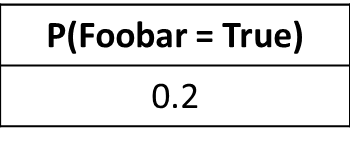
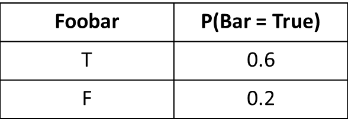
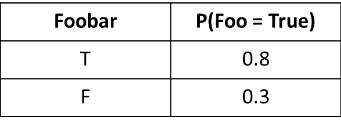
1. **Bayesian Networks, Maximum Likelihood Estimates, and Naïve Bayes**

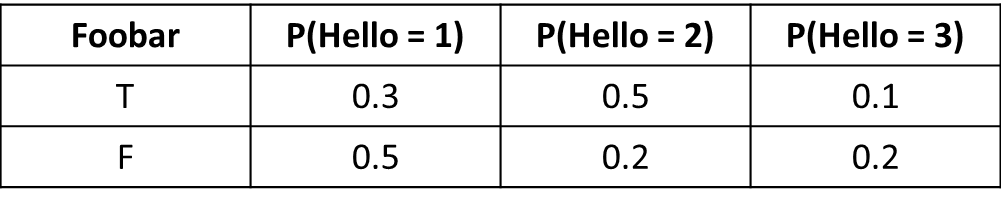
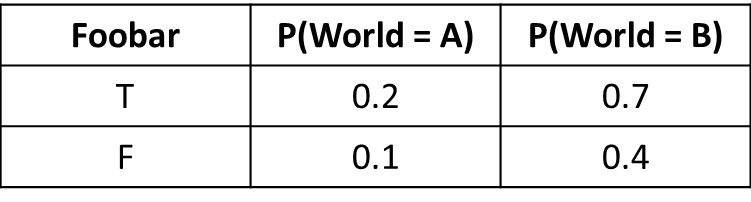
You are developing a Naïve Bayes system to predict whether or not documents are members of a class called *Foobar*, whatever that means. Predictions will be based on the values of four features, called *Foo*, *Hello*, *World*, and *Bar*, which are assumed to be conditionally independent given *Foobar*. *Foobar* is a Boolean class (i.e., it is either True or False); *Foo* and *bar* are Boolean features; and *Hello* and *World* are discrete features (the domain of *Hello* is {1, 2, 3, 4} and the domain of *World* is {A, B, C}).

A maximum likelihood approach has already been used to estimate the probabilities of values for each feature given the value of *Foobar*, as well as the prior probability of *Foobar*. The learned probability estimates are indicated in the conditional probability tables shown alongside the Bayesian network below. As is typical, to minimize the number of indicated parameters, the conditional probability tables leave out one possible value of each random variable, since they can be determined from the given information.



C:\Users\Carl\Documents\AI\F17\Quizzes\bn_f17_q2.tiff

****



Assume that the system is used to predict the category (*foobar* or ) of a new example, **e**, with the following feature values: **e** = [*Foo* = True, *Hello* = 2, *World* = C, *Bar* = False].

1. What is the computed probability of seeing these feature values, assuming that *Foobar* is True? In other words, compute P(**e** | *foobar*). Show your work.
2. What is the computed probability of seeing these feature values, assuming that *Foobar* is False? In other words, compute P(**e** | )? Show your work.
3. What are the computed probabilities of *Foobar* being True and False, given the feature values of **e**? In other words, compute P(*foobar* | **e**) and P( | **e**)? Show your work and simplify your final answers.
4. **Machine Learning Concepts**

Briefly answer each of the following questions related to machine learning (answers should contain one or at most two sentences each).

1. Related to machine learning, explain the concept of feature selection.
2. Related to decision trees, would it ever make sense for the same feature / attribute to be tested twice along a single path from a root to a leaf?
3. Briefly explain the conclusion of the No Free Lunch theorem.
4. Briefly explain why it typically takes longer to apply a k-nearest neighbor system than to train it. Is this a good thing or a bad thing?
5. Consider figures (i) through (iv) below. In each of them, the x's represent positive examples of some class, and the o's represent negative examples of the same class. Each example is fully represented as a two-dimensional numeric feature vector. Which of these classification tasks can theoretically be represented by a perceptron? Which of them can theoretically be represented by a neural network with one or more hidden layers? (Specify all that apply.)

