General Inference

A Bayesian Network gives us a reasonable way to specify the joint probability of a network of many random variables. Before we celebrate, realize that we still don't know how to use such a network to answer probability questions. There are many techniques for doing so. I am going to introduce you to one of the great ideas in probability for computer science: we can use sampling to solve inference questions on Bayesian networks. Sampling is frequently used in practice because it is relatively easy to understand and easy to implement.

Rejection Sampling

As a warmup consider what it would take to sample an assignment to each of the random variables in our Bayes net. Such a sample is often called a “joint sample” or a “particle” (as in a particle of sand). To sample a particle, simply sample a value for each random variable one at a time based on the value of the random variable’s parents. This means that if X_i is a parent of X_j, you will have to sample a value for X_i before you sample a value for X_j.

1. Sample from $P(\text{Uni} = 1)$: Bern(0.8). Sampled value for Uni is 1.
2. Sample from $P(\text{Influenza} = 1 \mid \text{Uni} = 1)$: Bern(0.2). Sampled value for Influenza is 0.
3. Sample from $P(\text{Fever} = 1 \mid \text{Influenza} = 0)$: Bern(0.05). Sampled value for Fever is 0.
4. Sample from $P(\text{Tired} = 1 \mid \text{Uni} = 1, \text{Influenza} = 0)$: Bern(0.8). Sampled value for Tired is 0.

Thus the sampled particle is: [Uni = 1, Influenza = 0, Fever = 0, Tired = 0]. If we were to run the process again we would get a new particle (with likelihood determined by the joint probability).

Now our strategy is simple: we are going to generate N samples where N is in the hundreds of thousands (if not millions). Then we can compute probability queries by counting. Let $\text{count}(\mathbf{X} = \mathbf{k})$ be notation for the number of particles where random variables $\mathbf{X}$ take on values $\mathbf{k}$. Recall that the bold notation $\mathbf{X}$ means that $\mathbf{X}$ is a vector with one or more elements. By the “frequentist” definition of probability: