The homework pertains to material covered in the first three lectures. The assignment should be typed, with properly labeled computer output except for some algebraic calculations that could be hand-written. You are encouraged to work together to a reasonable degree, but the write up should be your own.

1. For each of the following scenarios, say whether the binomial distribution with $n=30$ is a reasonable probability model for X. If yes, explain why you think the assumptions of binomial are satisfied. If no, explain how you think the assumptions are violated.
   a. An unfair coin is flipped 30 times. Let $X$ be the number of heads out of 30 that appear.
   b. A baseball team in an amateur league plays 3-games series against each of the other teams in the league, for a season total of 30 games. Let $X$ be the number of games the team wins out of 30 played games.
   c. A survey firm draws a random sample of 30 households in State College. Two adults from each household are asked, "Are you generally optimistic about the next four years with George W. Bush as President?". Let $X$ be the number who said "Yes".
   d. A survey firm draws a random sample of 15 households in State College containing two or more adults. In each one, one adult member of the household is asked, "Are you generally optimistic about the next four years with George W. Bush as President?". Let $X$ be the number who said "Yes".

2. For each of the following scenarios, say whether a Poisson model for the sample $(X_1, X_2, ..., X_n)$ may be appropriate.
   a. In Massachusetts state lottery game Megabucks, any player may win with probability 1 out of 2 million independently of other players. Also, suppose that the number of players each week is always equal to a constant $N$. Let $X_1, X_2, ..., X_n$ be the number of Megabucks winners in the first $n$ weeks of 2003.
   b. Now suppose the scenario as above, but the number of players $N$ varies substantially from week to week.
   c. Let $X_1, X_2, ..., X_{12}$ be the number of cars arriving at a tool booth on major highway during 12 consecutive hours in a single day.
   d. Let $X_1, X_2, ..., X_{48}$ be the number of cars arriving at the toll booth between hours of 1 and 2 pm on 48 Fridays in 2005 (excluding holidays).
3. Let $X$ be the number of vegans (strict vegetarians) observed in a random sample of $n = 30$ graduate students at Penn State. It is reasonable to assume that $X \sim \text{Bin}(30, p)$. The true proportion $p$ is unknown, but it's likely to be small.

(a) Assume for now that $p = .04$. Find the mean, the variance, and the standard deviation of $X$. Find the probabilities of the events $X = 0$, $X = 1$, $X = 2$, $X = 3$, and $X \geq 4$.

(b) The classical approximate 95% confidence interval for $p$ given in most textbooks is

$$
\hat{p} \pm 1.96 \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}},
$$

where $\hat{p} = x/n$. If the true $p$ was actually .04, could this interval actually cover the true parameter 95% of the time? Why or why not? (Hint: Based on part (a), how often would the interval become degenerate at zero?)

(c) Suppose that we observe two vegans in the sample. Plot the loglikelihood function for $p$ over a range of values from $p = .01$ to $p = .20$. Find the ML estimate $\hat{p}$ and the value of the loglikelihood at $\hat{p}$. Calculate the approximate 95% confidence interval for $p$ based on the expected information.

(d) Plot the loglikelihood function with respect to the transformed parameter

$$
\phi = \log \frac{p}{1 - p}
$$

over a range of $\phi$-values from $\phi = \log .01/.99 = -4.60$ to $\phi = \log .20/.80 = -1.39$. Comment on the appearance of this plot relative to the one in part (b). Calculate the approximate 95% confidence interval for $\phi$ based on the expected information, and transform the endpoints back to the $p$-scale.

(e) Now find an approximate 95% interval for $p$ based on the LR method, and compare it to the two intervals you found previously. Which of these interval(s) seem most appropriate, and why?
4. Five children were given a vitamin C supplement for an entire year. The numbers of colds experienced by these children over the course of the year was 2, 0, 1, 1, and 0.

   (a) Assuming a Poisson model, find the ML estimate and an asymptotic normal 95% confidence interval for \( \lambda \) based on the observed or expected information. Is this a sensible interval? Why or why not?

   (b) Find asymptotic normal 95% confidence intervals using the transformations \( \phi = \log(\lambda) \) and \( \phi = \lambda^{1/3} \).

   (c) Also, for purposes of comparison, calculate the standard normal-theory 95% confidence interval given by

   \[
   \bar{x} \pm t \sqrt{\frac{S^2}{n}}
   \]

   where \( S^2 \) is the sample variance and \( t \) is the 97.5th percentile of the \( t \)-distribution with \( n - 1 \) degrees of freedom. This interval is based on an assumption that the data come from a normal population. Would you trust the normal-theory confidence interval for these data? Explain.

   (d) Plot the loglikelihood

   i. as a function of \( \lambda \) over the range \( \lambda = 0.20 \) to \( \lambda = 2.0 \),

   ii. as a function of \( \phi = \log \lambda \) over the range \( \phi = \log(0.2) = -1.6 \) to \( \phi = \log(2.0) = .69 \), and

   iii. as a function of \( \phi = \lambda^{1/3} \) over the range \( \phi = 0.2^{1/3} = .58 \) to \( \phi = 2.0^{1/3} = 1.26 \).

   Based on these plots, which confidence interval for \( \lambda \) do you prefer?

   (e) Now find an approximate 95% confidence interval for \( \lambda \) using the LR method.

5. Find and state the Bayes’ Theorem (any stats book should have this, or search the web). How does this relate to conditional probabilities. What is meant by “prior” and what by “posterior” probability of an event?