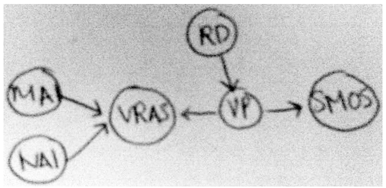
CS181 Assignment 4: Probabilistic Graphical Models and Inference

Alisa Nguyen and Joy Ming

12 April 2013

1. Problem 1

   (a) The directed graphical model (Bayesian network) that shows whether or not there is a virus on the computer can be shown as the following. In our model, the variables MAI, NAI and RD are assumed to be independent from all other variables and do not have any incoming edges tot heir nodes. Whether or not a virus is present tells you exactly whether silly messages appear on the screen (by definition of the problem), so there is a non-noisy edge going directly from VP to $SMOS$. Whether a virus is reported by our antivirus software also depends on whether a virus is present, but to have a report by our antivirus software we must first have antivirus software installed. Lastly, whether or not there was a recent download from an untrusted site has an effect on whether our computer has a virus (virus more likely if true).



The associated conditional probability tables are:

- $P(SMOS|VP)$

SMOS—VP	0	1
0	α	β
1	$1-\alpha$	$1-\beta$

  Example: $\alpha = 1, \beta = 1$

- $P(VP|RD)$

VP—RD	0	1
0	γ	δ
1	$1-\gamma$	$1-\delta$

  Example: $\gamma = 0.8, \delta = 0.2$.

- $P(VRAS|VP, MAI, NAI)$

VRAS—VP,MAI,NAI	0,0,0	0,0,1	0,1,0	1,0,0	1,0,1	1,1,0	0,1,1	1,1,1
0	a	b	c	d	e	f	g	h
1	$1-a$	$1-b$	$1-c$	$1-d$	$1-e$	$1-f$	$1-g$	$1-h$

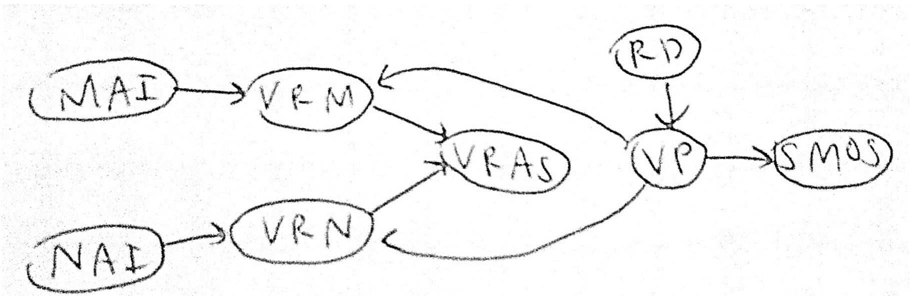
  Example: Each column must sum to 1.
  $a = 0.6, b = 0.2, c = 0.8, d = 0.5, e = 0.3, f = 0.45, g = 0.728, h = 0.586$.

- $P(MAI) = i$. Example: $i = 0.5$
- $P(NAI) = j$. Example: $j = 0.5$

- $P(RD) = k$. Example: $k = 0.583$

(b) An edge that could have possibly been added to the graph is the one between MAI and NAI. Most antivirus software warn the user before downloading to remove all other antivirus software before using the new one, so it is very improbable that a user would have both McAfee and Norton antivirus active at the same time. Although when creating our graphical model, we assumed that whether one antivirus software was present is independent from another, it is plausible to not assume so and add that edge in between MAI and NAI, going both ways. Having one of the antivirus softwares installed in real life significantly lowers the probability that any other one is also installed.

(c) The Bayesian network that includes these new variables would look like:



(d) Including nodes VRM and VRN helps the process because it allows us to take into account more information, most notably the accuracies of different antivirus software in detecting viruses. Based on the different weights we give to VRM and VRN, we gain more information about how accurate $VRAS$ when it either reports a virus is present or not, hopefully removing some of the noise from part (a). However, including these nodes also hinders the modeling process because the model now is more complex and takes up more space. Because we also modified the arrow originally pointing to $VRAS$ from VP to go to both VRM and VRN instead, using these two nodes leaves the $VRAS$ node almost unnecessary, allowing us to reduce the complexity slightly. Still, the tradeoff between representing enough information and complexity of the model still remains an important question, as it was when we were using decision trees and neural networks to model as well.

The new conditional probability tables are:

- $P(VRM|MAI, VP)$

VRM—MAI, VP	0,0	0,1	1,0	1,1
0	α	β	γ	δ
1	$1-\alpha$	$1-\beta$	$1-\gamma$	$1-\delta$

Example: $\alpha = 0.8, \beta = 0.6, \gamma = 0.4, \delta = 0.2$

- $P(VRN|NAI, VP)$

VRM—NAI, VP	0,0	0,1	1,0	1,1
0	a	b	c	d
1	$1-a$	$1-b$	$1-c$	$1-d$

Example: $a = 0.7, b = 0.5, c = 0.3, d = 0.1$

- $P(VRAS|VRM, VRN)$

VRAS—VRM, VRN	0,0	0,1	1,0	1,1
0	a	b	c	d
1	$1-a$	$1-b$	$1-c$	$1-d$

Example: $a = 0.99, b = 0.8, c = 0.7, d = 0.05$

2. Problem 2

   (a) Conditonal Independence

   In order to determine conditional independence, we used the D-separation algorithm. Two nodes A and B are conditionally independent given C in the case that all paths are "blocked" from A to B. A path is blocked if it includes a node such that arrows in the path meet head-to-tail or tail-to-tail at C. It is also blocked if arrows meet head-to-head at a node in the path between A and B and neither the node or its descendants is C.

   - We found that in example (a), $\boxed{\text{no variables}}$ are conditionally independent of A given B because no node had all paths to A blocked. All the paths could travel through C, which was not blocked. In other words, given knowledge of B, we can induce dependence from A to any other node in the graph.
   - We found that in example (b), $\boxed{\text{variables } C \text{ and } F}$ were conditionally independent of A given J because they are the only ones that pass through I, at which point the arrows meet head-to-head and neither the node nor its descendants is J, the given node.

   (b) Factored Distributions

   - For graph (a), the joint probability distribution is

     $$p(A,B,C,D,E,F,G,H,I) =$$

     $$p(A|B,C)p(B|D)p(C|E,F)p(D|G)p(E|G)p(F|H)p(G)p(H)p(I|G,H)$$

     This is equivalent to:

     $$\frac{1}{Z}\psi_1(A,B,C)\psi_2(D,B)\psi_3(C,E,F)\psi_4(D,G)\psi_5(G,E)\psi_6(G,I,H)\psi_7(F,H)$$

     where Z is the normalizaioton constant.
   - For graph (b), the joint probability distribution is

     $$p(A,B,C,D,E,F,G,H,I,J) =$$

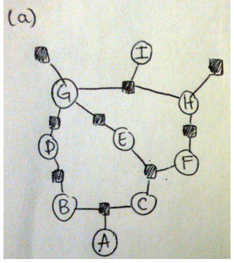
     $$p(A)p(B)p(C)p(D|B)p(E|B)p(F|C)p(G|A,D)p(H|D,E)p(I|E,G)p(J|G)$$

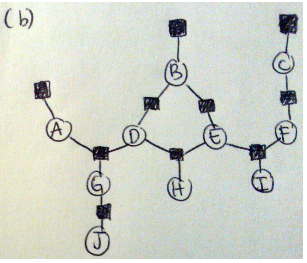
     This is equivalent to:

     $$\frac{1}{Z}\psi_1(A,D,G)\psi_2(G,J)\psi_3(D,B)\psi_4(D,E,H)\psi_5(E,B)\psi_6(E,F,I)\psi_7(F,C)$$

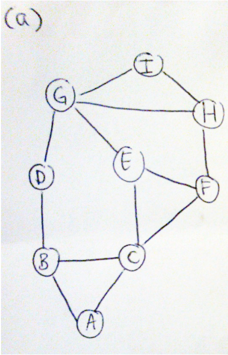
     where Z is the normalizaioton constant.

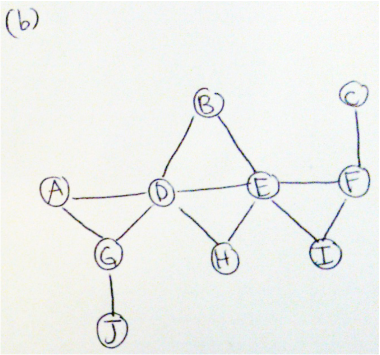
   (c) Factor Graphs

   - 



   (d) Undirected Graphs

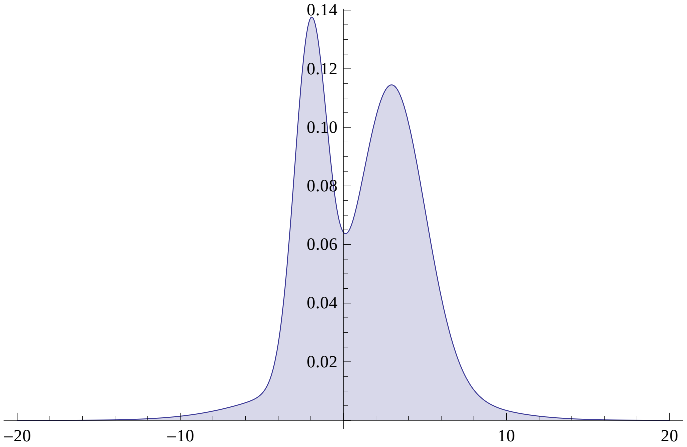
- 



3. Problem 3

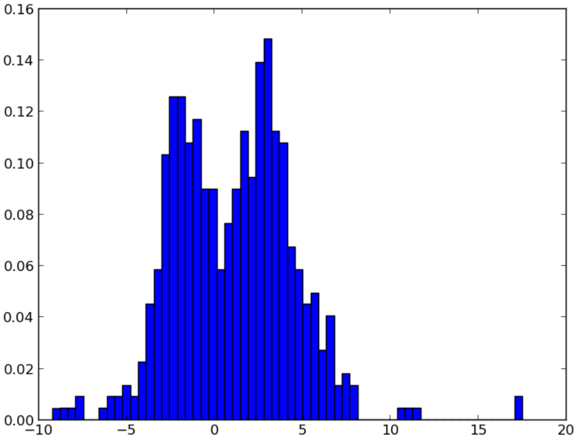
   (a) Density Function

$$f(x) = 0.2 \cdot N(x|\mu = 1, \sigma^2 = 25) + 0.3 \cdot N(x|\mu = -2, \sigma^2 = 1) + 0.5 \cdot N(x|\mu = 3, \sigma^2 = 4)$$



   (b) Direct Sampling

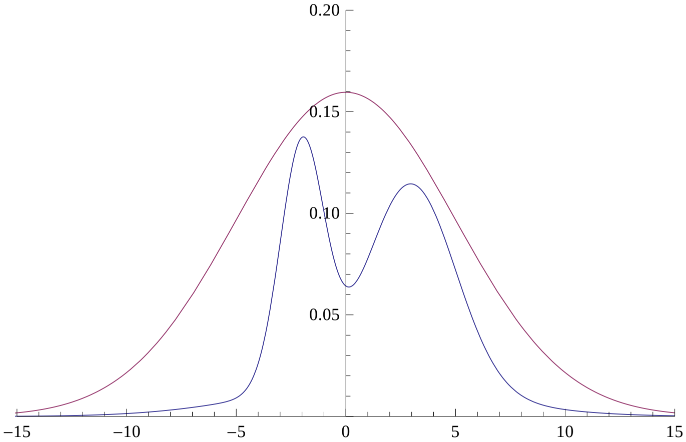
   In order to apply direct sampling to the data, we applied the function `direct()`, where we uniformly chose a random value between 0 and 1 which determined which normal distribution to generate a value from. We picked from the first normal distribution term with probability 0.2, the second with probability 0.3 and the last with probability 0.5. Our results from 500 samples drawn this way are recorded in the following histogram.



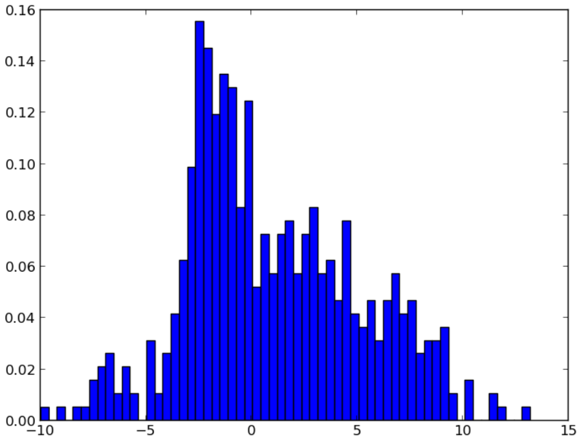
(c) Rejection Sampling

In order to apply rejection sampling, we applied the function `rejection()`, where we drew a variable based on an upper bound function, which we determined to be normal distribution with mean 0, variance 25, and constant 2, resulting in $q(x) = N(x|\mu = 0, \sigma^2 = 25)$ and $c = 2$. $c \cdot q(x)$ is graphed by the purple line below and our density function $f(x)$ is graphed by the blue line below

$$c \cdot q(x) = 2 \cdot N(x|\mu = 0, \sigma^2 = 25)$$

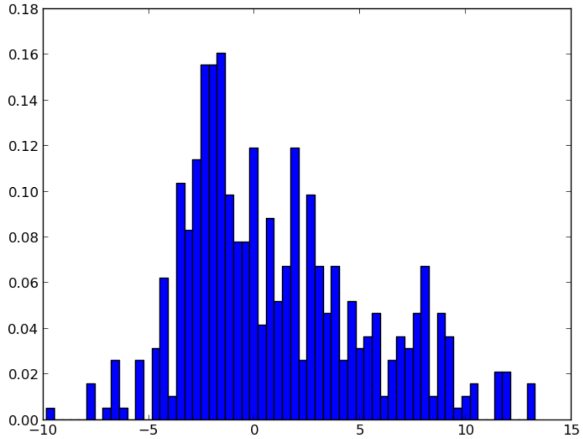


When drawing samples to make a histogram, we get $\boxed{456}$ rejections before we get 500 acceptances. The histogram is pictured below:

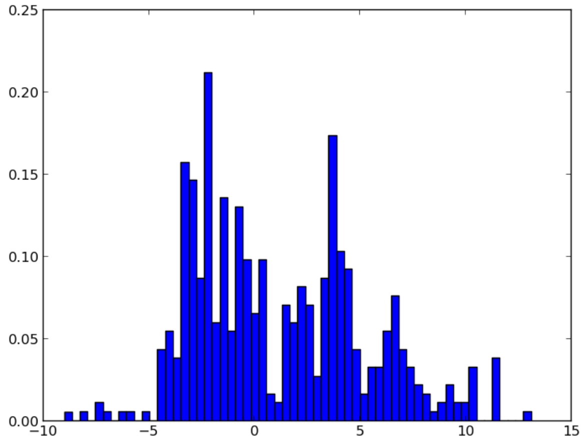


(d) Metropolis Hasting

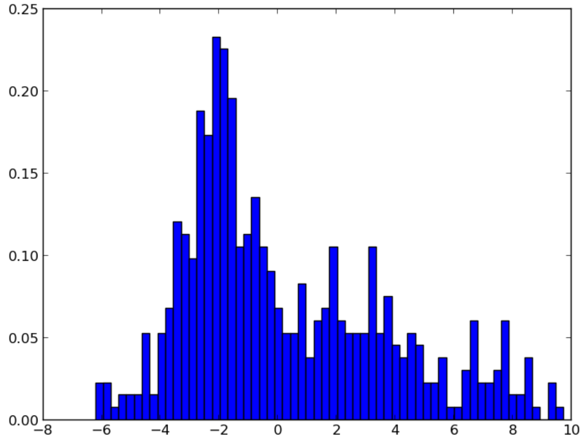
In order to apply the Metropolis Hasting sampling, we initalized our `metrohast()` function and drew a variable based on a Normal function with mean that is the previous x and variance 25 that is accepted based on the acceptance probability to get 304 acceptances out of the 500 samples. This is an acceptance rate of $\boxed{0.608}$ This results in the histogram below:



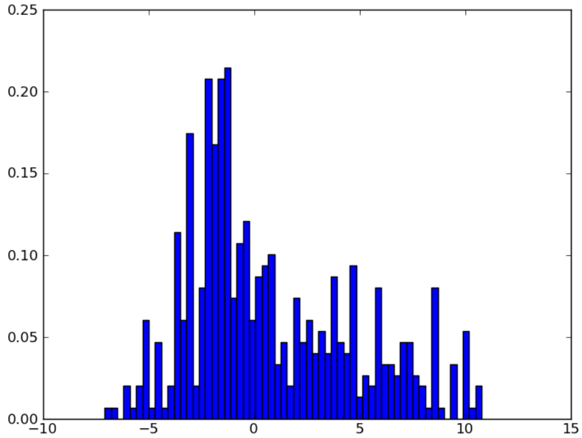
We picked variance 25 at first because it was a value we happened to use for our bounding function the first time around. If we change the variance to be different variables, we get different results. For example if we change the variance to be 100, we get 212 acceptances and an acceptance rate of 0.424 as shown in the following histogram:



Though the 0.424 acceptance rate is slightly better compared to when variance is 25, we wanted to try a smaller number. And if we change the variance to be 1, we get 431 acceptances out of 500, an acceptance rate of 0.862 and the following histogram:



It seems that the lower variance leads to more acceptances out of the 500 trials, and the higher the variance the less acceptances. We tried one more with variance 36, and got 272 acceptances, an acceptance rate of 0.544 and the following histogram:



As discussed in class, the ideal acceptance rate is about 0.5, and based on our experimental results, a variance of 36 was closest to that result. It is ideal to reject half and accept half the samples because we want to converge towards the target distribution as quickly as possible. If the acceptance rate is too big, the convergence takes too long and the distribution is not quite close enough (see graph with variance 100). If the acceptance rate is too small, the convergence doesn't actually happen, instead, the distribution skips and jumps over important features of the target distribution (see graph with variance 1). In order to get an ideal smooth distribution, we need a "just-right" variance, which we found to be about 36.