1. In class, we discussed the Central Limit Theorem (CLT), which states that the sum of independent, identically distributed (iid) random variables with finite variances will asymptotically be Gaussian. This is one reason that approximate Gaussian methods are so common in statistical analyses. In this problem, we will explore some features of the CLT for both discrete and continuous random variables.

   (a) Using MATLAB, plot the distribution of a Poisson random variable with rate parameter values of 1, 5, 15 and 25. Compare this distribution using QQ plots against a normal distribution with the appropriate mean and variance. What statements can be made about the quality of the Gaussian approximation to the Poisson as a function of the parameter?
   Useful MATLAB functions: **sqrt, exp, cumsum**

   (b) Repeat part A replacing the Poisson distribution with Binomial distributions with different values of n and p. Try a grid with $n = [15, 25, 50]$ and $p = [0.1, 0.3, 0.5]$. How does the quality of the Gaussian approximation vary as a function of n and p? How could you simulate a Gaussian random variable using a penny and how many flips would you need?
   Useful MATLAB functions: **nchoosek**

   (c) As mentioned in class, it is easy to show that the sum of i.i.d. Binomial and Poisson random variables are themselves Binomial and Poisson, respectively. But the CLT, states that they should be asymptotically Gaussian as the number of variables being summed becomes large. Based on your answer to parts A and B, why isn?t this a contradiction? In other words, how can the distribution of these sums asymptotically be both Binomial and Gaussian or both Poisson and Gaussian?

   (d) Now, let's examine the CLT in action for uniform random variables. One way to estimate the distribution of a random variable is to sample many times from a probability density and then plot the distribution of these samples. This procedure is called Monte Carlo simulation. Generate $10,000$ samples from a uniform $[0, 1]$ statistical law and construct a histogram showing the empirical CDF of these samples. Do the same for the $10,000$ sums of $2, 3, 4, 5,$ and 6 samples. Compare these empirical distributions to the appropriate Gaussian approximations via a QQ plot. How could you simulate a Gaussian random variable using only samples from a uniform?
   Useful MATLAB functions: **cumsum, rand, bar, hist**

2. (a) Suppose that $Y = aN + b$ where $a > 0$ and b is some constant. Show that the PDF of Y, $f_Y(v)$, satisfies

$$f_Y(v) = \frac{1}{a} f_N\left(\frac{1}{a}(v - b)\right).$$

(*hint: first calculate the CDF of* Y *and use equivalence of events*).

(b) Suppose that N is a continuous random variable with PDF $f_N(u)$. Now suppose that $Y = X + N$ where X is another random variable that is statistically independent of N. Carefully argue that

$$f_{Y|X}(v|u) = f_N(v-u).$$

(*hint: use equivalence of events after you have conditioned upon the fact that* $X = u$, *and exploit statistical independence*).

3. Suppose that N is Gaussian with expectation μ_N and variance σ_N^2.

(a) Show that the random variable $Y = aN + b$ is Gaussian with expectation $a\mu_N + b$ and variance $a^2\sigma_N^2$. (*hint: it is not enough to simply argue that* Y *has a specific mean and a specific variance. I want you to show that her whole PDF is that of a Gaussian. Use problem 2a and the Gaussian PDF*).

(b) Let $Y = X + N$ where X is a random variable that is statistically independent of N. Show that

$$f_{Y|X}(v|u) = \frac{1}{\sqrt{2\pi\sigma_N^2}} \exp\left(-\frac{(v-u-\mu_N)^2}{2\sigma_N^2}\right).$$

(c) Let $\mu_N = 0$ and $\sigma_N^2 = 3$. Use Matlab and plot the conditional PDF of Y given $X = -10$, $f_{Y|X}(v|-10)$ and the conditional PDF of Y given $X = 10$, $f_{Y|X}(v|10)$, for v ranging from -30 to 30..

(d) Suppose X is equally likely to be 10 or -10. Give an expression for $f_Y(v)$ as a function of v, and plot $f_Y(v)$ for v ranging from -30 to 30.

(e) Now suppose X is a random variable with PDF given by

$$f_X(u) = \begin{cases} c(2+u), & u \in [-2,-1) \\ c(2-u), & u \in [1,2] \\ 0 & \text{otherwise.} \end{cases}$$

Suppose that we observe that $y = 13.3$. Provide an expression for the conditional PDF of X given $y = 13.3$, $f_{X|Y}(u|13.3)$, up to a proportionality constant (i.e. the constant cannot vary with u). Use Matlab to plot $f_{X|Y}(u|13.3)$, up to a proportionality constant, for u ranging from -10 to 10. (*hint: first create separate M-file that provides the PDF of* X *and the PDF for* N. *Then combine them in your main script.*)

4. Suppose that $Y_1, Y_2, \ldots, Y_n$ are independent, identically distributed random variables where each Y_i is Poisson with an unknown parameter λ. All we know is that $3 \le \lambda \le 8$.

(a) Assume for a moment we are Bayesian that nature first selects the random variable X describing the parameter λ, and then afterwards generates n independent samples $Y_1, Y_2, \ldots, Y_n$ from a Poisson PMF of parameter given by the outcome of X. In Matlab, first draw X from a uniform statistical model on $[3, 8]$. Then take that value of X and use it as the parameter λ for generating $n = 10$ independent Poisson random variables $Y_1, Y_2, \ldots, Y_{10}$. Provide your code and give the values of $X, Y_1, \ldots, Y_{10}$. (*hint: use the Matlab command* **rand**).

(b) Provide an expression for the probability that $Y_1 = k_1$ and $Y_2 = k_2$ $\ldots$ and $Y_n = k_n$ given that $X = u$:

$$P_{Y_1,Y_2,\ldots,Y_n|X}(k_1, k_2, \ldots, k_n|u) = ?$$

(c) suppose someone told us that it is twice as likely for λ to lie between 3 and 7 as it is to lie between 7 and 8. Provide a PDF on X, $f_X(u)$, for $3 \leq u \leq 8$, and plot it with Matlab.

(d) Suppose now we observe $Y_1 = k_1$, $\ldots$, $Y_n = k_n$ and we would like to infer $f_{X|Y_1,\ldots,Y_n}(u|k_1 \ldots, k_n)$. Show that it can be expressed as

$$f_{X|Y_1,\ldots,Y_n}(u|k_1 \ldots, k_n) = \frac{g(u)u^{(\sum_{i=1}^{n} k_i)}}{C(k_1, \ldots, k_n)}.$$

and specify the functions $g(u)$ and $C(k_1, \ldots, k_n)$. The latter function C is sometimes called a 'proportionality constant' because it does not vary with u.

(e) Now suppose that the first 10 observations were

i	Y_i
1	5
2	4
3	6
4	3
5	2
6	3
7	4
8	3
9	5
10	2

and we would like to learn the parameter λ from the data. In Matlab, plot each of the below five functions functions, for $3 \leq u \leq 8$:

$$f_X(u)$$
$$f_{X|Y_1}(u|5)$$
$$f_{X|Y_1,Y_2}(u|5, 4)$$
$$f_{X|Y_1,\ldots,Y_5}(u|5, 4, 6, 3, 2)$$
$$f_{X|Y_1,\ldots,Y_{10}}(u|5, 4, 6, 3, 2, 3, 4, 3, 5, 2)$$

up to a proportionality constant (*i.e. you don't need to calculate* $C(k_1, \ldots, k_n)$*: it is only a constant to guarantee the area under the curve is* 1). As we get more data, notice how the posterior belief on X, given by $f_{X|Y_{1:n}=y_{1:n}}$, "sharpens".

5. Now we will take the exact same problem setup as problem 4, except we will use a frequentist perspective. Here, we assume that λ is unknown but **fixed and non-random**.

   (a) Define $\bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$. Specify the mean and variance of $\bar{Y}_n$ in terms of λ.

   (b) Use the Chebyshev inequality to argue to provide an expression C for which

   $$\mathbb{P}\left(\left|\bar{Y}_n - \lambda\right| > \epsilon\right) \leq C(\epsilon, \lambda, n)$$

   for some function C that depends upon ϵ, λ, and n.

   (c) For this specific problem, provide a function $D(\lambda, n)$ for which $C(\epsilon, \lambda, n) \leq D(\epsilon, n)$ (*hint: this is easy. read the beginning of problem 4*).

   (d) Define the interval $I_{\bar{Y}_n,\epsilon}$ as $I_{\bar{Y}_n,\epsilon} = (\bar{Y}_n - \epsilon, \bar{Y}_n + \epsilon)$. Use equivalence of events and complement of events to argue that

   $$\mathbb{P}\left(\lambda \notin I_{\bar{Y}_n,\epsilon}\right) = \mathbb{P}\left(\left|\bar{Y}_n - \lambda\right| > \epsilon\right) \leq D(\epsilon, n).$$

   (e) Now let ϵ be a function of n, i.e. $\epsilon \equiv \epsilon_n$ and give an expression of ϵ in terms of n to guarantee a 95% confidence interval, i.e. that

   $$\mathbb{P}\left(\lambda \in I_{\bar{Y}_n,\epsilon}\right) > 0.95$$

   (f) Plot the 95% confidence intervals for the same $Y_1, \ldots, Y_{10}$ values given in part (e) of problem 4, for the scenarios when $n = 1, 2, 5, 10$. What similarities do you see in terms of "sharpening" here, as it compares to 4e?