Lecture 3: Conditional probability, random variables, and probability distribution functions

Jason Mezey
jgm45@cornell.edu
Feb. 4, 2016 (Th) 8:40-9:55
Announcements

• First Computer lab is today (Thurs.!! Not Fri.):
  • In Ithaca, 5-6PM (!!) in MNLB30A (Mann Library Basement)
  • In NYC, TODAY’S LAB IS 4-5PM (!!) in Belfer 204-A (I will send out an email announcement...), we are working on scheduling rooms for future labs from 4-5PM (stay tuned)

• Homework #1 is posted on the class website:
  • Answers must be typed (!!) - please talk to us if this is a problem...
  • You must email your answers to your appropriate TA (Mahya - Ithaca; Jin - NYC) by 11:59PM on Mon. 2/8 (otherwise it is late!)
  • Homeworks are “open book” and you may work together but you MUST hand in your own work

• Jason’s office hours today (!!) 3:30-5PM (normally Thurs. 3-5PM) in 101 Biotech Suite AND Dept. Genetic Med. conference
Summary of lecture 3:

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

• In this lecture, we will add the following critical building blocks: conditional probabilities and independence, random variables, and we will begin discussing probability distributions
Conceptual Overview

System

Question

Inference

Sample

Prob. Models

Statistics

Assumptions
Review 1

\[ Pr(\mathcal{F}) \]

Experiment \[ \Omega \] \[ \mathcal{F} \]

(Sample Space) (Sigma Algebra)

This concept is often introduced to us as output. Function or ability function.

To use sample spaces in probability, we need a way to map these sets to the real numbers.

5 Probability Functions

\[ X \]

A

\[ A \]

\[ 1 \]

\[ X \]

F

To do this, we define a measure, \[ \{ A \} \]

where \( (\text{to numbers}) \):

We are going to define a from class).

the values taken by

This

\[ \text{function} \]

before we consider the specifics of how we define a measure.

\[ \text{for example, we can have the function} \]

\[ f \]

\[ \text{which maps} \]

\[ \text{sample spaces} \]

\[ \text{to} \]

\[ \text{numbers} \].

Before we consider the specifics of how we define a function, let's consider the intuitive definition of a function:

\[ X \]

\[ X \]

\[ F \]

\[ t \]

\[ \text{where} \]

\[ \text{intuitive def.} \]

\[ \text{operator that takes an input and produces an output.} \]

Before we consider the specifics of how we define a function, let's consider the intuitive definition of a function:

\[ X \]

\[ F \]

\[ t \]

\[ \text{where} \]

\[ \text{intuitive def.} \]

\[ \text{operator that takes an input and produces an output.} \]
Review II

• **Experiment** - a manipulation or measurement of a system that produces an outcome we can observe

• **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” / \{5’, 5’3”, 5’3.5”, ...\}

• **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. $A_1, A_2, ... \in \mathcal{F}$ 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\} / see last lecture
  • any actual measurement OR we could use $\mathbb{R}$ / see last lecture
Review III

- **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 \( A \subset \Omega \), \( Pr(A) \geq 0 \)

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

3. For \( A_1, A_2, \ldots \in \Omega \), if \( A_i \cap A_j = \emptyset \) (disjoint) for each \( i \neq j \): \( Pr(\bigcup_i^\infty A_i) = \sum_i^\infty Pr(A) \)

- 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 triple \( (\Omega, \mathcal{F}, Pr) \) is referred to as a measurable space
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!!):

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

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

The formal definition of the conditional probability of \( A_i \) given \( A_j \) is:

\[
Pr(A_i|A_j) = \frac{Pr(A_i \cap A_j)}{Pr(A_j)}
\]
An example of conditional prob.

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

\[
\begin{array}{ccc}
H_{1st} & H_{2nd} & T_{2nd} \\
HH & HT & TT \\
TH & & \\
T_{1st} & & \\
\end{array}
\]

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

\[\Pr(H_{1st}) = \Pr(HH \cup HT), \ Pr(H_{2nd}) = \Pr(HH \cup TH)\]
\[\Pr(T_{1st}) = \Pr(TH \cup TT), \ Pr(T_{2nd}) = \Pr(HT \cup TT)\]
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>

- By defining the conditional probability as 'the probability of an event given that another event has taken place', this concept makes formal the case where an event that can define the conditional probability as ‘the probability of an event given that another event has taken place’.

- The concept of conditional probability conforms to the definition of a probability function, so if we think of the sample space as $HH$, $HT$, and $TT$, then after restricting ourselves to these two cases, we have to pick the old sample space $HH, HT, TT$, and the new sample space be one.

- The formula in equation (1) comes from making the total probability of the first row in the original sample space equal to one, and after rescaling the total to be one (to be a probability function):

$$Pr(H_{2nd}|H_{1st}) = \frac{Pr(H_{2nd} \cap H_{1st})}{Pr(H_{1st})} = \frac{Pr(HH)}{Pr(HH \cup HT)} = \frac{0.25}{0.5} = 0.5$$
Independence

- The definition of independence is another concept that is not particularly intuitive at first glance, but it turns out it directly follows our intuition of what “independence” should mean and from the definition of conditional probability.

- Specifically, we intuitively think of two events as “independent” if knowing that one event has happened does not change the probability of a second event happening.

- i.e., the first event provides us no insight into what will happen second.
Independence

- This requires that we define independence as follows:

  If $A_i$ is independent of $A_j$, then we have:
  \[
  Pr(A_i | A_j) = Pr(A_i)
  \]

- Why is this? It follows from the definition of conditional prob.:

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

- This in turn produces the following relation for independent events:

  \[
  Pr(A_i \cap A_j) = Pr(A_i)Pr(A_j)
  \]
Example of independence

• 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>
<th>( H_{1st} )</th>
<th>( T_{1st} )</th>
<th>( H_{2nd} )</th>
<th>( T_{2nd} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( H_{1st} )</td>
<td>( \Pr(H_{1st} \cap H_{2nd}) )</td>
<td>( \Pr(H_{1st} \cap T_{2nd}) )</td>
<td>( \Pr(H_{1st}) )</td>
<td>( \Pr(H_{1st} \cap H_{2nd}) )</td>
<td>( \Pr(H_{1st} \cap T_{2nd}) )</td>
<td>( \Pr(H_{1st}) )</td>
</tr>
<tr>
<td>( T_{1st} )</td>
<td>( \Pr(T_{1st} \cap H_{2nd}) )</td>
<td>( \Pr(T_{1st} \cap T_{2nd}) )</td>
<td>( \Pr(T_{1st}) )</td>
<td>( \Pr(T_{1st} \cap H_{2nd}) )</td>
<td>( \Pr(T_{1st} \cap T_{2nd}) )</td>
<td>( \Pr(T_{1st}) )</td>
</tr>
</tbody>
</table>

\[
\begin{array}{ccc}
H_{1st} & 0.25 & 0.25 & 0.5 \\
T_{1st} & 0.25 & 0.25 & 0.5 \\
\end{array}
\]

In this model, \( H_{1st} \) and \( H_{2nd} \) are independent, i.e. \( \Pr(H_{1st} \cap H_{2nd}) = \Pr(H_{1st})\Pr(H_{2nd}) \)
### Example of non-independence

- Consider the sample space of “two coin flips” and the following probability model:

<table>
<thead>
<tr>
<th></th>
<th>$H_{2nd}$</th>
<th>$T_{2nd}$</th>
<th>$Pr(H_{1st})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_{1st}$</td>
<td>$Pr(H_{1st} \cap H_{2nd})$</td>
<td>$Pr(H_{1st} \cap T_{2nd})$</td>
<td>$Pr(H_{1st})$</td>
</tr>
<tr>
<td>$T_{1st}$</td>
<td>$Pr(T_{1st} \cap H_{2nd})$</td>
<td>$Pr(T_{1st} \cap T_{2nd})$</td>
<td>$Pr(T_{1st})$</td>
</tr>
<tr>
<td></td>
<td>$Pr(H_{2nd})$</td>
<td>$Pr(T_{2nd})$</td>
<td></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>0.4</td>
<td>0.1</td>
</tr>
<tr>
<td>$T_{1st}$</td>
<td>0.1</td>
<td>0.4</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.5</td>
</tr>
</tbody>
</table>

In this model $H_{1st}$ and $H_{2nd}$ are not independent, i.e. $Pr(H_{1st} \cap H_{2nd}) \neq Pr(H_{1st})Pr(H_{2nd})$
Next Essential Concept: Random Variables
Next Essential Concept: Random Variables

\[ X = x, \ Pr(X) \]

\[ X = \text{Random Variable} \]

\[ X(\omega), \omega \in \Omega \]

\[ Pr(\mathcal{F}) \]

Experiment

\[ \Omega \]

(Sample Space)

\[ \mathcal{F} \]

(Sigma Algebra)
A probability function / measure takes the Sigma Algebra to the reals and provides a model of the uncertainty in our system / experiment:

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

When we define a probability function, this is an assumption (!!!), i.e. what we believe is an appropriate probabilistic description of our system / experiment.

We would like to have a concept that connects the actual outcomes of our experiment to this probability model.

What’s more, we are often in situations where we are interested in outcomes that are a function of the original sample space.

For example, “Heads” and “Tails” accurately represent the outcomes of a coin flip example but they are not numbers (and therefore have no intrinsic ordering, etc.).

We will define a random variable for this purpose.

In general, the concept of a random variable is a “bridging” concept between the actual experiment and the probability model, this provides a numeric description of sample outcomes that can be defined many ways (i.e. this provides great versatility).
Random variables II

- **Random variable** - a real valued function on the sample space:

\[ X : \Omega \rightarrow \mathbb{R} \]

- Intuitively:

\[ \Omega \rightarrow [X(\omega), \omega \in \Omega] \rightarrow \mathbb{R} \]

- Note that these functions are not constrained by the axioms of probability, e.g. not constrained to be between zero or one (although they must be measurable functions and admit a probability distribution on the random variable!!)

- We generally define them in a manner that captures information that is of interest

- As an example, let’s define a random variable for the sample space of the “two coin flip” experiment that maps each sample outcome to the “number of Tails” of the outcome:

\[ X(HH) = 0, X(HT) = 1, X(TH) = 1, X(TT) = 2 \]
Random variables III

• Why we might want a concept like $X$:

  • This approach allows us to handle non-numeric and numeric sample spaces (sets) in the same framework (e.g., $\{H, T\}$ is non-numeric but a random variable maps them to something numeric)

  • We often want to define several random variables on the same sample space (e.g., for a “two coin flips” experiment “number of heads” and “number of heads on the first of the two flips”):

    $\begin{align*}
    X_1 : \Omega &\rightarrow \mathbb{R} \\
    X_2 : \Omega &\rightarrow \mathbb{R}
    \end{align*}$

• A random variable provides a bridge between the abstract sample space that is mapped by $X$ and the actual outcomes of the experiment that we run (the sample), which produces specific numbers $x$

• As an example, the notation $X = x$ bridges the abstract notion of what values could occur $X$ and values we actually measured $x$
Random variables IV

- A critical point to note: because we have defined a probability function on the sigma algebra, this “induces” a probability function on the random variable \( X \):

\[
Pr(\mathcal{F}) \rightarrow Pr(X)
\]

- We often use an “upper” case letter to represent the function and a “lower” case letter to represent the values we actually observe:

\[
Pr(X = x)
\]

- We will divide our discussion of random variables (which we will abbreviate r.v.) and the induced probability distributions into cases that are discrete (taking individual point values) or continuous (taking on values within an interval of the reals), since these have slightly different properties (but the same foundation is used to define both!!)
Next Essential Concept: Random Variables

\[ X = x, \Pr(X) \]

**Random Variable**

- \( X(\omega), \omega \in \Omega \)
- \( \Pr(\mathcal{F}) \)

**Experiment**

- \( \mathcal{X} \)
- \( \Omega \) (Sample Space)
- \( \mathcal{F} \) (Sigma Algebra)
Discrete random variables / probability mass functions (pmf)

- If we define a random variable on a discrete sample space, we produce a discrete random variable. For example, our two coin flip / number of Tails example:

\[ X(HH) = 0, \ X(HT) = 1, \ X(TH) = 1, \ X(TT) = 2 \]

- The probability function in this case will induce a probability distribution that we call a **probability mass function** which we will abbreviate as pmf.

- For our example, if we consider a fair coin probability model (assumption!) for our two coin flip experiment and define a “number of Tails” r.v., we induce the following pmf:

\[ Pr(HH) = Pr(HT) = Pr(TH) = Pr(TT) = 0.25 \]

\[ P_X(x) = Pr(X = x) = \begin{cases} 
Pr(X = 0) = 0.25 \\
Pr(X = 1) = 0.5 \\
Pr(X = 2) = 0.25 
\end{cases} \]
Discrete random variables / cumulative mass functions (cmf)

- An alternative (and important!) representation of a discrete probability model is a cumulative mass function which we will abbreviate (cmf):

\[ F_X(x) = Pr(X \leq x) \]

where we define this function for \( X \) from \(-\infty\) to \(+\infty\).

- This definition is not particularly intuitive, so it is often helpful to consider a graph illustration. For example, for our two coin flip / fair coin / number of Tails example:
Continuous random variables / probability density functions (pdf)

- For a continuous sample space, we can define a discrete random variable or a continuous random variable (or a mixture!)
- For continuous random variables, we will define analogous “probability” and “cumulative” functions, although these will have different properties
- For this class, we are considering only one continuous sample space: the reals (or more generally the multidimensional Euclidean space)
- Recall that we will use the reals as a convenient approximation to the true sample space
- For the reals, we define a probability density function (pdf): $f_X(x)$
- A pdf defines the probability of an interval of a random variable, i.e. the probability that the random variable will take a value in that interval

$$Pr(a \leq X \leq b) = \int_a^b f_X(x)dx$$
Probability density functions (pdf): normal example

- To illustrate the concept of a pdf, let’s consider the reals as the (approximate!) sample space of human heights, the normal (also called Gaussian) probability function as a probability model for human heights, and the random variable $X$ that takes the value “height” (what kind of function is this!?)

- In this case, the pdf of $X$ has the following form: 
  \[ f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \]
Continuous random variables / cumulative density functions (cdf)

- For continuous random variables, we also have an analog to the cmf, which is the cumulative density function abbreviated as cdf:

  \[ F_X(x) = \int_{-\infty}^{x} f_X(x) \, dx \]

- Again, a graph illustration is instructive

- Note the cdf runs from zero to one (why is this?)
Mathematical properties of continuous r.v.'s

- The pdf of $X$, a continuous r.v., does not represent the probability of a specific value of $X$, rather we can use it to find the probability that a value of $X$ falls in an interval $[a,b]$:

  $$Pr(a \leq X \leq b) = \int_a^b f_X(x) \, dx$$

- Related to this concept, for a continuous random variable, the probability of specific value (or point) is zero (why is this!?)

- For a specific continuous distribution the cdf is unique but the pdf is not, since we can assign values to non-measurable sets

- If this is the case, how would we ever get a specific value when performing an experiment!?
That’s it for today

• Next lecture, we will introduce random vectors, expectations, variances, covariances, and begin our discussion of parameterized probability models.