Bayesian Data Analysis

• Supporting document (Masson, 2011)

• Null hypothesis significance testing provides the probability of the observed outcome (or one that is even more extreme): $p(\text{Data } | \text{ H}_0)$

• There are serious concerns emerging about the stability and interpretation of $p$ values

• Dance of the $p$ values
Bayesian Data Analysis

- $p$ values are unstable across replications of a research study

- When the null hypothesis is true, what would the distribution of $p$ values across replications of a study look like?
Bayesian Data Analysis

• Interpretation of \( p \) values
  
  • natural tendency to interpret \( p \) value as an indication of how likely it is that the null hypothesis is true

• Null hypothesis significance testing provides the probability of the observed outcome (or one that is even more extreme), given that \( H_0 \) is true: \( p(\text{Data} \mid H_0) \)

• Notice that \( p(\text{Data} \mid H_0) \neq p(H_0 \mid \text{Data}) \)

What we get

What we want
Bayesian Data Analysis

• From $p(\text{Data} \mid \text{H}_0)$, we try to infer the status of $\text{H}_0$
  • we accept $\text{H}_1$ if the data disfavor $\text{H}_0$
  • or we fail to reject $\text{H}_0$, but then what?

• The plague of null effects: under null hypothesis testing, the null hypothesis cannot be accepted
  • even when data favoring the null hypothesis constitute a theoretically interesting outcome, hypothesis testing logic does not allow researchers to make effective use of such a result
Bayesian Data Analysis

• A **Bayesian** approach helps move us closer to what we want

• Bayes' theorem

\[
p(H \mid D) = \frac{p(D \mid H) \cdot p(H)}{p(D)}
\]

• Let's see how this works

  • medical testing scenario: a test for cancer
Bayesian Data Analysis

• For every 10,000 people in a population, suppose the following facts hold for the rate of cancer of some type and a medical test for that type of cancer:

<table>
<thead>
<tr>
<th>Test result</th>
<th>No cancer</th>
<th>Cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>990</td>
<td>80</td>
</tr>
<tr>
<td>Negative</td>
<td>8910</td>
<td>20</td>
</tr>
</tbody>
</table>

\[
p(C \mid Pos) = \frac{p(Pos \mid C) \cdot p(C)}{p(Pos)} = \frac{(0.8) \cdot (0.01)}{0.107} = 0.075
\]
Bayesian Data Analysis

• Rather than an emphasis on rejecting $H_0$, a Bayesian model selection method can be used

• $H_0$ and $H_1$ are characterized as opposing models of the data

• Bayesian approach evaluates the extent to which the data support the null model vs. the alternative model

• the result is a quantitative estimate of how much better one model rather than the other fits the data

• no cutoff (e.g., $p < .05$) for rejecting an hypothesis
Bayesian Data Analysis

- Bayes' theorem

\[ p(H \mid D) = \frac{p(D \mid H) \cdot p(H)}{p(D)} \]

posterior probability of Hypothesis given Data
Bayesian Data Analysis

• Bayes' theorem

\[ p(H \mid D) = \frac{p(D \mid H) \cdot p(H)}{p(D)} \]

• Define \textit{relative} posterior probabilities of null and alternative hypotheses with this formulation (odds)

\[
\frac{p(H_0 \mid D)}{p(H_1 \mid D)} = \frac{\frac{p(D \mid H_0) \cdot p(H_0)}{p(D)}}{\frac{p(D \mid H_1) \cdot p(H_1)}{p(D)}}
\]

\[
\frac{p(H_0 \mid D)}{p(H_1 \mid D)} = \frac{p(D \mid H_0) \cdot p(H_0)}{p(D \mid H_1) \cdot p(H_1)}
\]
Bayesian Data Analysis

\[
\frac{p(H_0 \mid D)}{p(H_1 \mid D)} = \frac{p(D \mid H_0)}{p(D \mid H_1)} \cdot \frac{p(H_0)}{p(H_1)}
\]

posterior odds  Bayes factor  prior odds

• Bayes factor reflects change in prior odds based on new data

• strength of evidence for \(H_0\) relative to \(H_1\)

• if we assume equal prior odds \([p(H_0) = p(H_1)]\), then posterior odds are equal to the Bayes factor
• Example of computing Bayes Factor
  • Madam X predicting coin flip outcomes

  $\text{H}_0: \ p_r(\text{correct}) = .5$

  $\text{H}_1: \ p_r(\text{correct}) = .7$

  $BF_{01} = \frac{p(H_0 | D)}{p(H_1 | D)} = \frac{p(D | H_0) \cdot p(H_0)}{p(D | H_1) \cdot p(H_1)}$

  $= \frac{.17620 \cdot .5}{.03082 \cdot .5} = 5.717$

  $p(H_0 | D) = \frac{BF_{01}}{1 + BF_{01}} \quad p(H_1 | D) = 1 - p(H_0 | D)$

  $= \frac{5.717}{1 + 5.717} = .851$

  $= 1 - .851 = .149$
Bayesian Data Analysis

• But wait… maybe Madam X’s powers are not necessarily equivalent to $p_r = .7$
• what about .6 or even .9?
• consider this version of $H_1$: $p_r = .6$ or .7 or .8 or .9
• how likely is each of these values to be correct?
  • suppose they are equally likely
Bayesian Data Analysis

• With this more complex $H_1$, $p(D|H_1)$ needs to be computed with each possible value of $p_r$ in mind

• set up the distribution of outcomes for each $p_r$ value, and from that, obtain the probability of $D$ (10 correct out of 20 predictions) in each case
Bayesian Data Analysis

$p_r = .6$

$p = .117$

$p = .031$

$p = .002$

$p = .000$

$p_r = .7$

$p = .000$

$p_r = .8$

$p = .000$

$p_r = .9$
Bayesian Data Analysis

• Considering these four, equally likely possibilities \( (p_r = .6 \text{ or } .7 \text{ or } .8 \text{ or } .9) \), how likely is Madam X to get 10 correct predictions out of 20 tries?

• we apply the multiplication and addition rules of probability

\[
p(D \mid H_1) = .25(.117) + .25(.031) + .25(.002) + .25(.000) = .0375
\]

\[
BF_{01} = \frac{p(H_0 \mid D)}{p(H_1 \mid D)} = \frac{p(D \mid H_0) \cdot p(H_0)}{p(D \mid H_1) \cdot p(H_1)} = \frac{.1762 \cdot .5}{.0375 \cdot .5} = 4.699
\]

\[
p(H_0 \mid D) = \frac{BF_{01}}{1 + BF_{01}} = \frac{4.699}{1 + 4.699} = .825
\]

\[
p(H_1 \mid D) = 1 - p(H_0 \mid D) = 1 - .825 = .175
\]
Bayesian Data Analysis

\[
\frac{p(H_0 \mid D)}{p(H_1 \mid D)} = \frac{p(D \mid H_0)}{p(D \mid H_1)} \cdot \frac{p(H_0)}{p(H_1)}
\]

- Complexity in computing \( p(D \mid H_1) \)
  - integrate across all reasonable values of effect size
- An approximation to this result can be obtained using components of the ANOVA summary table
  - prior distribution of possible effect sizes is centered at observed effect size, normal shape
Bayesian Data Analysis

- Two steps are required

- First, compute the difference between how well the $H_0$ and $H_1$ models fit the data, taking into account the extra parameters used by $H_1$
- consider an example with two independent groups
Bayesian Data Analysis

$M_D = 19$

$M_P = 22$
Bayesian Data Analysis

• The difference in the fit of the models is quantified this way:

\[ \Delta \text{BIC}_{10} = n \cdot \ln \left( \frac{\text{SSE}_1}{\text{SSE}_0} \right) + (k_1 - k_0) \cdot \ln(n) \]

• BIC = Bayesian Information Criterion
• \( \ln \) = natural logarithm
• \( n \) = number of observations
• \( \text{SSE}_1 \) = variability not explained by \( H_1 \)
• \( \text{SSE}_0 \) = variability not explained by \( H_0 \)
• \( k_1 - k_0 \) = difference in number of free parameters for \( H_1 \) and \( H_0 \)
Bayesian Data Analysis

• The second step is to transform $\Delta \text{BIC}_{10}$ into an estimate of the Bayes factor

$$\frac{p(D \mid H_0)}{p(D \mid H_1)} = BF_{01} \approx e^{\Delta \text{BIC}_{10} / 2}$$

(e.g., 4 $\Rightarrow$ 4/1)

Odds to win 2014-15 Stanley Cup

<table>
<thead>
<tr>
<th>Team</th>
<th>Odds (lose/win)</th>
<th>$p(\text{lose})$</th>
<th>$p(\text{win})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chicago Blackhawks</td>
<td>7/1</td>
<td>7/8 = .875</td>
<td>1 – .875 = .125</td>
</tr>
<tr>
<td>New York Rangers</td>
<td>8/1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anaheim Ducks</td>
<td>17/2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Bayesian Data Analysis

• The second step is to transform $\Delta BIC_{10}$ into an estimate of the Bayes factor

$$\frac{p(D \mid H_0)}{p(D \mid H_1)} = BF_{01} \approx e^{\Delta BIC_{10} / 2} \quad (\text{e.g., } 4 \Rightarrow 4/1)$$

Reminder:

$$\frac{p(H_0 \mid D)}{p(H_1 \mid D)} = \frac{p(D \mid H_0)}{p(D \mid H_1)} \cdot \frac{p(H_0)}{p(H_1)}$$

$$p_{BIC}(H_0 \mid D) = \frac{BF_{01}}{1 + BF_{01}} \quad p_{BIC}(H_1 \mid D) = 1 - p_{BIC}(H_0 \mid D)$$
Bayesian Data Analysis

• Application to comparison of two groups

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>109.52</td>
<td>1</td>
<td>109.52</td>
<td>4.31</td>
<td>.043</td>
</tr>
<tr>
<td>Error</td>
<td>1218.96</td>
<td>48</td>
<td>25.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1328.48</td>
<td>49</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[
\Delta \text{BIC}_{10} = n \cdot \ln \left( \frac{\text{SSE}_1}{\text{SSE}_0} \right) + (k_1 - k_0) \cdot \ln(n)
\]

\[
= 50 \cdot \ln \left( \frac{1218.96}{1328.48} \right) + 1 \cdot \ln(50) = -0.390
\]

\[
\text{BF}_{01} \approx e^{\frac{\Delta \text{BIC}_{10}}{2}}
\]

\[
= e^{-0.390/2}
\]

\[
= 0.8228
\]

\[
\text{p}_{\text{BIC}}(H_0 \mid D) = \frac{\text{BF}_{01}}{1 + \text{BF}_{01}} = \frac{.8228}{1 + .8228} = .451
\]

\[
\text{p}_{\text{BIC}}(H_1 \mid D) = 1 - \text{p}_{\text{BIC}}(H_0 \mid D) = 1 - .451 = .549
\]
Bayesian Data Analysis

• Application to a one-factor ANOVA design

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>No distr.</th>
<th>Conver.</th>
<th>Cell phone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>640.00</td>
<td>2</td>
<td>320.00</td>
<td>7.55</td>
<td>.001</td>
<td>8.0</td>
<td>12.0</td>
<td>16.0</td>
</tr>
<tr>
<td>Error</td>
<td>2417.56</td>
<td>57</td>
<td>42.41</td>
<td></td>
<td></td>
<td>6.8</td>
<td>5.4</td>
<td>7.2</td>
</tr>
<tr>
<td>Total</td>
<td>3057.56</td>
<td>59</td>
<td></td>
<td></td>
<td></td>
<td>20</td>
<td>20</td>
<td>20</td>
</tr>
</tbody>
</table>

\[
\Delta BIC_{10} = n \cdot \ln \left( \frac{\text{SSE}_1}{\text{SSE}_0} \right) + (k_1 - k_0) \cdot \ln(n)
\]

\[
= 60 \cdot \ln \left( \frac{2417.56}{3057.56} \right) + 2 \cdot \ln(60) = -5.903
\]

\[
BF_{01} \approx e^{\Delta BIC_{10}/2}
\]

\[
= e^{-5.903/2} = 0.0523
\]

\[
p_{\text{BIC}}(H_0 \mid D) = \frac{BF_{01}}{1 + BF_{01}} = \frac{0.0523}{1 + 0.0523} = 0.0497
\]

\[
p_{\text{BIC}}(H_1 \mid D) = 0.9503
\]
Bayesian Data Analysis

• Interpretation of $p_{\text{BIC}}$ values

<table>
<thead>
<tr>
<th>$p_{\text{BIC}}(H_i \mid D)$</th>
<th>Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>.50 - .75</td>
<td>weak</td>
</tr>
<tr>
<td>.75 - .95</td>
<td>positive</td>
</tr>
<tr>
<td>.95 - .99</td>
<td>strong</td>
</tr>
<tr>
<td>&gt; .99</td>
<td>very strong</td>
</tr>
</tbody>
</table>
Bayesian Data Analysis

• Interpreting a null effect

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>160.00</td>
<td>2</td>
<td>80.00</td>
<td>1.89</td>
<td>.16</td>
</tr>
<tr>
<td>Error</td>
<td>2417.56</td>
<td>57</td>
<td>42.41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>2577.56</td>
<td>59</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[
\Delta BIC_{10} = n \cdot \ln \left( \frac{\text{SSE}_1}{\text{SSE}_0} \right) + (k_1 - k_0) \cdot \ln(n)
\]

\[
= 60 \cdot \ln \left( \frac{2417.56}{2577.56} \right) + 2 \cdot \ln(60) = 4.344
\]

\[
BF_{01} \approx e^{\Delta BIC_{10}/2}
\]

\[
= e^{4.344/2} = 8.7742
\]

\[
\rho_{\text{BIC}}(H_0 \mid D) = \frac{BF_{01}}{1 + BF_{01}} = \frac{8.7742}{1 + 8.7742} = .8977
\]

\[
\rho_{\text{BIC}}(H_1 \mid D) = .1023
\]
Bayesian Data Analysis

• Application to repeated-measures ANOVA

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects</td>
<td>38.49</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>20.40</td>
<td>2</td>
<td>10.20</td>
<td>6.50</td>
<td>.016</td>
</tr>
<tr>
<td>Error</td>
<td>15.67</td>
<td>10</td>
<td>1.57</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>74.56</td>
<td>17</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$\Delta BIC_{10} = n \cdot \ln \left( \frac{SSE_1}{SSE_0} \right) + (k_1 - k_0) \cdot \ln(n)$

$= 12 \cdot \ln \left( \frac{15.67}{36.07} \right) + 2 \cdot \ln(12) = -5.035$

$BF_{01} \approx e^{\Delta BIC_{10}/2}$

$= e^{-5.035/2}$

$= 0.0807$

$p_{BIC} = (H_0 | D) = \frac{BF_{01}}{1 + BF_{01}} = \frac{0.0807}{1 + 0.0807} = 0.0747$

$p_{BIC}(H_1 | D) = 0.9253$

Note: In repeated-measures design, $n$ in $\Delta BIC$ formula is defined as $n(k-1)$
Bayesian Data Analysis

• Bayesian approach resolves various problems with $p$ values under the NHST system
  • it provides what researchers want: $p(H | D)$
  • effective evaluation of validity of the null hypothesis

• Easy to apply in practice (ANOVA-generated info)

• Need to develop new standards of evidence by exploring what Bayesian analysis produces for a wide range of data configurations