Bayes' Theorem

Rev. Thomas Bayes
1702-1761
Conditional Probability

\[ \Pr(B \mid A) = \frac{\Pr(B,A)}{\Pr(A)} \]
\[ \Pr(A \mid B) = \frac{\Pr(A,B)}{\Pr(B)} \]
Conditional Probability

\[ \Pr(B \mid A) = \frac{\Pr(B,A)}{\Pr(A)} \]
\[ \Pr(A \mid B) = \frac{\Pr(A,B)}{\Pr(B)} \]

\[ \Pr(A,B) = \Pr(B \mid A) \cdot \Pr(A) \]
\[ \Pr(B,A) = \Pr(A \mid B) \cdot \Pr(B) \]
Conditional Probability

\[ Pr(B \mid A) = \frac{Pr(B,A)}{Pr(A)} \]
\[ Pr(A \mid B) = \frac{Pr(A,B)}{Pr(B)} \]

\[ Pr(A,B) = Pr(B \mid A) \cdot Pr(A) \]
\[ Pr(B,A) = Pr(A \mid B) \cdot Pr(B) \]

\[ Pr(A \mid B) \cdot Pr(B) = Pr(B \mid A) \cdot Pr(A) \]
Conditional Probability

\[
\begin{align*}
\Pr(B \mid A) &= \frac{\Pr(B,A)}{\Pr(A)} \\
\Pr(A \mid B) &= \frac{\Pr(A,B)}{\Pr(B)} \\
\Pr(A,B) &= \Pr(B \mid A) \cdot \Pr(A) \\
\Pr(B,A) &= \Pr(A \mid B) \cdot \Pr(B) \\
\Pr(A \mid B) \cdot \Pr(B) &= \Pr(B \mid A) \cdot \Pr(A) \\
\Pr(A \mid B) &= \frac{\Pr(B \mid A) \cdot \Pr(A)}{\Pr(B)}
\end{align*}
\]

BAYES RULE
False Positives

- If a patient has a disease the test returns a positive 99% of the time
- If a patient does not have the disease, the test returns positive 5% of the time
- 0.1% of the population has the disease
- What is the probability that someone who tested positive has the disease?
Suppose A = has disease
B = tested positive

\[
P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\neg A)P(\neg A)}
\]

\[
P(A|B) = \frac{0.99 \cdot 0.001}{0.99 \cdot 0.001 + 0.05 \cdot 0.999} 
\approx 0.019
\]
What is the probability that the sun exploded?
Bayes' Theorem

\[ P(\theta|y) = \frac{\int_{-\infty}^{\infty} P(y|\theta) P(\theta) \, d\theta}{P(y)} = \frac{P(y|\theta) P(\theta)}{P(y)} \]
$P(\theta) = Unif$

$P(\theta|y) = ??$
What is $P(y \mid \theta)$?

$L = P(y \mid \theta) = \text{Binom}(y \mid n, \theta)$
\[ P(\theta|y) = \frac{\text{Binom}(y|n, \theta) \text{Unif}(\theta|0,1)}{\int_0^1 \text{Binom}(y|n, \theta) \text{Unif}(\theta|0,1) \, d\theta} \]

\[ P(\theta|y) = \frac{\binom{n}{y} \theta^y (1-\theta)^{n-y} \cdot 1}{\int_0^1 \binom{n}{y} \theta^y (1-\theta)^{n-y} \cdot 1 \, d\theta} \]
\[ P(\theta|y) = \frac{\theta^y (1-\theta)^{n-y}}{\int_0^1 \theta^y (1-\theta)^{n-y}} \]

Mean? Variance? Confidence Interval?

\[ Beta(x|\alpha, \beta) = \frac{x^{\alpha-1} (1-x)^{\beta-1}}{\int_0^1 x^{\alpha-1} (1-x)^{\beta-1}} \]

\[ P(\theta|y) = Beta(x|y+1, n-y+1) \]
• Posterior is a PDF
• $\theta$ is a random variable
• Interested in full distribution
Priors

- Makes it possible to calculate a posterior density of the model parameter rather than the likelihood of the data
- Provides a way of incorporating information that is external to the data set(s) at hand
- Inherently sequential

Previous Posterior = New Prior
Data updates the prior

\[ n=0 \]
Where do Priors come from?

- Uninformative / vague
  - Chosen to have minimal information content, allows the likelihood to dominate the analysis

- Previous analyses
  - Must be equivalent
  - Variance inflation

- “The literature”
  - Meta-analysis

- Expert knowledge
Where do Priors come from?

- Uninformative / vague
  - Chosen to have minimal information content, allows the likelihood to dominate the analysis.

- Previous analyses
  - Must be equivalent
  - Variance inflation

- "The literature"
  - Meta-analysis

- Expert knowledge

Prior specification must be “blind” to the data in the analysis!!

No “double dipping” -- leads to falsely overconfident results
How do I choose a prior PDF?

• Analogous to how we choose the data model
  – Range restrictions, shape, etc.

• Conjugacy
  – A prior is conjugate to the likelihood if the posterior PDF is in the same family as the prior
  – Allow for closed-form analytical solutions to either full posterior or (in multiparameter models) for the conditional distribution of that parameter.
  – Modern computational methods no longer require conjugacy
Example: Tree mortality rate

- Data: observed $n=4$ trees, $y=1$ died this year

\[
L = P(y | \theta) = \text{Binom}(y | \theta, n)
\]

- Prior: last year observed $n_0 = 2$ trees, $y_0 = 1$ died

\[
Prior = P(\theta) = \text{Beta}(\theta | y_0, n_0 - y_0)
\]
\[ P(\theta|y) \propto \text{Binom}(y|\theta, n) \text{Beta}(\theta|y_o, n_0-y_0) \]

\[ P(\theta|y) \propto \binom{n}{y} \theta^y (1-\theta)^{n-y} \times \frac{\theta^{y_o-1} (1-\theta)^{n_0-y_0-1}}{B(y_o, n_0-y_0)} \]

\[ P(\theta|y) \propto \theta^{y+y_o-1} (1-\theta)^{n-y+n_0-y_0-1} \]

\[ P(\theta|y) = \text{Beta}(\theta|y+y_0, n+n_0-y-y_0) \]

Beta-Binomial Model
\[ P(\theta|y) = \text{Beta}(\theta|y + y_0, n + n_0 - y - y_0) = \text{Beta}(\theta|2,4) \]
How much impact does the prior have on the analysis?