Moths & Meadows

William Berk, Patrick Waters, and Sean McGregor
Why are Moths Important?

- Primary consumers
- Indicator of ecosystem diversity and health
- Easy to sample, good for climate change studies
Gathering New Data: William Berk
Study Objectives

• How does this affect the flora and fauna of HJA?
• Start by studying moths.
• Observation
• Results could inform future actions.
Meadow Selection

- Selected on the basis of size and location.
- Assisted by ArcGIS shapefiles.
- Meadows larger than average were “large.”
- Meadows further from each other than average were “isolated.”
Study Method

• Three traps, three batteries.
  • One in the meadow.
  • One in the forest.
  • One on the border.
• Set the traps before nightfall.
• Collect the traps after daybreak.
• Recharge the batteries and empty the traps.
• Repeat for each area.
  • Area 1: July 31, 2008.
  • Area 2: August 6, 2008.
  • Area 3: August 7, 2008.
  • Area 4: August 12, 2008.
Moth Sorting

• Sorted by species to best of ability.
• Labeled individually with letters.
• Nine “Other” species also recorded.
• Results from each trap carefully logged.
Results

• 360 different species collected.
• Further study is necessary, but trends are already clear.

<table>
<thead>
<tr>
<th></th>
<th>M1</th>
<th>B1</th>
<th>F1</th>
<th>M2</th>
<th>B2</th>
<th>F2</th>
<th>M3</th>
<th>B3</th>
<th>F3</th>
<th>M4</th>
<th>B4</th>
<th>F4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other2</td>
<td>10</td>
<td>11</td>
<td>22</td>
<td>20</td>
<td>22</td>
<td>14</td>
<td>9</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>KK</td>
<td></td>
<td></td>
<td>1</td>
<td>4</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DZ</td>
<td></td>
<td></td>
<td>18</td>
<td>4</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>AO</td>
<td>29</td>
<td>2</td>
<td></td>
<td>4</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AP</td>
<td>26</td>
<td>7</td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GY</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td>25</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LK</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LX</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MJ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ML</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Conclusions

• Different environments for different moths.
• Dependence on environments can be extreme.
• Change could be beneficial or hazardous, depending on the species.
Further Study

• Properly identify species.
• Compare with existing data.
• Study other insects in the area.
• Study other attributes of the meadows.
Moth Metapopulation Models

- A population is divided spatially into metapopulations
- Each location can experience local extinction or colonization
- A graph can be used to represent which pairs meadows moths can travel between

Model Life Cycle of a Rare Caterpillar Living in the $i^{th}$ Meadow

- To become a moth and successfully reproduce it must win a Bernoulli trial at probability $p_i$.
- If it wins, it chooses a site to reproduce at. The chance of site $j$ being chosen is $T_{ij}$ (this mandates that $\Sigma_j T_{ij}=1$).
- Each victorious moth begets $r$ caterpillars at its chosen site.
Example Run of Metapopulation Model
Data mining with the Box Algorithm

• Goal: Find the best n-dim’l box such that at least m data points are contained, and a sample in the box tests positive p% of the time
Moth Habitat Preference

• Suppose moths are sampled from many locations with different parameters

• We can use a box algorithm to find what habitat types each species prefers

• How do we measure our confidence that these moths prefer any particular habitat at all?
Certainty in habitat inference

• Method: Partition all habitats into mutually exclusive habitat types and find the type with the most average moths per sample
• Hypothesis: Moths are most likely to be found in this habitat type
• Null Hypothesis: Moths do not prefer a particular habitat type
• Unlikely: Moths prefer a different habitat type
Certainty in habitat inference

• Suppose the null hypothesis is true, then how unlikely was it to find data as correlated as ours?

• If it is unlikely that random data would generate so many moths in any habitat type, then we are confident that our data is not random
Our problem

• Suppose that each of M moths is independently found in any sample with equal chance

• Each sample is in some habitat type, and there are N habitat types, each i\textsuperscript{th} habitat type including \(n_i\) samples

• What is the probability that at least one habitat type averages at least A moths per sample?
A simpler problem

• How many ways are there to distribute $M$ moths to $N$ samples so that no sample has more than $L$ moths?

• Consider the $m^{\text{th}}$ coefficient of $f(x)$

$$f(x) = m! \left( \frac{x^0}{0!} + \frac{x^1}{1!} + \frac{x^2}{2!} + \ldots + \frac{x^L}{L!} \right) \cdot \left( \frac{x^0}{0!} + \frac{x^1}{1!} + \frac{x^2}{2!} + \ldots + \frac{x^L}{L!} \right) \cdot \ldots \cdot \left( \frac{x^0}{0!} + \frac{x^1}{1!} + \frac{x^2}{2!} + \ldots + \frac{x^L}{L!} \right)$$

• Expand the product

• Identify choosing the $i^{\text{th}}$ term of the $j^{\text{th}}$ sum with putting $i$ moths in the $j^{\text{th}}$ sample
A simpler problem continued

\[ f(x) = M! \left( \frac{x^0}{0!} + \frac{x^1}{1!} + \frac{x^2}{2!} + \ldots + \frac{x^L}{L!} \right) \cdot \left( \frac{x^0}{0!} + \frac{x^1}{1!} + \frac{x^2}{2!} + \ldots + \frac{x^L}{L!} \right) \cdot \ldots \cdot \left( \frac{x^0}{0!} + \frac{x^1}{1!} + \frac{x^2}{2!} + \ldots + \frac{x^L}{L!} \right) \]

- There are \( M!/n_1!n_2!\ldots n_N! \) ways to distribute \( n_1, n_2, \ldots, n_N \) moths to samples 1, 2, \ldots, N
- The power rule for derivatives allows us to write our answer

\[ g(M) = \frac{d^M}{dx^M} \prod_{i=1}^{N} \left( \frac{x^0}{0!} + \frac{x^1}{1!} + \frac{x^2}{2!} + \ldots + \frac{x^L}{L!} \right) \bigg|_{x=0} \]

- \( g(M) \) is the number of ways to distribute M moths to N samples so that no sample has more than L moths
Back to our original problem

- We can modify our previous answer to give the number of ways to distribute $M$ moths to samples in $N$ habitat types such that there are $n_i$ samples in each $i^{th}$ type and no type averages more than $A$ moths per sample.

$$F(M) = \left. \frac{d^M}{dx^M} \left( \prod_{i=1}^{N} \sum_{j=0}^{\left\lfloor n_iA \right\rfloor} \frac{(n_i x)^j}{j!} \right) \right|_{x=0}$$

- Where $\left\lfloor n_iA \right\rfloor$ is the greatest integer less than $n_iA$. 
Solution

• In the null hypothesis, since each moth is independently found in any sample with equal chance

• In this case, the probability that at least one habitat type averages more than \( A \) moths per sample is

\[
P = 1 - \frac{\frac{d^M}{dx^M} \left( \prod_{i=1}^{N} \sum_{j=0}^{n_i A} \frac{(n_i x)^j}{j!} \right) \bigg|_{x=0}}{\left( \sum_{i=1}^{N} n_i \right)^M}
\]
Application of box method

• For example, if the box algorithm is run on Jeff Miller’s data for the abundance of Achytonix praeacuta, and the habitat characteristics elevation and slope aspect are resolved into three category types each, then we find that:
  • Achytonix praeacuta is most likely to be found in habitats with $0 < \text{aspect} < \frac{2\pi}{3}$ and $1440m < \text{elevation} < 2623m$.
  • Such a sample on average 0.26 of these moths
  • If these moths had no preference, at least one habitat type would average at least this many moths only 0.079% of the time
Computer Science and Moths & Meadows
Advantages of Databases

- Scalable >65,000 x IV
- Faster
- Works on any platform
- Interfaces with visualization programs
- Advanced SQL queries
- It is Free
Select all species found at trap number 1 in the genus Semiothisa translates to:

```
SELECT species.species FROM species, abundance, samples, traps WHERE 
species.binomial_genus="Semiothisa" AND 
species.species_number=abundance.species_number AND 
abundance.sample_number=samples.sample_number;
```
The Web Framework

Not the Jazz Musician
Database+Framework=kml
Moth Guilds

• No single definition
• Settled on “common food plants”
Apriori Algorithm

Baskets have diapers, chips, and beer

Support 11.4/100

Confidence 95.1/100

Baskets with diapers and chips also have beer

Gabriola-dyari + Semiothisa-signaria -> Nepytia-umbrosaria

Support 11.4

Confidence 95.1
Semiothisa-signaria <- Probole-alienaria (10.7, 96.3)
Semiothisa-signaria <- Eustroma-fasciata (11.7, 86.4)
Semiothisa-signaria <- Stamnoctenis-pearsalli (11.2, 96.5)
Nepytia-umbrosaria <- Achytonix-epipaschia (12.5, 89.2)
Pero-mizon <- Achytonix-epipaschia (12.5, 82.8)
Pero-mizon <- Scopula-junctaria (13.0, 82.8)
Semiothisa-signaria <- Scopula-junctaria (13.0, 87.1)
Semiothisa-signaria <- Epirrhoe-alternata (13.5, 85.2)
Pero-mizon <- Pseudorthodes-irrorata (12.7, 85.6)
Semiothisa-signaria <- Pseudorthodes-irrorata (12.7, 86.2)
Pero-mizon <- Hemeroplanis-finitima (12.5, 84.7)
Semiothisa-signaria <- Hemeroplanis-finitima (12.5, 86.0)
Nepytia-umbrosaria <- Gabriola-dyari (16.5, 86.5)
Semiothisa-signaria <- Anagoga-occiduaria (15.3, 93.2)
Pero-mizon <- Clemensia-albata (15.3, 85.9)
Semiothisa-signaria <- Clemensia-albata (15.3, 89.6)
Semiothisa-signaria <- Hydriomena-renunciata (17.1, 87.0)
Pero-mizon <- Ceratodalia-gueneata (16.6, 82.7)
Semiothisa-signaria <- Ceratodalia-gueneata (16.6, 82.7)
Caripeta-divisata <- Campaea-perlata (14.3, 80.0)
Perizoma-grandis <- Campaea-perlata (14.3, 84.4)
Semiothisa-signaria <- Campaea-perlata (14.3, 92.2)
Pero-mizon <- Euchlaena-tigrinaria (15.0, 80.4)
Semiothisa-signaria <- Euchlaena-tigrinaria (15.0, 95.8)
Semiothisa-signaria <- Diarsia-esurialis (18.2, 84.2)
Apriori Does Not Answer the Guild Question

• “Basket” analogy has too much baggage
  – Neglects temporal effects
    • Emergence times related to climate
    • Guilds do not necessarily emerge at the same time
    • Niches are too complex
Attempt 2: Topic Models with LDA

• LDA - “Latent Dirichlet Allocation”
  – Need more than 30 minutes to explain the theory
  – Comes from the world of document modeling
  – Every sample is a “document”
  – Uses probability voodoo to generate “topics” which have a probability of generating each species of moth
  – Each sample then has a probability of being generated from a topic
Topics

Topic Cow:
(From fictional farmers almanac)
Milk 0.204568
Beef 0.141886
Steak 0.055830
Grass 0.054236
Free-range 0.051049
Cruelty-free 0.038300
Cage-free 0.032988
Veal 0.031394
Nebraska 0.031394
Cheese 0.019708
Butter 0.017052
Heifer 0.014927
Bull 0.014396
State 0.012802
Rancher 0.011740
Veal 0.011208
Mignon 0.011208
Filet 0.011208
Sheep 0.010677
Burger 0.009615

Topic 0th:
(From moth data)
Homorthodes-hanhami 0.204568
Hemeroplanis-finitima 0.141886
Homorthodes-furfurata 0.055830
Eudrepanulatrix-rectifascia 0.054236
Aseptis-ethnica 0.051049
Semiothisa-signaria 0.038300
Lacinipolia-patalis 0.032988
Nadata-gibbosa 0.031394
Leucania-farcta 0.031394
Synedoida-adumbrata 0.019708
Synedoida-ochracea 0.017052
Pero-mizon 0.014927
Euchlaena-johnsonaria 0.014396
Perizoma-costiguttata 0.012802
Synedoida-sabulosa 0.011740
Lacinipolia-strigicollis 0.011208
Grammia-ornata 0.011208
Nemoria-darwiniata 0.011208
Semiothisa-californiaria 0.010677
Zosteropoda-hirtipes 0.009615
Topic Overlays
Conclusions

• Check back with Jeff Miller and various computer science graduate students
Acknowledgments

• Thanks to
  – Jeff Miller
  – Steven Highland
  – Desiree Tullos & Julia Jones
  – Weng-Keen Wong & Tom Dietterich
  – Yevgeniy Kovchegov
For More Information

See: www.atlasoflife.com/

Questions?