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)

\[ p(x_1=h, x_2=o, x_3=r, x_4=s, x_5=e, x_6=s, \ldots) \]
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
  - \( p(\text{LANG}=\text{english}) \ast p(\text{SENT}=x \mid \text{LANG}=\text{english}) \)
  - \( p(\text{LANG}=\text{polish}) \ast p(\text{SENT}=x \mid \text{LANG}=\text{polish}) \)
  - \( p(\text{LANG}=\text{xhosa}) \ast p(\text{SENT}=x \mid \text{LANG}=\text{xhosa}) \)

*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

likelihood (what we had before)
Let’s try it!

<table>
<thead>
<tr>
<th>Prior Prob</th>
<th>Likelihood</th>
<th>Joint Probability</th>
<th>Probability of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.7</td>
<td>0.00001</td>
<td>0.000007</td>
<td>0.000007</td>
</tr>
<tr>
<td>0.2</td>
<td>0.00004</td>
<td>0.000008</td>
<td>0.000008</td>
</tr>
<tr>
<td>0.1</td>
<td>0.00005</td>
<td>0.000005</td>
<td>0.000005</td>
</tr>
</tbody>
</table>

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

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

From a set of trigram dice (actually 3 sets, one per language).

Best compromise: 0.000008

Total over all ways of getting SENT=x: 0.000020
Let’s try it!

Joint probability

\[ p(\text{LANG}=\text{english}, \text{SENT}=x) = 0.000007 \]
\[ p(\text{LANG}=\text{polish}, \text{SENT}=x) = 0.000008 \quad \text{best compromise} \]
\[ p(\text{LANG}=\text{xhosa}, \text{SENT}=x) = 0.000005 \]

Total probability of getting SENT=x one way or another!

\[ p(\text{SENT}=x) = 0.000020 \]

Posterior probability

\[ p(\text{LANG}=\text{english} \mid \text{SENT}=x) = \frac{0.000007}{0.000020} = \frac{7}{20} \]
\[ p(\text{LANG}=\text{polish} \mid \text{SENT}=x) = \frac{0.000008}{0.000020} = \frac{8}{20} \quad \text{best} \]
\[ p(\text{LANG}=\text{xhosa} \mid \text{SENT}=x) = \frac{0.000005}{0.000020} = \frac{5}{20} \]

Given the evidence SENT=x, the possible languages sum to 1.

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

…
Let’s try it!

$$p(L\text{ANG}=\text{english}, \text{SENT}=x) = p(L\text{ANG}=\text{polish}, \text{SENT}=x) = p(L\text{ANG}=\text{xhosa}, \text{SENT}=x) = 0.000007$$

$$p(L\text{ANG}=\text{polish}, \text{SENT}=x) = 0.000008$$

$$p(L\text{ANG}=\text{xhosa}, \text{SENT}=x) = 0.000005$$

$$p(\text{SENT}=x) = 0.000020$$

**best compromise**

**total over all ways of getting x**
General Case ("noisy channel")

In a general case ("noisy channel"), we have:

- **Input**: $a$, $p(A=a)$
- **Messing up**: $a$ into $b$, $p(B=b | A=a)$
- **Decoder**: "decoder"
- **Output**: $b$
- **Most likely reconstruction**: $p(A=a | B=b)$

Language $\rightarrow$ text

text $\rightarrow$ speech

Spelled $\rightarrow$ misspelled

English $\rightarrow$ French

**Mathematical Formulation**:

$$p(A=a | B=b) = \frac{p(A=a) \cdot p(B=b | A=a)}{\sum_{a'} p(A=a') \cdot p(B=b | A=a')}$$
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
  - \( p(A = a_1 | B = b) \)
  - \( p(A = a_2 | B = b) \)
  - \( p(A = a_3 | B = b) \)

  *a posteriori*

- 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!):
  - \( p(A = a_1) \)
  - \( p(A = a_2) \)
  - \( p(A = a_3) \)
  *\( p(B = b | A = a_1) \)
  *\( p(B = b | A = a_2) \)
  *\( p(B = b | A = a_3) \)

*a priori*  \[ likelihood \]
Speech Recognition

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

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

- which we find as follows (we need prior probs!):
  - $p(\text{MEAN}=\text{gimme}) \times p(\text{SOUND}=\text{uhh} \mid \text{MEAN}=\text{gimme})$
  - $p(\text{MEAN}=\text{changeme}) \times p(\text{SOUND}=\text{uhh} \mid \text{MEAN}=\text{changeme})$
  - $p(\text{MEAN}=\text{loveme}) \times p(\text{SOUND}=\text{uhh} \mid \text{MEAN}=\text{loveme})$
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”
- \( p(\text{negative test} \mid \text{hoof}) = x \approx 0 \) “false neg”
  so \( p(\text{positive test} \mid \text{hoof}) = 1-x \approx 1 \)

- What is \( p(\text{hoof} \mid \text{positive test})? \)
  - don’t panic - still very small! < 1/51 for any \( x \)

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