Here’s my solution to the hard exercises in chapter 2 of McElreath’s Statistical Rethinking, 1st edition. When writing this up, came across a [very relevant article](https://www.theguardian.com/world/2020/feb/28/red-pandas-are-actually-two-separate-species-study-finds). We’ll solve these problems in two ways: using the counting method and using Bayes rule.

\(\DeclareMathOperator{\dbinomial}{Binomial} \DeclareMathOperator{\dbernoulli}{Bernoulli} \DeclareMathOperator{\dpoisson}{Poisson} \DeclareMathOperator{\dnormal}{Normal} \DeclareMathOperator{\dt}{t} \DeclareMathOperator{\dcauchy}{Cauchy} \DeclareMathOperator{\dexponential}{Exp} \DeclareMathOperator{\duniform}{Uniform} \DeclareMathOperator{\dgamma}{Gamma} \DeclareMathOperator{\dinvpamma}{Invpamma} \DeclareMathOperator{\invlogit}{InvLogit} \DeclareMathOperator{\logit}{Logit} \DeclareMathOperator{\ddirichlet}{Dirichlet} \DeclareMathOperator{\dbeta}{Beta}\)

**Counting method**

Let’s generate a dataset with all the features necessary to solve all the questions: twins at first birth, twins at second birth, and testing positive for species A.

N <- 100000

dfa <- tibble(

species = 'A',

t1 = rbinom(N, 1, 0.1),

t2 = rbinom(N, 1, 0.1),

pa = rbinom(N, 1, 0.8)

)

dfb <- tibble(

species = 'B',

t1 = rbinom(N, 1, 0.2),

t2 = rbinom(N, 1, 0.2),

pa = rbinom(N, 1, 1 - 0.65)

)

df <- dfa %>% bind\_rows(dfb)

All of the problems can now be solved by simply filtering out any events not consisent with our observations, then summarising the remaining events.

h1 <- df %>%

filter(t1 == 1) %>%

summarise(mean(t2 == 1)) %>%

pull()

h2 <- df %>%

filter(t1 == 1) %>%

summarise(mean(species == 'A')) %>%

pull()

h3 <- df %>%

filter(t1 == 1, t2 == 0) %>%

summarise(mean(species == 'A')) %>%

pull()

h4a <- df %>%

filter(pa == 1) %>%

summarise(mean(species == 'A')) %>%

pull()

h4b <- df %>%

filter(pa == 1, t1 == 1, t2 == 0) %>%

summarise(mean(species == 'A')) %>%

pull()

| Solutions | | |
| --- | --- | --- |
| **exercise** | **bayes** | **counting** |
| h1 | 0.1666667 | 0.1669936 |
| h2 | 0.3333333 | 0.3360484 |
| h3 | 0.3529412 | 0.3635856 |
| h4a | 0.6956522 | 0.6963991 |
| h4b | 0.5443787 | 0.5656231 |

For H1 we expect the probability to be between 0.1 and 0.2, since those are the two possible birth rates. Also, since we observed a twin birth already, it makes sense that it is closer to 0.2 since species B is more likely to birth twins. In other words, in H2 we expect the species to be less likely to be species A. Birthing a singleton infant is fairly common, so we wouldn’t expect this observation to change our inference very much in H3.

**Bayes rule**

Let’s also work out the solutions analytically using Bayes rule. Let’s start with H2 since it’s useful for calculating H1.

\[  
\begin{align}  
\mathbb P(A \mid T\_1)  
&=  
\frac{\mathbb P(T\_1 \mid A) \mathbb P(A)}{\mathbb P(T\_1)}  
\\  
&=  
\frac{\mathbb P(T\_1 \mid A) \mathbb P(A)}{\mathbb P(T\_1 \mid A) \mathbb P(A) + \mathbb P(T\_1 \mid B) \mathbb P(B)}  
\\  
&=  
\frac{0.1 \cdot 0.5}{0.1 \cdot 0.5 + 0.2 \cdot 0.5}  
\\  
&=  
\frac{0.05}{0.05 + 0.1}  
\\  
&=  
\frac{1}{3}  
\end{align}  
\]

Now we can use our solution to H2 and plug it into the appropriate place in the formula for H1. Note that \(\mathbb P(T\_2 \mid A)\) is the same as \(\mathbb P(T\_1 \mid A)\) by the assumptions of the problem. Similarily, once we know the species, whether the first birth was twins is irrelevant to the probability of twins in the second birth, i.e. \(\mathbb P(T\_2 \mid T\_1, A) = \mathbb P(T\_2 \mid A)\).

\[  
\begin{align}  
\mathbb P(T\_2 \mid T\_1)  
&=  
\mathbb P(T\_2 \mid T\_1, A) \mathbb P(A \mid T\_1)  
+  
\mathbb P(T\_2 \mid T\_1, B) \mathbb P(B \mid T\_1)  
\\  
&=  
\mathbb P(T\_2 \mid A) \mathbb P(A \mid T\_1)  
+  
\mathbb P(T\_2 \mid B) \mathbb P(B \mid T\_1)  
\\  
&=  
\frac{1}{10} \cdot \frac{1}{3} + \frac{2}{10} \cdot \frac{2}{3}  
\\  
&=  
\frac{5}{30}  
\\  
&=  
\frac{1}{6}  
\end{align}  
\]

For H3, let’s use the notation \(-T\_i\) to mean singleton infants (i.e. not twins).

\[  
\begin{align}  
\mathbb P(A \mid T\_1, – T\_2)  
&=  
\frac{\mathbb P(- T\_2 \mid T\_1, A) \mathbb P(A \mid T\_1)}{\mathbb P(- T\_2 \mid T\_1)}  
\\  
&=  
\frac{\mathbb P(- T\_2 \mid A) \mathbb P(A \mid T\_1)}{\mathbb P(- T\_2 \mid T\_1)}  
\\  
&=  
\frac{(1 – 0.1) \cdot \frac{1}{3}}{1 – 0.15}  
\\  
&=\frac{0.3}{0.85}  
\\  
&=  
\frac{6}{17}  
\end{align}  
\]

This is about 0.353.

Now for H4a.

\[  
\begin{align}  
\mathbb P(A \mid P\_A)  
&=  
\frac{\mathbb P(P\_A \mid A) \mathbb P(A)}{\mathbb P(P\_A)}  
\\  
&=  
\frac{\mathbb P(P\_A \mid A) \mathbb P(A)}{\mathbb P(P\_A \mid A) \mathbb P(A) + \mathbb P(P\_A \mid B) \mathbb P(B)}  
\\  
&=  
\frac{0.8 \cdot 0.5 }{0.8 \cdot 0.5 + 0.35 \cdot 0.5}  
\\  
&=  
\frac{0.4 }{0.4 + 0.175}  
\\  
&=  
\frac{0.4 }{0.575}  
\end{align}  
\]

This is about 0.696.

Finally H4b.

\[  
\begin{align}  
\mathbb P(A \mid P\_A, T\_1, -T\_2)  
&=  
\frac{\mathbb P(P\_A \mid A, T\_1, -T\_2) \mathbb P(A \mid T\_1, -T\_2)}{\mathbb P(P\_A \mid T\_1, -T\_2)}  
\\  
&=  
\frac{\mathbb P(P\_A \mid A) \mathbb P(A \mid T\_1, -T\_2)}{\mathbb P(P\_A \mid A) \mathbb P(A \mid T\_1, -T\_2) + \mathbb P(P\_A \mid B) \mathbb P(B \mid T\_1, -T\_2)}  
\\  
&=  
\frac{\frac{4}{5} \cdot \frac{6}{17} }{\frac{4}{5}\cdot \frac{6}{17} + \frac{7}{20} \cdot \frac{11}{17}}  
\\  
&=  
\frac{\frac{24}{85} }{\frac{24}{85} + \frac{77}{340}}  
\\  
&=  
\frac{\frac{24}{85} }{\frac{92 + 77}{340}}  
\\  
&=  
\frac{24}{85} \cdot \frac{340}{169}  
\\  
&=  
\frac{92}{169}  
\end{align}  
\]

This is about 0.544.

Please enable JavaScript to view the [comments powered by Disqus.](https://disqus.com/?ref_noscript)