Bayes’ Theorem
Remember Language ID?

- Let $p(X) =$ probability of text $X$ in English
- Let $q(X) =$ probability of text $X$ in Polish
- Which probability is higher?

- (we’d also like bias toward English since it’s more likely \textit{a priori} – ignore that for now)

"Horses and Lukasiewicz are on the curriculum."

\[ p(x_1=h, x_2=o, x_3=r, x_4=s, x_5=e, x_6=s, \ldots) \]
Bayes’ Theorem

- \( p(A \mid B) = p(B \mid A) \times p(A) / p(B) \)

- Easy to check by removing syntactic sugar
- **Use 1**: Converts \( p(B \mid A) \) to \( p(A \mid B) \)
- **Use 2**: Updates \( p(A) \) to \( p(A \mid B) \)

- Stare at it so you’ll recognize it later
Language ID

- Given a sentence $x$, I suggested comparing its prob in different languages:
  - $p(\text{SENT}=x \mid \text{LANG}=\text{english})$ (i.e., $p_{\text{english}}(\text{SENT}=x)$)
  - $p(\text{SENT}=x \mid \text{LANG}=\text{polish})$ (i.e., $p_{\text{polish}}(\text{SENT}=x)$)
  - $p(\text{SENT}=x \mid \text{LANG}=\text{xhosa})$ (i.e., $p_{\text{xhosa}}(\text{SENT}=x)$)

- But surely for language ID we should compare
  - $p(\text{LANG}=\text{english} \mid \text{SENT}=x)$
  - $p(\text{LANG}=\text{polish} \mid \text{SENT}=x)$
  - $p(\text{LANG}=\text{xhosa} \mid \text{SENT}=x)$
Language ID

- For language ID we should compare
  - $p(\text{LANG}=\text{english} \mid \text{SENT}=x)$
  - $p(\text{LANG}=\text{polish} \mid \text{SENT}=x)$
  - $p(\text{LANG}=\text{xhosa} \mid \text{SENT}=x)$

- For ease, multiply by $p(\text{SENT}=x)$ and compare
  - $p(\text{LANG}=\text{english}, \text{SENT}=x)$
  - $p(\text{LANG}=\text{polish}, \text{SENT}=x)$
  - $p(\text{LANG}=\text{xhosa}, \text{SENT}=x)$

- Must know prior probabilities; then rewrite as
  - $p(\text{LANG}=\text{english}) \times p(\text{SENT}=x \mid \text{LANG}=\text{english})$
  - $p(\text{LANG}=\text{polish}) \times p(\text{SENT}=x \mid \text{LANG}=\text{polish})$
  - $p(\text{LANG}=\text{xhosa}) \times p(\text{SENT}=x \mid \text{LANG}=\text{xhosa})$

- * likelihood (what we had before)
  - a posteriori

- sum of these is a way to find $p(\text{SENT}=x)$; can divide back by that to get posterior probs

- a priori
Let’s try it!

```
best 0.7  p(LANG=english) * p(SENT=x | LANG=english)  0.000001
0.2  p(LANG=polish)  * p(SENT=x | LANG=polish)   0.000004
0.1  p(LANG=xhosa)  * p(SENT=x | LANG=xhosa)   0.000005 best

Prior prob

from a very simple model: a single die whose sides are the languages of the world

= p(LANG=english, SENT=x)  0.000007
= p(LANG=polish, SENT=x)  0.000008 best compromise
= p(LANG=xhosa, SENT=x)  0.000005

Joint probability

= 0.000005

Probability of evidence

= 0.00020  total over all ways of getting SENT=x

```

“First we pick a random LANG, then we roll a random SENT with the LANG dice.”

from a set of trigram dice (actually 3 sets, one per language)
Let’s try it!

"First we pick a random LANG, then we roll a random SENT with the LANG dice."

\[
\begin{align*}
\text{joint probability} &= p(\text{LANG}=\text{english}, \text{SENT}=x) \\
&= p(\text{LANG}=\text{polish}, \text{SENT}=x) \\
&= p(\text{LANG}=\text{xhosa}, \text{SENT}=x) \\
\text{probability of evidence} &= p(\text{SENT}=x) \\
&= 0.000007 \\
&= 0.000008 \\
&= 0.000005 \\
\text{add up} &
\begin{align*}
p(\text{LANG}=\text{english} \mid \text{SENT}=x) &= \frac{0.000007}{0.000020} = \frac{7}{20} \\
p(\text{LANG}=\text{polish} \mid \text{SENT}=x) &= \frac{0.000008}{0.000020} = \frac{8}{20} \quad \text{best} \\
p(\text{LANG}=\text{xhosa} \mid \text{SENT}=x) &= \frac{0.000005}{0.000020} = \frac{5}{20} \\
\end{align*}
\end{align*}
\]

given the evidence \(\text{SENT}=x\), the possible languages sum to 1

(normalize by a constant so they'll sum to 1)
**General Case (“noisy channel”)**

```
language → text
```
```
text → speech
```
```
spelled → misspelled
```
```
English → French
```

```
“noisy channel”
```
```
 mess up a into b
```
```
p(A=a)
```
```
p(B=b | A=a)
```

```
“decoder”
```
```
most likely reconstruction of a
```
```
maximize p(A=a | B=b)
```
```
= p(A=a) p(B=b | A=a) / (B=b)
```
```
= p(A=a) p(B=b | A=a)
```
```
/ Σa', p(A=a’) p(B=b | A=a’)
```

Language ID

- For language ID we should compare
  - $p(\text{LANG}=\text{english} \mid \text{SENT}=x)$
  - $p(\text{LANG}=\text{polish} \mid \text{SENT}=x)$
  - $p(\text{LANG}=\text{xhosa} \mid \text{SENT}=x)$

- For ease, multiply by $p(\text{SENT}=x)$ and compare
  - $p(\text{LANG}=\text{english}, \text{SENT}=x)$
  - $p(\text{LANG}=\text{polish}, \text{SENT}=x)$
  - $p(\text{LANG}=\text{xhosa}, \text{SENT}=x)$

- which we find as follows (we need prior probs!):
  - $p(\text{LANG}=\text{english})$ * $p(\text{SENT}=x \mid \text{LANG}=\text{english})$
  - $p(\text{LANG}=\text{polish})$ * $p(\text{SENT}=x \mid \text{LANG}=\text{polish})$
  - $p(\text{LANG}=\text{xhosa})$ * $p(\text{SENT}=x \mid \text{LANG}=\text{xhosa})$

- a posteriori
- a priori
- likelihood
General Case ("noisy channel")

- Want most likely A to have generated evidence B
  
  - \( p(A = a_1 | B = b) \)
  
  - \( p(A = a_2 | B = b) \)
  
  - \( p(A = a_3 | B = b) \)

- For ease, multiply by \( p(B=b) \) and compare
  
  - \( p(A = a_1, B = b) \)
  
  - \( p(A = a_2, B = b) \)
  
  - \( p(A = a_3, B = b) \)

- which we find as follows (we need prior probs!):
  
  - \( p(A = a_1) \) * \( p(B = b | A = a_1) \)
  
  - \( p(A = a_2) \) * \( p(B = b | A = a_2) \)
  
  - \( p(A = a_3) \) * \( p(B = b | A = a_3) \)

\( a \ posteriori \)

\( a \ priori \)

\( likelihood \)
Speech Recognition

- For baby speech recognition we should compare
  - $p(\text{MEANING}=\text{gimme} \mid \text{SOUND}=\text{uhh})$
  - $p(\text{MEANING}=\text{changeme} \mid \text{SOUND}=\text{uhh})$
  - $p(\text{MEANING}=\text{loveme} \mid \text{SOUND}=\text{uhh})$

- For ease, multiply by $p(\text{SOUND}=\text{uhh})$ & compare
  - $p(\text{MEANING}=\text{gimme}, \text{SOUND}=\text{uhh})$
  - $p(\text{MEANING}=\text{changeme}, \text{SOUND}=\text{uhh})$
  - $p(\text{MEANING}=\text{loveme}, \text{SOUND}=\text{uhh})$

- which we find as follows (we need prior probs!):
  - $p(\text{MEAN}=\text{gimme}) \ast p(\text{SOUND}=\text{uhh} \mid \text{MEAN}=\text{gimme})$
  - $p(\text{MEAN}=\text{changeme}) \ast p(\text{SOUND}=\text{uhh} \mid \text{MEAN}=\text{changeme})$
  - $p(\text{MEAN}=\text{loveme}) \ast p(\text{SOUND}=\text{uhh} \mid \text{MEAN}=\text{loveme})$

\[a \text{ priori}\]
\[a \text{ posteriori}\]

likelihood
A simpler view? Odds Ratios

- What A values are probable, given that B=b?
- Bayes’ Theorem says:
  - $p(A=a_1 \mid B=b) = p(A=a_1) \cdot p(B=b \mid A=a_1) / p(B=b)$
  - $p(A=a_2 \mid B=b) = p(A=a_2) \cdot p(B=b \mid A=a_2) / p(B=b)$
- Therefore
  $$\frac{p(A=a_1 \mid B=b)}{p(A=a_2 \mid B=b)} = \frac{p(A=a_1)}{p(A=a_2)} \cdot \frac{p(B=b \mid A=a_1)}{p(B=b \mid A=a_2)}$$
  
  posterior odds ratio = prior odds ratio * likelihood ratio
A simpler view? Odds Ratios

\[
\frac{p(A=a_1 | B=b)}{p(A=a_2 | B=b)} = \frac{\frac{p(A=a_1)}{p(A=a_2)}}{\frac{p(B=b | A=a_1)}{p(B=b | A=a_2)}}
\]

- **posterior odds ratio**
- **prior odds ratio**
- **likelihood odds ratio**

| \(p(\text{SENT}=x | \text{LANG} = \text{english})\) | \(0.7\) | \(p(\text{LANG} = \text{english})\) | \(0.00001\) |
|---------------------------------|-------|-----------------|---------|
| \(p(\text{SENT}=x | \text{LANG} = \text{polish})\) | \(0.2\) | \(p(\text{LANG} = \text{polish})\) | \(0.00004\) |
| \(p(\text{SENT}=x | \text{LANG} = \text{xhosa})\) | \(0.1\) | \(p(\text{LANG} = \text{xhosa})\) | \(0.00005\) |

- A **priori**, English is 7 times as probable as Xhosa (7:1 odds)
- But **likelihood** of English is only 1/5 as large (1:5 odds)
- So a **posteriori**, English now \(7 \times \frac{1}{5} = 1.2\) times as probable (7:5 odds)
  - That is: \(p(\text{English}) = 7/12\), \(p(\text{Xhosa}) = 5/12\) if no other options
Life or Death!

- \( p(\text{hoof}) = 0.001 \) so \( p(\neg \text{hoof}) = 0.999 \)

- \( p(\text{positive test} \mid \neg \text{hoof}) = 0.05 \) “false pos”
- \( p(\text{negative test} \mid \text{hoof}) = \varepsilon \geq 0 \) “false neg”
  
  So \( p(\text{positive test} \mid \text{hoof}) = 1 - \varepsilon \)

- What is \( p(\text{hoof} \mid \text{positive test}) \)?
  - Consider the hoof: \( \neg \text{hoof} \) odds ratio
  - Prior odds ratio \( 1:999 \) (improbable!)
  - Likelihood ratio at most \( 1:0.05 \), or equivalently \( 20:1 \)
  - So posterior odds ratio at most \( 20:999 \), or about \( 1:50 \)
    - That is, \( p(\text{hoof} \mid \text{positive test}) \) at most about \( 1/51 \)

Does Epitaph have hoof-and-mouth disease? He tested positive – oh no! False positive rate only 5%