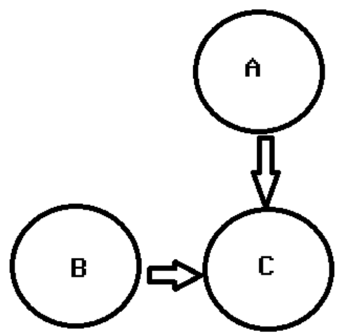
ANLY-511: Practice for Midterm Test

The real test will have about the same number of problems. But some problems may be shorter or longer and some topics may be covered on the test that don't appear on this practice test. Give yourself $\frac{x}{2}$ *minutes to do a problem that counts towards* x *%.*

1. (20 %) Consider the Bayesian network in the graph below. There is a random variable on each node that can have the values T and F. Arrows indicate conditioning.



Here are some probabilities:

$$\mathcal{P}(A=T)=.6,\ \mathcal{P}(B=F)=.5,\ \mathcal{P}(C=T)=.39$$

$$\mathcal{P}(C=T|A=T\&B=T)=.1,\ \mathcal{P}(C=T|A=T\&B=F)=.4,,\mathcal{P}(C=T|A=F\&B=T)=.8\,.$$

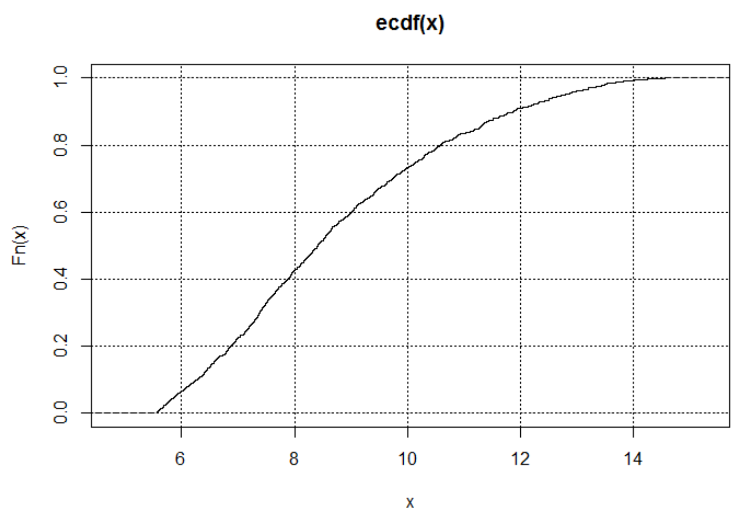
Find $\mathcal{P}(A=T\&B=T)$, $\mathcal{P}(C=T|A=F\&B=F)$, $\mathcal{P}(C=T|A=T)$.

2. (20 %) You are given a sample x from a certain distribution and sanother sample y, perhaps of different size, from another distribution. The two samples are already sorted from smallest to largest. You are interested in determining whether the two distributions are possibly the same. You already know that they have very similar means and standard deviations. Which of the following approaches can be used to do this? Which are useless or wrong? Which cannot be implemented in `R` ? Write a one line explanation in each case.

1. Plot y against x. If this is close to a straight line, the two distributions might be the same.
2. Make `qqnorm` plots of both. If these look similar, the distributions might be the same.
3. Plot the ecdf of x and the ecdf of y in the same graph. If they almost fall on top of each other, the two distributions might be the same.
4. Subtract the two sample from one another and make a boxplot of the differences. If these differences are all very small, the two distributions might be the same.
5. Nothing more is needed, since the mean and standard deviations are so close.

3. (20 %) Let n, p, q be given, where n is a positive integer and $0 < p < 1$, $0 < q < 1$. Let X have a $B(n,p)$ distribution and let $Y|X = k$ have a $B(k,q)$ distribution.
a) Find a formula for $\mathcal{P}(X = Y = 1)$ in terms of n, p, q.
b) Find $\mathcal{E}(Y|X)$.
c) Find $\mathcal{E}(Y)$.

4. (20 %) Suppose a sample $(x_1, \ldots, x_n)$ is given. The $2k$ % **Winsorization** of this sample consists of a sample of the same length, where all data below the $(50-k)$th %ile of the sample are set to the $(50-k)$th percentile and all data above the $(50+k)$th percentile are set to the $(50+k)$th percentile. All other data are unchanged.
Suppose a sample has the empirical cdf (ecdf) given below. Draw the ecdf of the 80% winsorized sample in the same plot.



5. (5 %) Pick a random number X from $\{1, \ldots, 20\}$, with uniform probabilities. Let A be the event "X is divisible by 3" and let B be the event "X is divisible by 5". Compute $\mathcal{P}(A|B)$ and $\mathcal{P}(B|A)$.

6. (5 %) Let X, Y be random variables with finite expected values. It is known that $\mathcal{E}(X - \mathcal{E}(X)) = 0$. Is $\mathcal{E}((X - \mathcal{E}(X))|Y) = 0$? Explain your answer.

7. (5 %) Suppose the random variable X has the cdf F. Express the cdf of the random variable $Y = 3X - 1$ in terms of F.

8. (5 %) Write down a transition matrix P for a Markov chain X_i with three states $\{A, B, C\}$ such that $\mathcal{P}(X_2 = A|X_0 = A) = \frac{1}{4}$.