Lecture 3: Probability, conditional probability, and random variables

Jason Mezey
jgm45@cornell.edu
Jan. 28, 2020 (T) 8:40-9:55
Announcements I: Registration

• All students (undergrad / grad) at Cornell, Weill, Tech, Rockefeller, MSK should now be registered

• If you are “other” (=not a student) at Cornell (Ithaca) with permission you can register - I am still looking into whether we can waive the fee… please stay tuned

• If you are a “other” (=not a student) at Weill or MSK I am still waiting for the Weill grad school to get us permissions / instructions on registration (again, please stay tuned…)
Announcements II: Piazza

- It is essential that everyone get up on Piazza by end of the day today

- If you cannot sign-up yourself, you need to email me directly and I will enroll you (jgm45 at cornell dot edu)

- If you have already emailed me (e.g., last night, this morning) I have your email and will get you up today

- I will send out a Piazza test email tomorrow (Weds) morning - if you do not receive the email by Weds. Afternoon, please email me asap…
• For CMS - if you can enroll yourself in the class please do so

• For everyone else please Piazza me that you need to get up on CMS and include your email (again if you have done this already - I have your email and we are working on it)

• Our goal is to get everyone up on CMS so you can SUBMIT your first homework (due next week) on CMS

• Your first homework assignment will be AVAILABLE on the class site tomorrow (Weds)!
Announcements IV: Homework #1

• Homework #1 will be posted on the class website TOMORROW (Weds.)

• Problems are divided into “easy”, “medium, and “hard”

• All homeworks are “open book” and you may work together but you MUST hand in your own work (i.e., a copy of someone’s written answer will not be accepted)

• Answers must be typed (!!) - please talk to us if this is a problem...

• We will ask you to upload your answers to CMS

• You must upload / email your homework by 11:59PM, Tues., Feb 4 (otherwise it is late - no excuses!!)
Announcements V: Labs

• Computer lab this week (!!) - Bring your laptop!

• Note that in Ithaca (= Labs taught by Rachel!):
  • Lab 1 will meet 5-6PM on Thurs. / Lab 2 Fri. 8-9AM
  • The location of these labs will be Weill 226 (!!)

• Note that in NYC (= Labs taught by Scott!):
  • Lab 1 will meet 4-5PM on Thurs. and will meet in:
    1300 York, LC 2nd Floor, R [D236], S [D234]
  • We are working on setting up a second lab session (Lab 2)
    that will meet Fri. 9-10AM, location TBD (stay tuned)!

• Note that while labs are required, the first three labs are
  OPTIONAL if you are already highly proficient at R

• I’m going to send an information “test” by Piazza tomorrow
  (Weds) for you to help you assess whether you can skip the
  first three labs
Summary of lecture 3:

• Last lecture, we introduced critical concepts for modeling genetic systems, including rigorous definitions of experiments, sample spaces

• In this lecture, we will consider sigma algebras and probability functions and then add the following critical building blocks: conditional probabilities and independence
Review: Thoughts about what probability is modeling

- We are attempting to model the results of a non-ideal experiment to understand a system
- Such experiments include extensive amounts of uncontrolled aspects (important for the system!) that we usually cannot specify
- What we may be able to do is provide a reasonable model of how these uncontrolled aspects impact the results of the experiments
• **Set** - any collection, group, or conglomerate

• **Element** - a member of a set

• **Set Operations:**
  - **Union** ($\cup$) ≡ an operator on sets which produces a single set containing all elements of the sets.
  - **Intersection** ($\cap$) ≡ an operator on sets which produces a single set containing all elements common to all of the sets.

• **Important Definitions:**
  - **Element of** ($\in$) ≡ an object within a set, e.g. $H \in \{H, T\}$
  - **Subset** ($\subset$) ≡ a set that is contained within another set, e.g. $\{H\} \subset \{H, T\}$
  - **Complement** ($A^c$) ≡ the set containing all other elements of a set other than $A$, e.g. $\{H\}^c = \{T\}$.
  - **Disjoint Sets** ≡ sets with no elements in common.

• **A Special Set:** **Empty Set** ($\emptyset$) ≡ the set with no elements (the empty set is unique and is sometimes represented as \{ \}).
Review: Sample Spaces

- **Sample Space** \( (\Omega) \) - set comprising all possible outcomes associated with an experiment

- Examples (Experiment / Sample Space):
  - “Single coin flip” / \{H, T\}
  - “Two coin flips” / \{HH, HT, TH, TT\}
  - “Measure Heights” / any actual measurement OR we could use \( \mathbb{R} \)

- **Events** - a subset of the sample space

- Examples (Sample Space / Examples of Events):
  - “Single coin flip” / \( \emptyset, \{H\}, \{H,T\} \)
  - “Two coin flips” / \{TH\}, \{HH, TH\}, \{HT, TH, TT\}
  - “Measure Heights” / \{1.7m\}, \{1.5m, ..., 2.2m\} OR \[1.7m], (1.5m, 1.8m)
Review: Probability functions (intuition)

- **Probability Function** (intuition) - we would like to construct a function that assigns a number to each event such that it matches our intuition about the “chance” the event will happen (as a result of an experiment)

- To be useful, we need to assign a number not just to each individual element of the set but to EVERY event

- To accomplish this, we will need the concept of a **Sigma Algebra** (or **Sigma Field**)
Sample Spaces / Sigma Algebra

- **Sigma Algebra** ($\mathcal{F}$) - a collection of events (subsets) of $\Omega$ of interest with the following three properties: 1. $\emptyset \in \mathcal{F}$, 2. $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$, 3. $A_1, A_2, \ldots \in \mathcal{F}$ then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

  *Note that we are interested in a particular Sigma Algebra for each sample space...*

- Examples (Sample Space / Sigma Algebra):
  - $\{H, T\} / \emptyset, \{H\}, \{T\}, \{H, T\}$
  - $\{HH, HT, TH, TT\} /$ \hspace{1cm} \begin{align*}
  \emptyset, \{HH\}, \{HT\}, \{TH\}, \{TT\}, \{HH, HT\}, \{HH, TH\}, \{HH, TT\}, \{HT, TH\}, \{HT, TT\},
  \{TH, TT\}, \{HH, HT, TH\}, \{HH, HT, TT\}, \{HH, TH, TT\}, \{TH, HT, TT\}, \{HH, TH, HT, TT\}
  \end{align*}
  - $\mathbb{R}$ / more complicated to define the sigma algebra of interest (see next slide…)
- Note that the pair $(\Omega, \mathcal{F})$ is referred to as a measurable space
The (appropriate) Sigma Algebra on the Reals

• For probability, we need an appropriate Sigma Algebra on the Reals (remember there are many possible Sigma Algebra!)

• Interestingly, this Sigma Algebra does not include all subsets of the reals

• One problem is this would include “more sets than we need” for what we need in probability

• Another problem is these subsets include “non-measurable sets” such that if they were included, we could not define a probability measure (!!!)

• A way of describing the appropriate Sigma Algebra for the Reals is all open and closed intervals (where $a$ and $b$ may be any number) and all unions and intersections of these intervals:

\[ [a, b], (a, b], [a, b), (a, b) \]

• It seems like these should include all subsets of the Reals, but they don’t...
Probability functions

- **Probability Function** - maps a Sigma Algebra of a sample to a subset of the reals:

\[ Pr(\mathcal{F}) : \mathcal{F} \rightarrow [0, 1] \]

- Not all such functions that map a Sigma Algebra to [0,1] are probability functions, only those that satisfy the following Axioms of Probability (where an axiom is a property assumed to be true):

1. For \( \mathcal{A} \subset \Omega \), \( Pr(\mathcal{A}) \geq 0 \)

2. \( Pr(\Omega) = 1 \)

3. For \( \mathcal{A}_1, \mathcal{A}_2, \ldots \in \Omega \), if \( \mathcal{A}_i \cap \mathcal{A}_j = \emptyset \) (disjoint) for each \( i \neq j \): \( Pr(\bigcup_{i=1}^{\infty} \mathcal{A}_i) = \sum_{i=1}^{\infty} Pr(\mathcal{A}) \)

- Note that since a probability function takes sets as an input and is restricted in structure, we often refer to a probability function as a *probability measure*

- Note that the pair \( (\Omega, \mathcal{F}) \) is referred to as a measurable space and the triple \( (\Omega, \mathcal{F}, Pr) \) is referred to as a measure space
Probability function: example 1

- For “two coin flips” a probability function will assign a probability to each subset of the Sigma Field:

  $\emptyset, \{HH\}, \{HT\}, \{TH\}, \{TT\}, \{HH, HT\}, \{HH, TH\}, \{HH, TT\}, \{HT, TH\}, \{HT, TT\}, \{TH, TT\}\{HH, HT, TH\}, \{HH, HT, TT\}, \{HH, TH, TT\}, \{TH, HT, TT\}\{HH, TH, HT, TT\}$

- We could define a probability function as follows:

  \[
  Pr(\emptyset) = 0 \\
  Pr(HH) = 0.25, Pr(HT) = 0.25, Pr(TH) = 0.25, Pr(TT) = 0.25 \\
  Pr(HH \cup HT) = 0.5, Pr(HH \cup TH) = 0.5, Pr(HH \cup TT) = 0.5 \\
  Pr(HT \cup TH) = 0.5, Pr(HT \cup TT) = 0.5, Pr(TH \cup TT) = 0.5 \\
  Pr(HH \cup HT \cup TH) = 0.75, \text{ etc. } Pr(HH \cup HT \cup TH \cup TT) = 1.0
  \]

- Not that this is one possible probability model - what other possible probability models could be assumed for this system / experiment?
Probability function: example II

- The following is (one example) of a probability function (on the sigma algebra) for the two coin flip experiment:

\[ Pr(\emptyset) = 0 \]
\[ Pr(HH) = 0.25, Pr(HT) = 0.25, Pr(TH) = 0.25, Pr(TT) = 0.25 \]
\[ Pr(HH \cup HT) = 0.5, Pr(HH \cup TH) = 0.5, Pr(HH \cup TT) = 0.5 \]
\[ Pr(HT \cup TH) = 0.5, Pr(HT \cup TT) = 0.5, Pr(TH \cup TT) = 0.5 \]
\[ Pr(HH \cup HT \cup TH) = 0.75, \text{ etc. } Pr(HH \cup HT \cup TH \cup TT) = 1.0 \]

- The following is an example of a function (on the sigma algebra) of the two coin flip experiment but is not a probability function:

\[ Pr(\emptyset) = 0 \]
\[ Pr(HH) = 0.25, Pr(HT) = 0.25, Pr(TH) = 0.25, Pr(TT) = 0.25 \]
\[ Pr(HH \cup HT) = 0.5, Pr(HH \cup TH) = 0.5, Pr(HH \cup TT) = 0 \]
\[ Pr(HT \cup TH) = 1.0, Pr(HT \cup TT) = 0.5, Pr(TH \cup TT) = 0.5 \]
\[ Pr(HH \cup HT \cup TH) = 0.75, \text{ etc. } Pr(HH \cup HT \cup TH \cup TT) = 1.0 \]
Probability functions on the Reals

• For a sample space that is the Reals, recall that a probability function has to assign (appropriate) probabilities to the Sigma Algebra: open and closed (and half open-closed) intervals and unions, intersections

• Taking just the one to one mapping of a function on the reals would not assign intervals (subsets of the Sigma Algebra) a probability

• ...unless we are careful about the construction and how we make use of a function (see “Random Variables” next lecture!)
Probability functions on the Reals

• In any realistic case, our true sample outcomes will be discrete

• However we often model cases where the Reals have nice properties and mathematical tools available that we can leverage

• An intuitive way to conceptualize how we make use of the Reals is when we use them for the sample space (!!)

• Note: statisticians sometimes conceptualize sample spaces as abstract and a mapping to the Reals for random variables… we’ll discuss later

• Example: we often use the real numbers as the sample space for an experiment where the sample outcome is human heights

• Two questions:
  • What approximations are we making?
  • Why are these approximations reasonable?
Probability functions on the Reals

• For a sample space that is the Reals, recall that a probability function has to assign (appropriate) probabilities to the Sigma Algebra: open and closed (and half open-closed) intervals and unions, intersections

• Taking just the one to one mapping of a function on the reals would not assign intervals (subsets of the Sigma Algebra) a probability

• ... unless we are careful about the construction and how we make use of a function (see “Random Variables” next lecture!)
Additional thought about what probability is modeling

• Again, note that our eventual objective is to provide a reasonable model of how these uncontrolled aspects impact the results of the experiments

• More specifically, we assume that the impact of the uncontrolled aspects are random but where certain outcomes are more probable than others (note the assumption!!)

• This is what a probability function is built to model (= to provide the probability of random outcomes of an experiment)

• Note that while random is intuitive, it’s a problematic concept...
Essential concepts: conditional probability and independence

• As well as having an intuitive sense of what it means for something we observe to be random (within definable rules) we also have an intuitive sense about how the rules change once we observe specific outcomes or assume certain possibility applies

• This intuition is captured in conditional probability

• This is the essential concept in any area of probabilistic modeling, where the concept of independence directly follows

• In fact, almost anything we are doing in statistics, machine learning, etc. is really attempting to identify or leverage conditional probabilities

• As an example, we could consider the conditional probability that someone will be taller or shorter if they have a “T” at a particular position in the genome
Conditional probability

- We have an intuitive concept of *conditional probability*: the probability of an event, given another event has taken place.
- We will formalize this using the following definition (note that this is still a probability!!):

  The formal definition of the conditional probability of $A_i$ given $A_j$ is:

  $$ Pr(A_i|A_j) = \frac{Pr(A_i \cap A_j)}{Pr(A_j)} $$

- While not obvious at first glance, this is actually an intuitive definition that matches our conception of conditional probability.
An example of conditional prob.

- Consider the sample space of “two coin flips” and the following probability model: \( Pr\{HH\} = Pr\{HT\} = Pr\{TH\} = Pr\{TT\} = 0.25 \)

<table>
<thead>
<tr>
<th></th>
<th>( H_{2nd} )</th>
<th>( T_{2nd} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( H_{1st} )</td>
<td>( HH )</td>
<td>( HT )</td>
</tr>
<tr>
<td>( T_{1st} )</td>
<td>( TH )</td>
<td>( TT )</td>
</tr>
</tbody>
</table>

\[
\begin{array}{ccc}
\text{\( H_{1st} \)} & \text{\( Pr(H_{1st} \cap H_{2nd}) \)} & \text{\( Pr(H_{1st} \cap T_{2nd}) \)} & \text{\( Pr(H_{1st}) \)} \\
\text{\( T_{1st} \)} & \text{\( Pr(T_{1st} \cap H_{2nd}) \)} & \text{\( Pr(T_{1st} \cap T_{2nd}) \)} & \text{\( Pr(T_{1st}) \)} \\
\text{\( Pr(H_{2nd}) \)} & \text{\( Pr(T_{2nd}) \)} & \text{\( Pr(T_{2nd}) \)} & \text{\( Pr(T_{2nd}) \)}
\end{array}
\]

\( Pr(H_{1st}) = Pr(HH \cup HT), \ Pr(H_{2nd}) = Pr(HH \cup TH) \)

\( Pr(T_{1st}) = Pr(TH \cup TT), \ Pr(T_{2nd}) = Pr(HT \cup TT) \)
An example of conditional prob.

- Intuitively, if we condition on the first flip being “Heads”, we need to rescale the total to be one (to be a probability function):

<table>
<thead>
<tr>
<th></th>
<th>$H_{2nd}$</th>
<th>$T_{2nd}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_{1st}$</td>
<td>HH</td>
<td>HT</td>
</tr>
<tr>
<td>$T_{1st}$</td>
<td>TH</td>
<td>TT</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$H_{2nd}$</th>
<th>$T_{2nd}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_{1st}$</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>$T_{1st}$</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.5</td>
</tr>
</tbody>
</table>
An example of conditional prob.

- Intuitively, if we condition on the first flip being “Heads”, we need to rescale the total to be one (to be a probability function):

\[
\begin{array}{c|cc}
 & H_{2nd} & T_{2nd} \\
\hline
H_{1st} & HH & HT \\
T_{1st} & TH & TT \\
\end{array}
\]

\[
\begin{array}{c|ccc}
 & H_{2nd} & T_{2nd} \\
\hline
H_{1st} & 0.25 & 0.25 & 0.5 \\
T_{1st} & 0.25 & 0.25 & 0.5 \\
\end{array}
\]

\[
\begin{array}{c|c}
 & H_{2nd} \\
\hline
H_{1st} & 0.25 \\
T_{1st} & 0.25 \\
\end{array}
\]
An example of conditional prob.

- Intuitively, if we condition on the first flip being “Heads”, we need to rescale the total to be one (to be a probability function):

<table>
<thead>
<tr>
<th></th>
<th>$H_{2nd}$</th>
<th>$T_{2nd}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_{1st}$</td>
<td>$HH$</td>
<td>$HT$</td>
</tr>
<tr>
<td>$T_{1st}$</td>
<td>$TH$</td>
<td>$TT$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$H_{2nd}$</th>
<th>$T_{2nd}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_{1st}$</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>$T_{1st}$</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.5</td>
</tr>
</tbody>
</table>

$$Pr(H_{2nd}|H_{1st}) = \frac{Pr(H_{2nd} \cap H_{1st})}{Pr(H_{1st})} = \frac{Pr(HH)}{Pr(HH \cup HT)} = \frac{0.25}{0.5} = 0.27$$
That’s it for today

• Next lecture, we will finish our discussion of conditional probability (and independence) and introduce random, pdf’s and cdf’s, and random vectors