3.4 Solutions to Bayes’ theorem examples of Section 3.2

3.4.1 Girls named Lulu

Case 1 A family has two children. What is the probability that both are girls? Hopefully you said \( \frac{1}{4} \).

Case 2 A family has two children. Given that one of them is a girl, what is the probability that both are girls? Answer: \( \frac{1}{3} \).

- Define

\[
x = \begin{cases} 1 : & \text{no boys in the family} \\ 0 : & \text{boy in the family} \end{cases} \quad y = \begin{cases} 1 : & \text{no girls in the family} \\ 0 : & \text{girl in the family} \end{cases}
\]

\[
f_x(1) = \frac{1}{4}, \quad f_x(0) = \frac{3}{4}, \quad f_y(1) = \frac{1}{4}, \quad f_y(0) = \frac{3}{4}
\]

We want to know \( f_{x\mid y}(1\mid 0) \):

\[
f_{x\mid y}(1\mid 0) = \frac{f_y(0\mid 1)f_x(1)}{f_y(0)} = \frac{1 \times \frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}
\]

- Other approach: (B,B), (B,G), (G,B), (G,G).

Case 3 A family has two children. Given that one of them is a girl named Lulu, what is the probability that both are girls?

Here, the key is that Lulu is an unusual name for a girl.

- Define

\[
x = \begin{cases} 1 : & \text{no boys in the family} \\ 0 : & \text{boy in the family} \end{cases} \quad y = \begin{cases} 1 : & \text{no girl named Lulu in the family} \\ 0 : & \text{girl named Lulu in the family} \end{cases}
\]

\[
f_x(1) = \frac{1}{4}, \quad f_x(0) = \frac{3}{4}, \quad f_y(1) = \frac{1}{4}, \quad f_y(0) = \frac{3}{4}
\]

\[
\begin{array}{c|c|c}
\text{c_1} & \text{c_2} & f(c_1, c_2) \\
\hline
0 & 0 & (1-p)(1-p) \\
0 & 1 & (1-p)p \\
1 & 0 & p \\
1 & 1 & 0 \\
\end{array}
\]

Therefore \( f_{y\mid x}(0\mid 1) = 2p - p^2 \approx 2p \) since \( p \ll 1 \). This is somewhat intuitive, probabilities of independent, low probability events roughly add.

- What is \( f_{y\mid x}(0\mid 1) \)? Given that they are both girls, what is probability that there is a girl named Lulu in the family? Let \( p \) be the probability that someone names a girl Lulu, where we assume that \( p \ll 1 \) (do you know anyone named Lulu?). Then consider the random variables \( c_1 \) and \( c_2 \), where \( c_1 \) refers to the first child, \( c_2 \) to the second child:

\[
c_1 = \begin{cases} 0 : & \text{not named Lulu} \\ 1 : & \text{named Lulu} \end{cases}
\]

\[
\begin{array}{c|c|c}
\text{c_1} & \text{c_2} & f(c_1, c_2) \\
\hline
0 & 0 & (1-p)(1-p) \\
0 & 1 & (1-p)p \\
1 & 0 & p \\
1 & 1 & 0 \\
\end{array}
\]

- How about \( f_y(0) \) - the probability that there is a girl named Lulu in the family? Proceed similarly to above:

\[
c_1 = \begin{cases} 0 : & \text{boy} \\ 1 : & \text{girl, not named Lulu} \\ 2 : & \text{girl, named Lulu} \end{cases}
\]
\[
\begin{array}{c|c|c}
      & c_1 & c_2 \mid f(c_1, c_2) \\
0 & 0 & 0.25 \\
0 & 1 & 0.25(1 - p) \\
0 & 2 & 0.25p \\
1 & 0 & 0.25(1 - p) \\
1 & 1 & 0.25(1 - p)^2 \\
1 & 2 & 0.25p(1 - p) \\
2 & 0 & 0.25p \\
2 & 1 & 0.25p \\
2 & 2 & 0 \\
\end{array}
\]

Sanity check: (sum) 0.25(1 + 1 - p + 1 - p + 1 - 2p + p^2 + p - p^2 + 2p) = 0.25(4) = 1.
\(f_y(0) = 0.75p + 0.25p - 0.25p^2 \approx p\). Again, this is somewhat intuitive: the probability that any child is named Lulu is roughly 0.5p, and since this is a low probability event, the probabilities for two children roughly add.

Therefore
\[
\begin{align*}
f_x|y(1|0) & \approx 2p \times \frac{1}{p} = \frac{1}{2} \\
\end{align*}
\]

Even though you can follow all the steps, the end result probably still seems wrong.

An alternate, heuristic way to calculate this: assume 1 in 1000 girls is named Lulu. Of 100,000 families, 75,000 will have at least one girl: 50,000 will have a girl and a boy, and 25,000 will have two girls. Of the 50,000 girl/boy families, we expect 50 to have a girl named Lulu. Of the 25,000 girl/girl families, we expect 50 to have a girl named Lulu: 25 where the first-born is Lulu, 25 where the second born is Lulu. The probability is thus \(\frac{1}{2}\).

### 3.4.2 Disease diagnosis

Define the following random variables:
\[
x = \begin{cases} 
1 : \text{patient does not have cancer} \\
0 : \text{patient has cancer}
\end{cases}
\]
\[
y = \begin{cases} 
1 : \text{test provides a negative result} \\
0 : \text{test provides a positive result}
\end{cases}
\]

We want to know \(f_x|y(0|0)\):
\[
f_x|y(0|0) = \frac{f_{y|x}(0|0) f_x(0)}{f_y(0)}
\]
\(f_{y|x}(0|0) = 0.9\) (false negative rate = 10%)
\(f_y(0) = 0.008\)
\(f_y(0) = f_{y|x}(0|0) f_x(0) + f_{y|x}(0|1) f_x(1)\) (Total probability theorem: \(f(y) = \sum_x f(y|x) f(x)\))
\[
= 0.90 \times 0.008 + 0.07 \times 0.992
\]
\[
\therefore f_x|y(0|0) = \frac{0.9 \times 0.008 + 0.07 \times 0.992}{0.9 \times 0.008 + 0.07 \times 0.992} \approx 0.094
\]

The probability is 9.4% !!!! Reason: most positive results are due to false positives.

### 3.4.3 Prosecutor’s fallacy

Problems with the conviction:

- The events are \textit{not} independent. A more detailed study showed, chances of two cases of SIDS are 1 in 2.75 million - however, these are still small odds.
Inversion: the probability that two children die of SIDS is actually of no interest. What we want to know:

(1) Given that two children have died, what is the probability that they died of SIDS?

(2) Given that two children have died, what is the probability that they were murdered?

These were calculated, and (1) is nine times more likely than (2).

The conviction was eventually overturned.