$ is uniformly distributed on the interval between a and b. The function g should depend upon a and b.

   (b) Generate a matlab function M-file that is given two inputs, a and b (where $a \leq b$) and its output is a number U that is uniformly distributed on the interval between a and b. Inside your M-file, you should call **rand**.

   (c) Suppose that the heads probability of a coin flip is a random variable X, equally likely to be between 0.2 and 0.75. Once the probability of heads is established (i.e. $X = u$), coins are flipped independently for 10 times, where the probability of heads on each toss is u. Let Y represent the total number of heads on the ten tosses. Generate another M file for this problem that first draws X using code from part (b). In that M-file, add new code that simulates the 10 coin flips (*use your previous M-file code from HW1 to simulate coin flips of a known heads probability*) and calculates Y. Make the output of your M-file the value of X as well as the value of Y.

   (d) what is the probability that $Y = 5$?

   (e) what is the probability density on X given that $Y = 5$? After you have calculated your expression on pencil and paper, write an M-file that makes a Matlab plot to show this. Use the function **linspace(a,b)** to create equally spaced values of u between a and b. For example, the code **a = 0.1; b=0.8; uvec = linspace(0.1, 0.8);** will create a vector **uvec** of 100 with equally spaced points between 0.1 and 0.8. You can verify this by typing **min(uvec), max(uvec), length(uvec)**.

2. Suppose that Medtronic has a heart arrhythmia monitoring system that fails with probability 10^{-6}. Suppose that the US has approximately 3 million people with Medtronic monitors implanted. Use the Poisson approximation to the Binomial to approximate the probability that the number of people in the US whose Medtronic arrhythmia monitoring system fails is greater than 2.

3. (a) In the previous homework set, you showed using the law of conditional probability that the geometric random variable has the *memoryless* property:

$$\mathbb{P}(X > k + j | X > j) = \mathbb{P}(X > k).$$

Now I want you to derive rigorously without any math. Pick $k = 2$ and $j = 3$. Draw out the scenario. Remember from course how to re-describe the event $X > j$ for a geometric random variable. Argue this simply from the fact that we are flipping independent coins and using equivalence of events.

(b) Show using the law of conditional probabilities and integration that if Y is an exponential random variable of rate λ then we still get

$$\mathbb{P}(Y > k + j | Y > j) = \mathbb{P}(Y > k).$$

(c) Show the same result in part (c), using logic analogous to part (a). *hint: how does an exponential random variable relate to a geometric?*

4. (a) Show (perhaps with an example) that for any three numbers a, b, c, the following logical statements $C = \{\min(a, b) > c\}$, $A = \{a > c\}$ and $B = \{b > c\}$ satisfy the following relationship:

$$C = A \cap B.$$

In other words, pick any (a, b, c). If the logical statement C is true then so are the logical statements A and B. And vice versa. Show this. Start by picking some a, b, c values and watch the pattern. Then give the relationship generally.

(b) suppose X is an exponentially distributed random variable with rate λ_X and Y is an exponential random variable with rate λ_Y. Calculate $\mathbb{P}(X > c)$ and $\mathbb{P}(Y > c)$.

(c) let X and Y from part (b) be statistically independent and Z be the random variable given by $Z = \min(X, Y)$. Show that Z is exponentially distributed with rate $\lambda_X + \lambda_Y$. (*hint: use part (a) and exploit independence clelverly*)

5. In this problem, we will use a common approach to analyzing neural spiking data: discretize time into bins of fixed width and count the number of events that occur in each time bin.

(a) Bin the spike train data from **Iyengar_spike_times.mat** (originally sampled at 1ms time resolution) into time bins of width 1 ms, 10 ms, and 100 ms. Plot both the time series of spike counts and the distribution of spike counts as a histogram for each bin width.

The Poisson process is a commonly used simplistic model for neural spiking activity. One feature of this model is that if the spiking activity binned into 1 second blocks is a Poisson random variable with parameters $(\lambda, 1)$, then the number of spike counts in a bin of size Δ seconds is also Poisson with parameters (λ, Δ).

(b) Given the number of spikes observed in this 30 second observation interval, what is a good choice for a rate parameter (in spikes per second) for a Poisson model of this data? Plot your histogram distributions from part (a) against the model distributions of a Poisson with the appropriate rate parameter. How well do these Poisson models fit the data for each bin width? Why might the Poisson model not be able to capture the statistical structure in this data?

(c) When the bin width becomes small, the probability of observing more than one event under a Poisson model becomes very small and can be well approximated by a Bernoulli random variable. In class, we discussed the Poisson approximation to the Binomial probability model. Here we will build a Bernoulli approximation to a Poisson model. Assume you have a spiking neuron that fires according to a Poisson model with the rate parameter you calculated in part B. Using our basic probability laws, note that for a Poisson variable N, $Pr[N > 1] = 1 - Pr[N = 0] - Pr[N = 1]$. What is the largest bin width for which the probability of observing more than one spike is smaller than 0.05? What is the probability of observing more than one spike for a bin width of 1 ms? What would be an appropriate Bernoulli approximation to this model for a bin width of 1 ms?

6. Suppose that the position of a rat, denoted as the random variable X, lies somewhere uniformly at random on the $[0, 1]$ line. We record from a hippocampus 'place cell' neuron whose spiking depends on the position of the rat. Denote Y as the random variable specifying the total number of spikes recorded in 50ms of recording. The distribution of Y given $X = x$ is a Poisson distribution whose rate λ depends on x:

$$\lambda(x) = g(x)$$

for some function g.

(a) Find the PDF of X, $f_X(x)$ and the PMF of Y given $X = x$, $P_{Y|X}(y|x)$.

(b) Find the posterior PDF on X given that k spikes were recorded in $50ms$: $f_{X|Y}(x|y)$.

(c) Assume that $g(x)$ has a 'place cell' structure:

$$g(x) = \lambda_0 + \lambda_1 e^{-(x-x_0)^2}.$$

Fix $\lambda_0 = 3$, and $\lambda_1 = 2$. Plot $g(x)$ on $[0, 1]$ for $x_0 = 0.25, 0.5, 0.75$. Describe what λ_0, λ_1, and x_0 represent from the 'place cell' spiking context.

(d) Suppose that 20 spikes occurred in the $50ms$ of recording. Plot the posterior PDF $f_{X|Y}(x|20)$ on $[0, 1]$ using matlab. (*hint: remember that it must integrate to one*).