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).
D=Database entry is positive

Query 1: \( P(B|A = false, D = true) \)?
Do using full variable elimination first:

No irrelevant variables

\[
\begin{align*}
f_0(C) &= P(A = false|C) = \\
&\begin{array}{lr}
C & f_0(C) \\
\hline 
 t & 0.2 \\
 f & 0.85 \\
\end{array} \\
f_1(C) &= P(C) = \\
&\begin{array}{lr}
C & f_1(C) \\
\hline 
 t & 0.32 \\
 f & 0.68 \\
\end{array} \\
\begin{array}{ccc}
M & C & B \\
\hline 
n & n & n & 0.61 \\
n & n & f & 0.39 \\
n & f & n & 0.52 \\
\end{array} \\
f_2(B, M, C) &= P(B|M, C) = \\
&\begin{array}{cccc}
B & M & C \\
\hline 
n & n & n & 0.48 \\
n & n & f & 0.78 \\
n & f & n & 0.22 \\
n & f & f & 0.044 \\
n & f & f & 0.956 \\
\end{array} \\
f_3(M) &= P(M) = \\
&\begin{array}{lr}
M & P(M) \\
\hline 
n & 0.08 \\
n & 0.92 \\
\end{array} \\
f_4(R, B) &= P(R|B) = \\
&\begin{array}{lr}
R & f_4(R, B) \\
\hline 
n & n & 0.98 \\
n & n & f & 0.02 \\
n & n & f & 0.01 \\
n & f & f & 0.99 \\
\end{array} \\
f_5(R) &= P(D = true|R) = \\
&\begin{array}{lr}
R & f_5(R) \\
\hline 
n & n & 0.96 \\
n & n & f & 0.001 \\
\end{array}
\]
so that

\[
P(B|A = \text{false}, D = \text{true}) = \sum_{M,R,C} f_0(C)f_1(C)f_2(B,M,C)f_3(M)f_4(R,B)f_5(R) \\
= \sum_M f_3(M)\sum_C f_7(C)f_2(B,M,C)\sum R f_4(R,B)f_5(R) \\
\]

where

\[
f_7(C) = f_0(C)f_1(C) = \begin{array}{c|c}
    t & 0.2 \times 0.32 = 0.064 \\
    f & 0.85 \times 0.68 = 0.578 \\
\end{array}
\]

\[
f_6(B) = \sum_R f_4(R,B)f_5(R) = \begin{array}{c|c}
    t & 0.98 \times 0.96 + 0.02 \times 0.001 = 0.941 \\
    f & 0.01 \times 0.96 + 0.99 \times 0.001 = 0.011 \\
\end{array}
\]

\[
f_8(M,B) = \sum_C f_7(C)f_2(B,M,C) = \begin{array}{c|c|c}
    M & B & f_8(M,B) \\
    \hline
    t & t & 0.64 \times 0.61 + 0.578 \times 0.52 = 0.3396 \\
    t & f & 0.64 \times 0.39 + 0.578 \times 0.48 = 0.302 \\
    f & t & 0.064 \times 0.78 + 0.578 \times 0.044 = 0.075 \\
    f & f & 0.064 \times 0.22 + 0.578 \times 0.956 = 0.567 \\
\end{array}
\]

\[
f_9(B) = \sum_M f_4(M)f_8(B,M) = \begin{array}{c|c|c}
    t & .3396 \times 0.08 + 0.075 \times 0.92 = 0.096 \\
    f & 0.302 \times 0.08 + 0.567 \times 0.92 = 0.546 \\
\end{array}
\]

\[
f_10(B) = f_6(B)f_9(B) = \begin{array}{c|c|c}
    B & f_9(B) \\
    \hline
    t & 0.096 \times 0.941 = 0.090 \\
    f & 0.546 \times 0.011 = 0.0060 \\
\end{array}
\]

and so,

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

Now let us do the same calculation but this time we will be sampling both C and M from their respective priors, such that \(q(C = \text{true}) = P(C = \text{true}) = 0.32\) 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 \(i = 1, \ldots, N\).
\[ P(B|D = \text{true}, A = \text{false}) \propto \sum_{s_i = \{c_i, m_i\}} P(B, D = \text{true}, A = \text{false}|c_i, m_i) \]
\[ = \sum_{s_i} \sum_R P(D = \text{true}|R) P(R|B) P(B|c_i, m_i) P(A = \text{false}|c_i) \]
\[ \text{factor using the product rule and use conditional independence:} \]
\[ = \sum_{s_i} \sum_R P(D = \text{true}|R) P(R|B) P(B|c_i, m_i) P(A = \text{false}|c_i) \]
\[ \text{distribute sum over } R \]
\[ = \sum_{s_i} P(B|c_i, m_i) P(A = \text{false}|c_i) \sum_R P(D = \text{true}|R) P(R|B) \]
\[ \text{borrow factor } f_6 \text{ from above} \]
\[ = \sum_{s_i} P(B|c_i, m_i) P(A = \text{false}|c_i) f_6(B) \]

Now let us draw 100 samples \( s_i = \{c_i, m_i\} \), and compute the weight of each as \( P(B|c_i, m_i) \times P(A = \text{false}|c_i) \times f_6(B) \), which we do for both \( B = \text{true} \) and \( B = \text{false} \). 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:

<table>
<thead>
<tr>
<th>( c )</th>
<th>( m )</th>
<th>( N_{cm} )</th>
<th>( B = \text{true} )</th>
<th>( B = \text{false} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>true</td>
<td>true</td>
<td>0.32 * 0.08 * 100 = 2</td>
<td>0.61 * 0.2 * 0.941 = 0.115</td>
<td>0.39 * 0.2 * 0.0011 = 0.00086</td>
</tr>
<tr>
<td>true</td>
<td>false</td>
<td>0.32 * 0.92 * 100 = 30</td>
<td>0.52 * 0.2 * 0.941 = 0.098</td>
<td>0.48 * 0.2 * 0.0011 = 0.000106</td>
</tr>
<tr>
<td>false</td>
<td>true</td>
<td>0.68 * 0.08 * 100 = 5</td>
<td>0.78 * 0.85 * 0.941 = 0.624</td>
<td>0.22 * 0.85 * 0.0011 = 0.0021</td>
</tr>
<tr>
<td>false</td>
<td>false</td>
<td>0.68 * 0.92 * 100 = 63</td>
<td>0.044 * 0.85 * 0.941 = 0.035</td>
<td>0.956 * 0.85 * 0.0011 = 0.0089</td>
</tr>
</tbody>
</table>

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

\[
2 * 1.115 + 30 * 0.098 + 5 * 0.624 + 63 * 0.035 = 8.495 \text{ for } B = \text{true} \text{ and } \\
2 * 0.00086 + 30 * 0.00106 + 5 * 0.0021 + 63 * 0.0089 = 0.605 \text{ for } B = \text{false}, \text{ and thus} \\
\]

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

Notice how only 2 samples are drawn for the \( c = \text{true}, m = \text{true} \) case. This could cause problems (we might get unlucky and draw zero samples, in which case the result would be not as precise.

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

<table>
<thead>
<tr>
<th>( c )</th>
<th>( m )</th>
<th>( N_{cm} )</th>
<th>( B = \text{true} )</th>
<th>( B = \text{false} )</th>
<th>( P(m)/q(m) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>true</td>
<td>true</td>
<td>0.32 * 0.5 * 100 = 16</td>
<td>0.115</td>
<td>0.00086</td>
<td>0.04</td>
</tr>
<tr>
<td>true</td>
<td>false</td>
<td>0.32 * 0.5 * 100 = 16</td>
<td>0.098</td>
<td>0.000106</td>
<td>0.45</td>
</tr>
<tr>
<td>false</td>
<td>true</td>
<td>0.68 * 0.5 * 100 = 34</td>
<td>0.624</td>
<td>0.0021</td>
<td>0.04</td>
</tr>
<tr>
<td>false</td>
<td>false</td>
<td>0.68 * 0.5 * 100 = 34</td>
<td>0.035</td>
<td>0.0089</td>
<td>0.46</td>
</tr>
</tbody>
</table>
And therefore, the total weight $= N_{cm} \times P(B|c, m) P(A = false|c) f_0(B) \frac{P(m)}{q(M)}$ is
\[
16 \times 0.115 \times 0.46 + 16 \times 0.098 \times 0.46 + 16 \times 0.624 \times 0.46 + 16 \times 0.035 \times 0.46 = 2.19 \text{ for } B = true \text{ and } 16 \times 0.00086 \times 0.46 + 16 \times 0.00106 \times 0.46 + 16 \times 0.0021 \times 0.46 + 16 \times 0.0089 \times 0.46 = 0.154 \text{ for } B = false, 
\]
and thus
\[
P(B = true|D = true, A = false) = \frac{2.19}{2.19 + 0.154} = 0.934
\]
Finally, if we also sampled from $q(C = true) = 0.5$, we could have 25 samples in each category, each with weights $P(m) \times P(c)/(q(M) \times q(C))$:

| $c$ | $m$ | $N_{cm}$ | $P(B|c, m) P(A = false|c) f_0(B)$ | $P(c) \times P(m)/(q(m) \times q(c))$ |
|-----|-----|----------|-------------------------------|-------------------------------------|
| true | true | 0.5 \times 0.5 \times 100 = 25 | 0.115 | 0.00086 | 0.1024 |
| true | false | 0.5 \times 0.5 \times 100 = 25 | 0.098 | 0.00106 | 1.1776 |
| false | true | 0.5 \times 0.5 \times 100 = 25 | 0.624 | 0.0021 | 0.2176 |
| false | false | 0.5 \times 0.5 \times 100 = 25 | 0.035 | 0.0089 | 2.504 |

And therefore, the total weight $= N_{cm} \times P(B|c, m) P(A = false|c) f_0(B) \frac{P(m)}{q(M)}$ is
\[
25 \times 0.115 \times 0.1024 + 25 \times 0.098 \times 1.1776 + 25 \times 0.624 \times 0.2176 + 25 \times 0.035 \times 2.504 = 8.765 \text{ for } B = true \text{ and } 25 \times 0.00086 \times 0.1024 + 25 \times 0.00106 \times 1.1776 + 25 \times 0.0021 \times 0.2176 + 25 \times 0.0089 \times 2.504 = 0.6019 \text{ for } B = false, 
\]
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.