

Bayes Net Inference using Stochastic Sampling

Jesse Hoey

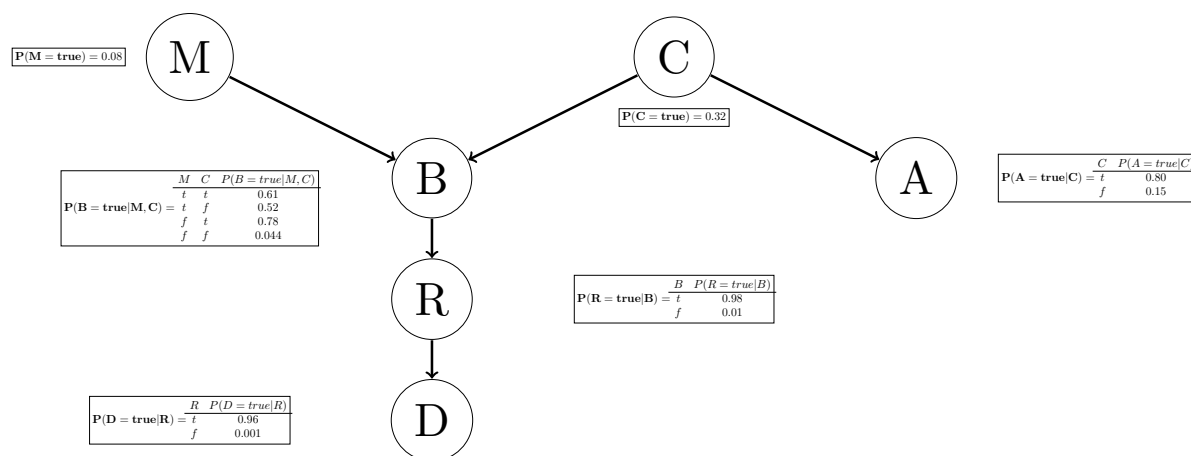
David R. Cheriton School of Computer Science

University of Waterloo

Waterloo, Ontario, CANADA, N2L3G1

jhoey@cs.uwaterloo.ca

Cancer is a disease which occurs with probability 0.32 in a certain population (e.g. older long-time smokers). A Bayes' net is used to diagnose the presence of cancer (C) in such patients by using two binary tests, test A (A) and test B (B) which report the presence or absence of cancer. Test A is a simple test the doctor can perform directly on the patient and see the results immediately. Test A has a true positive rate of 0.8 and a false positive rate of 0.15. That is, when cancer is present, test A detects it 80% of the time, and when cancer is not present, test A reports it 15% of the time. Test B, on the other hand, has greater precision (true positive rate 0.78 and false positive rate 0.044), but requires a complicated machine, which malfunctions (M) with probability 0.08. When the machine malfunctions, test B's true positive rate drops to 0.61 and its false positive rate rises to 0.52. Further, test B's results are not directly available to the doctor. Instead, they are read by a technician who writes them down as a report (R) in a logbook, which is then passed to a data entry person who enters the result in a database (D). The doctor reads the result from the database. The technician and the data entry person sometimes make mistakes, however. The technician's true and false positive rates are 0.98 and 0.01, respectively, while the data entry person's rates are 0.96 and 0.001 (for true and false positive rates, respectively).



M=Machine malfunctions
 C=Patient has cancer
 A=Test A reads positive
 B=Test B reads positive
 R=Report is positive

D=Database entry is positive

Query 1: $P(B|A = false, D = true)$?

Do using full variable elimination first:

No irrelevant variables

$$f_0(C) = P(A = false|C) = \begin{array}{c|c} C & f_0(C) \\ \hline t & 0.2 \\ f & 0.85 \end{array}$$

$$f_1(C) = P(C) = \begin{array}{c|c} C & f_1(C) \\ \hline t & 0.32 \\ f & 0.68 \end{array}$$

$$f_2(B, M, C) = P(B|M, C) = \begin{array}{c|c|c|c} M & C & B & f_1(B, M, C) \\ \hline t & t & t & 0.61 \\ t & t & f & 0.39 \\ t & f & t & 0.52 \\ t & f & f & 0.48 \\ f & t & t & 0.78 \\ f & t & f & 0.22 \\ f & f & t & 0.044 \\ f & f & f & 0.956 \end{array}$$

$$f_3(M) = P(M) = \begin{array}{c|c} M & P(M) \\ \hline t & 0.08 \\ f & 0.92 \end{array}$$

$$f_4(R, B) = P(R|B) = \begin{array}{c|c|c} R & B & f_2(R, B) \\ \hline t & t & 0.98 \\ t & f & 0.01 \\ f & t & 0.02 \\ f & f & 0.99 \end{array}$$

$$f_5(R) = P(D = true|R) = \begin{array}{c|c} R & f_3(R) \\ \hline t & 0.96 \\ f & 0.001 \end{array}$$

so that

$$\begin{aligned}
P(B|A = \text{false}, D = \text{true}) &\propto \sum_{M,R,C} f_0(C)f_1(C)f_2(B, M, C)f_3(M)f_4(R, B)f_5(R) \\
&= \sum_M f_3(M) \sum_C f_7(C)f_2(B, M, C) \sum_R f_4(R, B)f_5(R)
\end{aligned}$$

$$\begin{aligned}
\text{where } f_7(C) &= f_0(C)f_1(C) = \begin{array}{c} \overline{\overline{C}} \\ \overline{f_7(C)} \\ t \quad 0.2 * 0.32 = 0.064 \\ f \quad 0.85 * 0.68 = 0.578 \end{array} \\
&= f_6(B) \sum_M f_3(M) \sum_C f_7(C)f_2(B, M, C)
\end{aligned}$$

$$\begin{aligned}
\text{where } f_6(B) &= \sum_R f_4(R, B)f_5(R) = \begin{array}{c} \overline{\overline{B}} \\ \overline{f_6(B)} \\ t \quad 0.98 * 0.96 + 0.02 * 0.001 = 0.941 \\ f \quad 0.01 * 0.96 + 0.99 * 0.001 = 0.011 \end{array} \\
&= f_6(B) \sum_M f_4(M)f_8(B, M)
\end{aligned}$$

$$\begin{aligned}
\text{where } f_8(M, B) &= \sum_C f_7(C)f_2(B, M, C) = \begin{array}{c} \overline{\overline{M \quad B}} \\ \overline{f_8(M, B)} \\ t \quad t \quad 0.064 * 0.61 + 0.578 * 0.52 = 0.3396 \\ t \quad f \quad 0.064 * 0.39 + 0.578 * 0.48 = 0.302 \\ f \quad t \quad 0.064 * 0.78 + 0.578 * 0.044 = 0.075 \\ f \quad f \quad 0.064 * 0.22 + 0.578 * 0.956 = 0.567 \end{array} \\
&= f_6(B)f_9(B)
\end{aligned}$$

$$\begin{aligned}
\text{where } f_9(B) &= \sum_M f_4(M)f_8(B, M) = \begin{array}{c} \overline{\overline{B}} \\ \overline{f_9(B)} \\ t \quad 0.3396 * .08 + 0.075 * 0.92 = 0.096 \\ f \quad 0.302 * 0.08 + 0.567 * 0.92 = 0.546 \end{array} \\
&= f_{10}(B)
\end{aligned}$$

$$\begin{aligned}
\text{where } f_{10}(B) &= f_6(B)f_9(B) = \begin{array}{c} \overline{\overline{B}} \\ \overline{f_{10}(B)} \\ t \quad 0.096 * 0.941 = 0.090 \\ f \quad 0.546 * 0.011 = 0.0060 \end{array}
\end{aligned}$$

and so,

$$\begin{aligned}
P(B = \text{true}|A = \text{false}, D = \text{true}) &= \begin{array}{c} \overline{\overline{B}} \\ \overline{P(B|A = \text{false}, D = \text{true})} \\ t \quad 0.09/(0.09 + 0.006) = 0.936 \\ f \quad 0.006/(0.09 + 0.006) = 0.062 \end{array}
\end{aligned}$$

Now let us do the same calculation but this time we will be sampling both C and M from their respective priors, such that the proposal distributions are $q(C = \text{true}) = P(C = \text{true}) = 0.32$ and $q(M = \text{true}) = P(M = \text{true}) = 0.08$. We will use exact inference over only the variable R (and the query B), and stochastic inference using the samples $s_i = \{c_i, m_i\}$ for $i = 1, \dots, N$.

$$\begin{aligned}
P(B|D = true, A = false) &\propto \sum_{s_i=\{c_i, m_i\}} P(B, D = true, A = false|c_i, m_i) \\
&= \sum_{s_i} \sum_R P(B, D = true, A = false, R|c_i, m_i)
\end{aligned}$$

factor using the product rule and use conditional independence:

$$= \sum_{s_i} \sum_R P(D = true|R)P(R|B)P(B|c_i, m_i)P(A = false|c_i)$$

distribute sum over R

$$= \sum_{s_i} P(B|c_i, m_i)P(A = false|c_i) \sum_R P(D = true|R)P(R|B)$$

borrow factor f_6 from above

$$= \sum_{s_i} P(B|c_i, m_i)P(A = false|c_i)f_6(B)$$

Now let us draw 100 samples $s_i = \{c_i, m_i\}$, and compute the weight of each as $P(B|c_i, m_i) \times P(A = false|c_i) \times f_6(B)$, which we do for both $B = true$ and $B = false$. For proposal distributions $q(C)$ and $q(M)$, there will be $N_{cm} \approx q(C) \times q(M) \times N$ samples drawn for each of the four combinations of c and m , which we aggregate as:

c	m	N_{cm}	$P(B c, m)P(A = false c)f_6(B)$	
			$B = true$	$B = false$
<i>true</i>	<i>true</i>	$0.32 * 0.08 * 100 = 2$	$0.61 * 0.2 * 0.941 = 0.115$	$0.39 * 0.2 * 0.011 = 0.00086$
<i>true</i>	<i>false</i>	$0.32 * 0.92 * 100 = 30$	$0.52 * 0.2 * 0.941 = 0.098$	$0.48 * 0.2 * 0.011 = 0.000106$
<i>false</i>	<i>true</i>	$0.68 * 0.08 * 100 = 5$	$0.78 * 0.85 * 0.941 = 0.624$	$0.22 * 0.85 * 0.011 = 0.0021$
<i>false</i>	<i>false</i>	$0.68 * 0.92 * 100 = 63$	$0.044 * 0.85 * 0.941 = 0.035$	$0.956 * 0.85 * 0.011 = 0.0089$

And therefore, the total weight= $N_{cm} \times P(B|c, m)P(A = false|c)f_6(B)$ is

$2 * .115 + 30 * 0.098 + 5 * 0.624 + 63 * 0.035 = 8.495$ for $B = true$ and

$2 * 0.00086 + 30 * 0.00106 + 5 * 0.0021 + 63 * 0.0089 = 0.605$ for $B = false$, and thus

$$P(B = true|D = true, A = false) = \frac{8.495}{8.495 + 0.605} = 0.934$$

Notice how only 2 samples are drawn for the $c = true, m = true$ case. This could cause problems (we might get unlucky and draw zero samples, in which case the result would be not as precise. This becomes more important in situations where the parts of the model we are sampling from assign a small weight (Are unlikely) but the parts we are doing exact inference over assign high

weights. In the example above, when $c = false, m = true$, the probability of drawing such a sample is $0.68 * 0.08 = 0.05$, but each such sample has a large weight (of 0.624). Suppose we had gotten unlucky and drawn 0 samples for $c = false, m = true$, instead drawing 68 samples for $c = false, m = false$. Then the calculation would yield:

$$2 * .115 + 30 * 0.098 + 0 * 0.624 + 68 * 0.035 = 5.55 \text{ for } B = true \text{ and}$$

$$2 * 0.00086 + 30 * 0.00106 + 0 * 0.0021 + 68 * 0.0089 = 0.639 \text{ for } B = false, \text{ and thus}$$

$$P(B = true|D = true, A = false) = \frac{5.55}{5.55 + 0.639} = 0.898$$

So, instead, we draw from $q(M = true) = 0.5$, and then weight the samples using $P(m)/q(M)$:

c	m	N_{cm}	$P(B c, m)P(A = false c)f_6(B)$		$P(m)/q(m)$
			$B = true$	$B = false$	
<i>true</i>	<i>true</i>	$0.32 * 0.5 * 100 = 16$	0.115	0.00086	0.16
<i>true</i>	<i>false</i>	$0.32 * 0.5 * 100 = 16$	0.098	0.000106	1.84
<i>false</i>	<i>true</i>	$0.68 * 0.5 * 100 = 34$	0.624	0.0021	0.15
<i>false</i>	<i>false</i>	$0.68 * 0.5 * 100 = 34$	0.035	0.0089	1.84

And therefore, the total weight= $N_{cm} \times P(B|c, m)P(A = false|c)f_6(B) \frac{P(m)}{q(M)}$ is

$$16 * 0.115 * 0.16 + 16 * 0.098 * 1.84 + 34 * 0.624 * 0.16 + 34 * 0.035 * 1.84 = 8.76 \text{ for } B = true \text{ and}$$

$$16 * 0.00086 * 0.16 + 16 * 0.00106 * 1.84 + 34 * 0.0021 * 0.16 + 34 * 0.0089 * 1.84 = 0.602 \text{ for } B = false, \text{ and thus}$$

$$P(B = true|D = true, A = false) = \frac{8.76}{8.76 + 0.602} = 0.936$$

Finally, if we also sampled from $q(C = true) = 0.5$, we could have 25 samples in each category, each with weights $P(m) * P(c)/(q(M) * q(C))$:

c	m	N_{cm}	$P(B c, m)P(A = false c)f_6(B)$		$P(c) * P(m)/(q(m) * q(c))$
			$B = true$	$B = false$	
<i>true</i>	<i>true</i>	$0.5 * 0.5 * 100 = 25$	0.115	0.00086	0.1024
<i>true</i>	<i>false</i>	$0.5 * 0.5 * 100 = 25$	0.098	0.000106	1.1776
<i>false</i>	<i>true</i>	$0.5 * 0.5 * 100 = 25$	0.624	0.0021	0.2176
<i>false</i>	<i>false</i>	$0.5 * 0.5 * 100 = 25$	0.035	0.0089	2.504

And therefore, the total weight= $N_{cm} \times P(B|c, m)P(A = false|c)f_6(B) \frac{P(m)}{q(M)}$ is

$$25 * .115 * 0.1024 + 25 * 0.098 * 1.1776 + 25 * 0.624 * 0.2176 + 25 * 0.035 * 2.504 = 8.765 \text{ for } B = true \text{ and}$$

and

$$25 * 0.00086 * 0.1024 + 25 * 0.00106 * 1.1776 + 25 * 0.0021 * 0.2176 + 25 * 0.0089 * 2.504 = 0.6019 \text{ for } B = false, \text{ and thus}$$

$$P(B = true|D = true, A = false) = \frac{8.765}{8.765 + 0.6019} = 0.936$$

The key idea is we can use any proposal we want to get a better sample distribution, and then weight each sample according to how unlikely it is. So, the underrepresented classes get more samples, but with smaller weights, whereas the overrepresented classes get fewer samples with larger weights.

Query 2: $P(D|A = true)$?

Exercise: do this using variable elimination. The answer is $P(D = true|A = true) = 0.54$.

Let us solve this by sampling over B, D and R and doing exact inference over M and C . To do this, we will write

$$\begin{aligned}
 P(D|A = true) &= \sum_{s_i=\{b_i, r_i, d_i\}} P(D|A = true, d_i, r_i, b_i) \\
 &= \sum_{s_i=\{b_i, r_i, d_i\}} \sum_{B, R} P(D, B, R|A = true, d_i, r_i, b_i) \\
 &= \sum_{s_i=\{b_i, r_i, d_i\}} \sum_R P(D|R, d_i, r_i, b_i) \sum_B P(R|B, d_i, r_i, b_i) P(B|A = true, d_i, r_i, b_i) \\
 &= \sum_{s_i=\{b_i, r_i, d_i\}} \sum_R P(D|R, d_i, r_i) \sum_B P(R|B, r_i, b_i) P(B|A = true, b_i)
 \end{aligned}$$

Terms like $P(D|R, d_i, r_i) = 1$ if $D = d_i \wedge R = r_i$ and 0 otherwise, so they essentially just fold the sums over R and D into the sample counts. We can therefore draw samples b_i, r_i, d_i from some proposals over variables B, R, D , and weight each sample so generated by $P(b_i|A = true)$. To do this, let's first compute these weights by writing

$$P(B|A = true) = \sum_{M, C} P(B|M, C)P(M)P(C)P(A = true|C)$$

We have factors:

$$\begin{aligned}
 f_0(C) = P(A = true|C) &= \begin{array}{c} \overline{\overline{C}} \\ t \quad 0.8 \\ f \quad 0.15 \end{array} \\
 f_1(C) = P(C) &= \begin{array}{c} \overline{\overline{C}} \\ t \quad 0.32 \\ f \quad 0.68 \end{array} \\
 f_2(B, M, C) = P(B|M, C) &= \begin{array}{c} \overline{\overline{M}} \quad \overline{\overline{C}} \quad \overline{\overline{B}} \\ t \quad t \quad t \quad 0.61 \\ t \quad t \quad f \quad 0.39 \\ t \quad f \quad t \quad 0.52 \\ t \quad f \quad f \quad 0.48 \\ f \quad t \quad t \quad 0.78 \\ f \quad t \quad f \quad 0.22 \\ f \quad f \quad t \quad 0.044 \\ f \quad f \quad f \quad 0.956 \end{array} \\
 f_3(M) = P(M) &= \begin{array}{c} \overline{\overline{M}} \\ t \quad 0.08 \\ f \quad 0.92 \end{array}
 \end{aligned}$$

and so

$$P(B|A = true) \propto \sum_C f_1(C) f_0(C) \sum_M f_2(B, M, C) f_3(M)$$

$$= \sum_C f_7(C) \sum_M f_2(B, M, C) f_3(M)$$

where $f_4(C) = f_0(C) f_1(C) = \begin{array}{c} C \\ t \\ f \end{array} \begin{array}{c} f_4(C) \\ 0.8 * 0.32 = 0.256 \\ 0.15 * 0.68 = 0.102 \end{array}$

$$= \sum_C f_4(C) f_5(C, B)$$

where $f_5(C, B) = \sum_M f_2(B, M, C) f_3(M) = \begin{array}{c} B \ C \\ t \ t \\ t \ f \\ f \ t \\ f \ f \end{array} \begin{array}{c} f_5(C, B) \\ 0.61 * 0.08 + 0.78 * 0.92 = 0.7664 \\ 0.39 * 0.08 + 0.22 * 0.92 = 0.2336 \\ 0.52 * 0.08 + 0.044 * 0.92 = 0.082 \\ 0.48 * 0.08 + 0.956 * 0.92 = 0.918 \end{array}$

$$= f_6(B)$$

where $f_6(B) = \sum_C f_4(C) f_5(C, B) = \begin{array}{c} B \\ t \\ f \end{array} \begin{array}{c} f_6(B) \\ 0.7664 * 0.256 + 0.082 * 0.102 = 0.205 \\ 0.2336 * 0.256 + 0.918 * 0.102 = 0.153 \end{array}$

Now suppose we draw N samples over B using $q(B = true) = 0.5$, and then draw samples for R from $q(R) = P(R|B)$ and for D from $q(D) = P(D|R)$, then we get the following distribution of samples (with N_{brd} giving the expected number of samples drawn for each value of B, R, D) with weights given by $\frac{f_6(B)}{q(B)}$:

b	r	d	N_{brd}	$f_6(B)/q(B)$
<i>true</i>	<i>true</i>	<i>true</i>	$\frac{N}{2} * 0.98 * 0.96$	$0.205/2 = 0.1025$
<i>true</i>	<i>true</i>	<i>false</i>	$\frac{N}{2} * 0.98 * 0.04$	0.1025
<i>true</i>	<i>false</i>	<i>true</i>	$\frac{N}{2} * 0.02 * 0.001$	0.1025
<i>true</i>	<i>false</i>	<i>false</i>	$\frac{N}{2} * 0.02 * 0.999$	0.1025
<i>false</i>	<i>true</i>	<i>true</i>	$\frac{N}{2} * 0.01 * 0.96$	$0.153/2 = 0.0765$
<i>false</i>	<i>true</i>	<i>false</i>	$\frac{N}{2} * 0.01 * 0.04$	0.0765
<i>false</i>	<i>false</i>	<i>true</i>	$\frac{N}{2} * 0.99 * 0.001$	0.0765
<i>false</i>	<i>false</i>	<i>false</i>	$\frac{N}{2} * 0.99 * 0.999$	0.0765

Therefore, summing all these for $D = true$ gives $\frac{N}{2} * 0.0972$ and for $D = false$ gives $\frac{N}{2} * 0.082$ and so

$$P(D = true|A = true) = \frac{0.0972}{0.0972 + 0.082} = 0.54$$