Applied Bayesian Data Analysis — Exercise 3 A

Kim Albertsson

CERN and Luleå University of Technology

kim.albertsson@ltu.se

October 14, 2019

Question 1

What is $p(\mathrm{E} = \mathrm{Blue} | \mathrm{H} \in \{\mathrm{Red}, \mathrm{Blond}\})$?

Q: What is $p(\mathrm{E} = \mathrm{Blue} | \mathrm{H} \in \{\mathrm{Red}, \mathrm{Blond}\})$?

$$ \tag{1} $$

Joint probabilities

	Black	Brown	Red	Blond	$\sum$
Brown	0.11	0.20	0.04	0.01	0.37
Blue	0.03	0.14	0.03	0.16	0.36
Hazel	0.03	0.09	0.02	0.02	0.16
Green	0.01	0.05	0.02	0.03	0.11
$\sum$	0.18	0.48	0.12	0.21	1.0

Q: What is $p(\mathrm{E} = \mathrm{Blue} | \mathrm{H} \in \{\mathrm{Red}, \mathrm{Blond}\})$?

$$p(\mathrm{E} = \mathrm{Blue} | \mathrm{H} \in \{\mathrm{Red}, \mathrm{Blond}\}) = \frac{p(\mathrm{H} \in \{\mathrm{Red}, \mathrm{Blond}\} | \mathrm{E} = \mathrm{Blue}) p(\mathrm{E} = \mathrm{Blue})}{p(\mathrm{H} \in \{\mathrm{Red}, \mathrm{Blond}\})}$$

Joint probabilities

	Black	Brown	Red	Blond	$\sum$
Brown	0.11	0.20	0.04	0.01	0.37
Blue	0.03	0.14	0.03	0.16	0.36
Hazel	0.03	0.09	0.02	0.02	0.16
Green	0.01	0.05	0.02	0.03	0.11
$\sum$	0.18	0.48	0.12	0.21	1.0

Q: What is $p(\mathrm{E} = \mathrm{Blue} | \mathrm{H} \in \{\mathrm{Red}, \mathrm{Blond}\})$?

$$p(\mathrm{E} = \mathrm{Blue} | \mathrm{H} \in \{\mathrm{Red}, \mathrm{Blond}\}) =$$
$$\frac{p(\mathrm{H} \in \{\mathrm{Red}, \mathrm{Blond}\} | \mathrm{E} = \mathrm{Blue}) p(\mathrm{E} = \mathrm{Blue})}{p(\mathrm{H} \in \{\mathrm{Red}, \mathrm{Blond}\})} =$$
$$\frac{0.03 + 0.16}{0.12 + 0.21} =$$
$$0.58$$

Joint probabilities

	Black	Brown	Red	Blond	$\sum$
Brown	0.11	0.20	0.04	0.01	0.37
Blue	0.03	0.14	0.03	0.16	0.36
Hazel	0.03	0.09	0.02	0.02	0.16
Green	0.01	0.05	0.02	0.03	0.11
$\sum$	0.18	0.48	0.12	0.21	1.0

Question 2

Create a function to calculate $p(\mathrm{sick}|\,T)$ where T is any sequence drawn from $\{+,-\}$. Explain how the function operates.

The setting supposes the existence of a disease with prevalence 0.001 in a population ($p(\mathrm{sick}) = 0.001$). It also supposes the existence of a test for said disease with sensitivity 0.99 ($p(+|\mathrm{sick}) = 0.99$) and specificity 0.95 ($p(-|\mathrm{sick}) = 0.05$).

$$ \tag{2} $$

```python
p = {}

# Conditionals
p['+|:('] = 0.99 # True positive
p['+|:)'] = 0.05 # False positive



# Marginals
p[':('] = 0.001 # p sick
```

```python
p = {}

# Conditionals
p['+|:('] = 0.99 # True positive
p['+|:)'] = 0.05 # False positive
p['-|:('] = 1 - p['+|:(']
p['-|:)'] = 1 - p['+|:)']

# Marginals
p[':('] = 0.001 # p sick
p[':)'] = 1 - p[':('] # p healthy
p['+'] = p['+|:(']*p[':('] + p['+|:)']*p[':)']
p['-'] = 1 - p['+']
```

$$
\begin{aligned}
p(h|t_0 t_1 \ldots) &= p(h|t_0, t_1, \ldots) \\
&= \frac{p(t_0, t_1, \ldots|h)}{p(t_0, t_1, \ldots)} p(h) \\
&\text{assume independence} \\
&= \frac{p(t_0|h)p(t_1|h)p(\ldots|h)}{p(t_0)p(t_1)p(\ldots)} p(h) \\
&= \ldots \cdot \frac{p(t_1|h)}{p(t_1)} \cdot \frac{p(t_0|h)}{p(t_0)} \cdot p(h)
\end{aligned}
$$

```python
def proba_is_sick(test_scores):
    allowed_symbols = '-+'

    posterior = p[':(']
    for symbol in test_scores:
        assert(symbol in allowed_symbols)
        prior = posterior
        likelihood = p[f'{symbol}|:(']
        evidence = p[f'{symbol}']
        posterior = likelihood / evidence * prior
    return posterior
```

```
>>> proba_is_sick('')
0.001
>>> proba_is_sick('+')
0.019434628975265017
>>> proba_is_sick('+-')
0.00020477766395449 22
```