Statistical Analysis

Douglas W. Higinbotham, Ph.D.
Summer Opportunities

- Department of Energy SULI Program
  - [http://science.energy.gov/wdts/suli/](http://science.energy.gov/wdts/suli/)
  - Deadline 13 January 2017 !!
  - 10 weeks, $500 per week + Housing & Airfare
  - 18 yrs old, citizen or permanent resident
  - At least an 2.95 GPA

- National Science Foundation REU Programs
  - MANY REU Programs (one example)
    - [https://www.jlab.org/accel/reu/index.html](https://www.jlab.org/accel/reu/index.html)
    - Deadline 17 February 2017
    - 10 weeks, $500 per week + Housing & Air fare
    - Similar requirements to SULI

- I have mentored for both programs and I did an REU as an undergraduate
  - Example SULI project published in journal of undergraduate research
# The 21st Century Building Blocks

## Proton

<table>
<thead>
<tr>
<th>Quarks</th>
<th>Mass (MeV/c²)</th>
<th>Charge</th>
<th>Spin</th>
</tr>
</thead>
<tbody>
<tr>
<td>u</td>
<td>≈2.3</td>
<td>2/3</td>
<td>1/2</td>
</tr>
<tr>
<td>d</td>
<td>≈4.8</td>
<td>-1/3</td>
<td>1/2</td>
</tr>
<tr>
<td>s</td>
<td>≈95</td>
<td>-1/3</td>
<td>1/2</td>
</tr>
<tr>
<td>b</td>
<td>≈1.18</td>
<td>-1/3</td>
<td>1/2</td>
</tr>
<tr>
<td>t</td>
<td>≈173.07</td>
<td>2/3</td>
<td>1/2</td>
</tr>
<tr>
<td>c</td>
<td>≈1.275</td>
<td>2/3</td>
<td>1/2</td>
</tr>
<tr>
<td>g</td>
<td>≈125</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>H</td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

## Leptons

<table>
<thead>
<tr>
<th>Leptons</th>
<th>Mass (MeV/c²)</th>
<th>Charge</th>
<th>Spin</th>
</tr>
</thead>
<tbody>
<tr>
<td>e</td>
<td>&lt;2.2</td>
<td>-1</td>
<td>1/2</td>
</tr>
<tr>
<td>μ</td>
<td>&lt;0.17</td>
<td>-1</td>
<td>1/2</td>
</tr>
<tr>
<td>τ</td>
<td>&lt;15.5</td>
<td>-1</td>
<td>1/2</td>
</tr>
<tr>
<td>νₑ</td>
<td>&lt;0.17</td>
<td>0</td>
<td>1/2</td>
</tr>
<tr>
<td>νₑ</td>
<td>&lt;0.17</td>
<td>0</td>
<td>1/2</td>
</tr>
<tr>
<td>νₜ</td>
<td>&lt;0.17</td>
<td>0</td>
<td>1/2</td>
</tr>
</tbody>
</table>

## Gauge Bosons

<table>
<thead>
<tr>
<th>Gauge Bosons</th>
<th>Mass (MeV/c²)</th>
<th>Spin</th>
</tr>
</thead>
<tbody>
<tr>
<td>γ</td>
<td>0.511</td>
<td>1</td>
</tr>
<tr>
<td>Z</td>
<td>105.7</td>
<td>1</td>
</tr>
<tr>
<td>W</td>
<td>1.777</td>
<td>1</td>
</tr>
<tr>
<td>Z₀</td>
<td>0.12</td>
<td>1</td>
</tr>
<tr>
<td>Higgs</td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

---

*Douglas W. Higinbotham, Ph.D.*

*Duquesne University*

*Jefferson Lab*
Cloud chambers

• A radioactive source in the mixture.
• The clouds show where the particles went.
• Indirectly letting us “see” the passage of particles through matter.
Cosmic Radiation
(and other sources 1mREM per day)
Live Video Feed From CERN’s LHC

Douglas W. Higinbotham, Ph.D.
A Collision from CMS at CERN
An Incredibly Huge Particle Detector
All Models Are Wrong

“The most that can be expected from any model is that it can supply a useful approximation to reality: All models are wrong; some models are useful.”  - George Box (1919 – 2013)

“An ever increasing amount of computational work is being relegated to computers, and often we almost blindly assume that the obtained results are correct.”

- Simon Širca & Martin Horvat
Retrograde Motion of Mars As Seen From Earth
Earth vs. Sun Centered Models

Ptolemaic Model

Copernican Model

Douglas W. Higinbotham, Ph.D.
Phases & Elliptical Orbits

• At first, with orbits as perfect circles, Ptolemaic models were better at predicting the orbits of the planets then Copernican models.

• It was the phases of the Venus (Galileo 1610) along with the elliptical orbits of Kepler (1609) [fitting the “naked eye” data of Brahe (1574)] that proved to be the downfall of the Ptolemaic model.

Illustration by Galileo Galilei in Sidereus Nuncius (Starry Messenger) 1610.
Occam's Razor

- William Occam (1287 – 1347)
- One can always explain failing explanations with an ad hoc hypothesis, thus in Science, simpler theories are preferable to more complex ones. (e.g. the Sun centered vs. Earth centered)
- Layman’s version of Occam’s Razor is “the simplest explanation is usually the correct one” (i.e. KISS)
- In statistical versions of Occam's Razor, one uses a rigorous formulation instead of a philosophical argument. In particular, one must provide a specific definition of simple:
  - F test, Akaike information criterion, Bayesian information criterion, etc.
  - In statistical modeling of data too simple is under-fitting and too complicated is over-fitting.
Bayesian Priors

In 5 min. or less; or I’ll have to do it again next class...
In the Empire Strikes Back C3P0 calculates the odds of navigating an asteroid field.

C3P0 is an excellent robot and does a frequentist calculation to get the odds.

But from a Bayesian point of view, C3P0 is he actually saying Han SHOULD go in the asteroid field!
C3P ignored the fact Han Solo is one of the best pilots in the galaxy.

Never tell me the odds...

...without first establishing a Bayesian prior!

So whether it is a prior this you are adding to the problem (like orbits as perfect circles) or a prior you are ignoring (like Han is great pilot); Bayesian analysis is about being conscious of the additional information you are bringing or not bringing to the problem.

Douglas W. Higinbotham, Ph.D.
Let’s Make A Deal

• Behind Two of the Doors Are Goats ...
• Behind One of the Doors Is The Prize: A Pirate’s Booty!
Monty Hall Problem

Door #1

Door #2

Door #3

Douglas W. Higinbotham, Ph.D.
Monty Hall Problem

Door #1

Wrong

Door #2

Door #3

Douglas W. Higinbotham, Ph.D.
Monty Hall Problem

Door #1

Door #2

Door #3

Wrong

Douglas W. Higinbotham, Ph.D.
Monty Hall Problem

Door #1

Wrong

Door #2

Wrong

Door #3

Right!

Douglas W. Higinbotham, Ph.D.
Probability Says Switch

• Your human reaction is to stay with your original choice.
• I setup the problem before I even arrived today (i.e. you have 1/3 chance to pick correctly at the start)
• My opening a door is NOT random, so the remaining door has a 2/3 chance to be the winner!

If you are still not convinced: https://en.wikipedia.org/wiki/Monty_Hall_problem
What If I Opened 98 Out Of 100 Doors! Would you still not switch?!
Muonic Hydrogen Data

- High precision results from Muonic Lamb shift data give a proton radius of 0.84 fm.
- This result contradicts many other extractions which have determined the proton radius to be ~0.88 fm.

Can anyone figure out what is wrong with the nature cover picture?!
An Example of Statistical Analysis

• So is the radius of the proton 0.84 or 0.88 fm?! 
• Take elastic scattering data 
• Extract a proton radius
Measurement Is Often A Goldilocks Problem

From Deep Space

Too Far

A Modern Telescope

From Orbit

Just Right

Ruler & Some Geometry

On The Planet

Too Close

Theodolite*
What is *just right* for the proton?!

- **We use Plank’s constant** one to relate energy to length in natural units:
  - $Q^2$ of 1 GeV$^2 = 25.7$ fm$^{-2}$.
- Radius of the proton is $\sim 0.84 - 0.88$ fm.
- Thus one can immediately guesstimate that with electron scattering one needs:
  - $Q^2 < (1/0.88 \text{ fm})^2 < 1.2$ fm$^{-2}$ to get the radius of the proton (equivalent to 0.05 GeV$^2$).
  - $Q^2 > 1.2$ fm$^{-2}$ to understand the details of the edge of the proton (e.g. a pion cloud, CQCBM, etc.).
  - $Q^2 \gg 1.2$ fm$^{-2}$ to understand transition from hadronic to partonic (e.g. the bound light constitute quarks).

Guesstimation books by Larry Weinstein (ODU)

*Douglas W. Higinbotham, Ph.D.*
How do we make the electron scattering measurements?

- Beam of electrons from an accelerator (E)
- Place target material in the beam
  - Foils are easy, nearly point (typically thin) targets and thickness is easy to determine
  - Cryo-targets are challenging (e.g. boiling effects, energy loss)
- For elastic hydrogen measure scattered electron (E’) and/or proton.
  - Over determined reaction
- Spectrometers are used
  - Magnetic fields, wire-chambers, reconstructed tracks, sieve data, etc.
Jefferson Lab Hall A Left Spectrometer

Douglas W. Higginbotham, Ph.D.
Extrapolate The Slope of $G_E$ Using Low $Q^2$ Data

$$r_p \equiv \sqrt{\langle r^2 \rangle} = \left( -6 \left. \frac{dG_E(Q^2)}{dQ^2} \right|_{Q^2=0} \right)^{1/2}$$

The question is going to be what function to use for the fitting & extrapolating.

The answer to this question STRONGLY effects the answer!

For linear regression using a polynomial function one can use an F-test. For non-linear regression more complicated techniques are required.
Warning: Danger of Confirmation Bias

In psychology and cognitive science, confirmation bias is a tendency to search for or interpret information in a way that confirms one's preconceptions, leading to statistical errors.
Test of Additional Term

A textbook statistics problem is to quantify when to stop adding terms to a fit of experimental data. One way to do this is with an F-distribution test.

\[ F = \frac{\chi^2(j - 1) - \chi^2(j)}{\chi^2(j)} (N - j - 1) \]

where \( j \) is the order of the fit and \( N \) the number points being fit.

<table>
<thead>
<tr>
<th>( N - j - 1 )</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>6</th>
<th>8</th>
<th>12</th>
<th>20</th>
<th>60</th>
<th>120</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reject ( j^{th} ) order to 95% confidence level if ( F ) is smaller than</td>
<td>18.5</td>
<td>10.1</td>
<td>7.7</td>
<td>6</td>
<td>5.3</td>
<td>4.7</td>
<td>4.3</td>
<td>4</td>
<td>3.9</td>
</tr>
</tbody>
</table>

Quantifies a statement that adding a term doesn’t significantly improve a fit.

One is free to pick a different alpha, alpha=0.05 is just typical to prevent over-fitting.

(see James 2nd edition page 282, Bevington 3rd edition page 207, or Širca page 268)
Real World Example


\[ f(Q^2) = n_0 G_E(Q^2) \approx n_0 \left( 1 + \sum_{i=1}^{m} a_i Q^{2i} \right) \]

<table>
<thead>
<tr>
<th>N</th>
<th>j</th>
<th>( \chi^2 )</th>
<th>( \chi^2/\nu )</th>
<th>( n_0 )</th>
<th>( a_1 )</th>
<th>( a_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>2</td>
<td>13.71</td>
<td>0.623</td>
<td>1.002(2)</td>
<td>-0.119(4)</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>3</td>
<td>13.71</td>
<td>0.652</td>
<td>1.002(5)</td>
<td>-0.120(20)</td>
<td>0.00(2)</td>
</tr>
</tbody>
</table>

F-test rejects fitting with the more complex \( j=3 \) (\( j=m+1 \)) function, that does NOT mean \( a_2 = 0 \).

Pohl et.al’s 0.84 fm radius would predict an \( a_1 \) value of \(-0.1176\) since radius = \( \sqrt{-6a_1} \)

Douglas W. Higinbotham, Ph.D.
Linear and Quadratic Fits of Low $Q^2$ Data & Continued Fraction Fits To $Q^2$ of 1 GeV$^2$ (25.7 fm$^{-2}$)

All three results consistent with the 0.84 fm radius of the Muonic hydrogen Lamb shift.

NOTE: Publishing simple explanations that disagreed with complex 0.88 fm results proved to be amazingly challenging.
### Beyond Simple Fitting: Stepwise Regression

IEEE Rankings are based mostly on CPU usage (i.e. big data)

**2015**

<table>
<thead>
<tr>
<th>Language Rank</th>
<th>Types</th>
<th>Spectrum Ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Java</td>
<td>![Network Icon] ![Mobile Icon] ![Desktop Icon]</td>
<td>100.0</td>
</tr>
<tr>
<td>2. C</td>
<td>![Mobile Icon] ![Desktop Icon] ![Server Icon]</td>
<td>99.9</td>
</tr>
<tr>
<td>3. C++</td>
<td>![Mobile Icon] ![Desktop Icon] ![Server Icon]</td>
<td>99.4</td>
</tr>
<tr>
<td>4. Python</td>
<td>![Network Icon] ![Desktop Icon]</td>
<td>96.5</td>
</tr>
<tr>
<td>5. C#</td>
<td>![Network Icon] ![Mobile Icon] ![Desktop Icon] ![Server Icon]</td>
<td>91.3</td>
</tr>
<tr>
<td>6. R</td>
<td>![Desktop Icon]</td>
<td>84.8</td>
</tr>
<tr>
<td>7. PHP</td>
<td>![Network Icon] ![Desktop Icon]</td>
<td>84.5</td>
</tr>
<tr>
<td>8. JavaScript</td>
<td>![Network Icon] ![Mobile Icon] ![Desktop Icon]</td>
<td>83.0</td>
</tr>
<tr>
<td>9. Ruby</td>
<td>![Network Icon] ![Desktop Icon]</td>
<td>76.2</td>
</tr>
<tr>
<td>10. Matlab</td>
<td>![Desktop Icon]</td>
<td>72.4</td>
</tr>
</tbody>
</table>
Stepwise Regression of $G_E$ from 2014 Data

Douglas W. Higinbotham, Ph.D.

Pohl et.al’s 0.84 fm radius would predict a slope of -0.1176 !!
PRad: Jefferson Lab Hall B Proton Radius Experiment

Small angle and to measure even smaller $Q^2$ to minimize the extrapolation.
Expected Precision of PRad Data

$G_E$ from PRC 93 (2016) 065207

Expected PRad Points

$Q^2 [\text{fm}^{-2}]$
New Atomic Hydrogen Lamb Shift

PRELIMINARY RESULT SHOWN WITH EXPLICATE PERMISSION FROM RANDOLF POHL

Preliminary results from talk given at HC2NP and photos from Trento workshop.
for details see Pohl’s talk at https://indico.cern.ch/event/492464/timetable/#20160930.detailed

Final results should be out soon.

Preliminary results from talk given at HC2NP and photos from Trento workshop.
for details see Pohl’s talk at https://indico.cern.ch/event/492464/timetable/#20160930.detailed

Douglas W. Higinbotham, Ph.D.
LIGO Observatories (Searching for Gravitational Waves)
You Can **Hear the Signal “Chirp”**
Example Source: Orbital Decay of a Binary System

LIGO Hanford Observatory: GW150914

Black Hole is the Final State
Last Lecture: Real World Example

- Idea
- Defense
- Running
- Analysis
- Publications
- New Ideas