Quantitative Genomics and Genetics

BioCB 4830/6830; PBSB.5201.03

Lecture 5: Random Variables/Vectors & Expectations/Variances

Jason Mezey

Feb 6, 2024 (T) 8:40-9:55

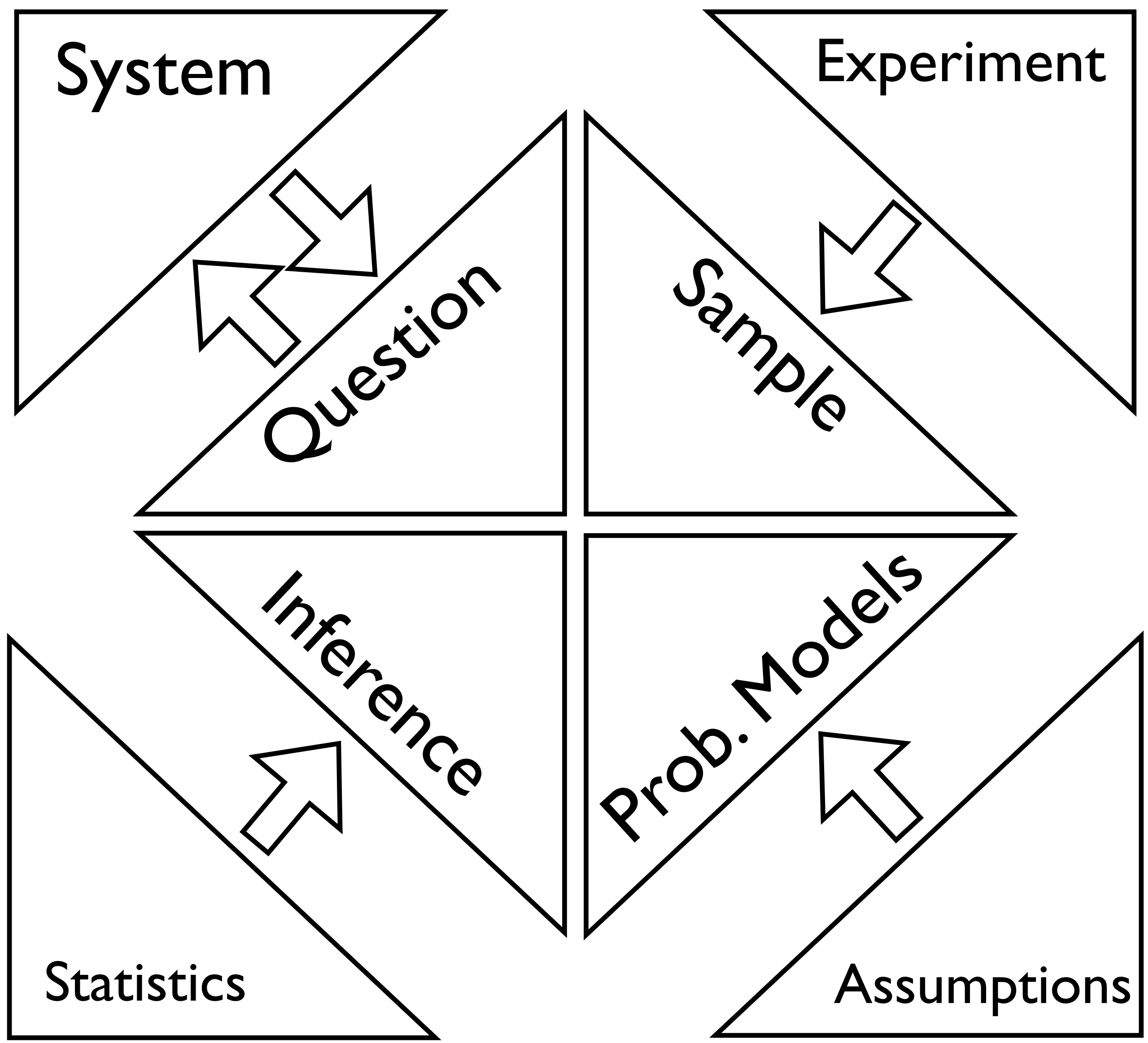
Announcements

- Registering for labs in Ithaca - everyone should be registered for EITHER Thurs or Fri lab, if not - PLEASE CANVAS EMAIL ME ASAP
- For Weill students - please continue to join by zoom this week and next unless you hear otherwise (!!)
- For Weill / NYC students - Weill recess (Feb 26-March 1) does not correspond to Ithaca spring break (April 1-5) - we will (likely) re-schedule lab and we will not have homework due during Weill recess (lectures will be recorded…
- Reminder: your 1st homework is due Fri (Feb 9) 11:59PM (!!) - note that after today's lecture, you will have everything you need to complete your homework!
- Office hours this week = we are moving them 11AM-1PM on Weds (Feb 7) - note we may move them any given week but I will always send an email / announce when these will be ahead of time!

Summary of lecture 5: Introduction to random vectors AND expectations / variances

- Last class, we introduced random variables
- Today we will continue our discussion of random variables and introduce random vectors!
- We will also begin our discussion of expectation, variances, and related

Conceptual Overview



Review: Experiments and Outcomes

- **Experiment -** a manipulation or measurement of a system that produces an outcome we can observe
- **Experiment Outcome -** a possible result of the experiment
- Example (Experiment / Outcomes):
  - Coin flip / "Heads" or "Tails"
  - Two coin flips / HH, HT, TH, TT
  - Measure heights in this class / 1.5m, 1.71m, 1.85m, …

Review: Sample Spaces

- **Sample Space** (Ω) - set comprising all possible outcomes associated with an experiment
- (Note: we have not defined a **Sample** - we will do this later!)
- 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: Sigma Algebra

- **Sigma Algebra** ($\mathcal{F}$) - a collection of events (subsets) of Ω of interest with the following three properties: **1**. $\emptyset \in \mathcal{F}$, **2**. $\mathcal{A} \in \mathcal{F} \text{ then } \mathcal{A}^c \in \mathcal{F}$, **3**. $\mathcal{A}_1, \mathcal{A}_2, ... \in \mathcal{F} \text{ then } \bigcup_{i=1}^{\infty} \mathcal{A}_i \in \mathcal{F}$

  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} /

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

  - $\mathbb{R}$ / more complicated to define the sigma algebra of interest (see next slide…)

Review: Probability functions I

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

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

- 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 $\mathcal{A} \subset \Omega, Pr(\mathcal{A}) \geqslant 0$
2. $Pr(\Omega) = 1$
3. For $\mathcal{A}_1, \mathcal{A}_2, ... \subset \Omega$, if $\mathcal{A}_i \cap \mathcal{A}_j = \emptyset$ (disjoint) for each $i \neq j$: $Pr(\bigcup_i^{\infty} \mathcal{A}_i) = \sum_i^{\infty} Pr(\mathcal{A}_i)$

- 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*

Review: 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 $\mathcal{A}_i$ given $\mathcal{A}_j$ is:

$$Pr(\mathcal{A}_i|\mathcal{A}_j) = \frac{Pr(\mathcal{A}_i \bigcap \mathcal{A}_j)}{Pr(\mathcal{A}_j)}$$

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

Review: Independence

- This requires that we define independence as follows:

If $\mathcal{A}_i$ is independent of $\mathcal{A}_j$, then we have:

$$Pr(\mathcal{A}_i|\mathcal{A}_j) = Pr(\mathcal{A}_i)$$

- This implies the following from the definition of conditional prob.:

$$Pr(\mathcal{A}_i|\mathcal{A}_j) = \frac{Pr(\mathcal{A}_i \bigcap \mathcal{A}_j)}{Pr(\mathcal{A}_j)} = \frac{Pr(\mathcal{A}_i)Pr(\mathcal{A}_j)}{Pr(\mathcal{A}_j)} = Pr(\mathcal{A}_i)$$

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

$$Pr(\mathcal{A}_i \cap \mathcal{A}_j) = Pr(\mathcal{A}_i)Pr(\mathcal{A}_j)$$

Review: Random variables I

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

$$X : \Omega \rightarrow \mathbb{R}$$

- Intuitively:

$$\Omega \longrightarrow \boxed{X(\omega), \omega \in \Omega} \longrightarrow \mathbb{R}$$

- 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:

$$X(HH) = 0, X(HT) = 1, X(TH) = 1, X(TT) = 2$$

Review: 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"):

$$X_1 : \Omega \to \mathbb{R}$$
$$X_2 : \Omega \to \mathbb{R}$$

$$\Omega \begin{array}{l} \longrightarrow X_1 \\ \longrightarrow X_2 \end{array}$$

  - 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*

Review: Random variables III

- 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*:

$$Pr : \mathcal{F} \rightarrow [0,1] \Rightarrow Pr : X \rightarrow [0,1]$$

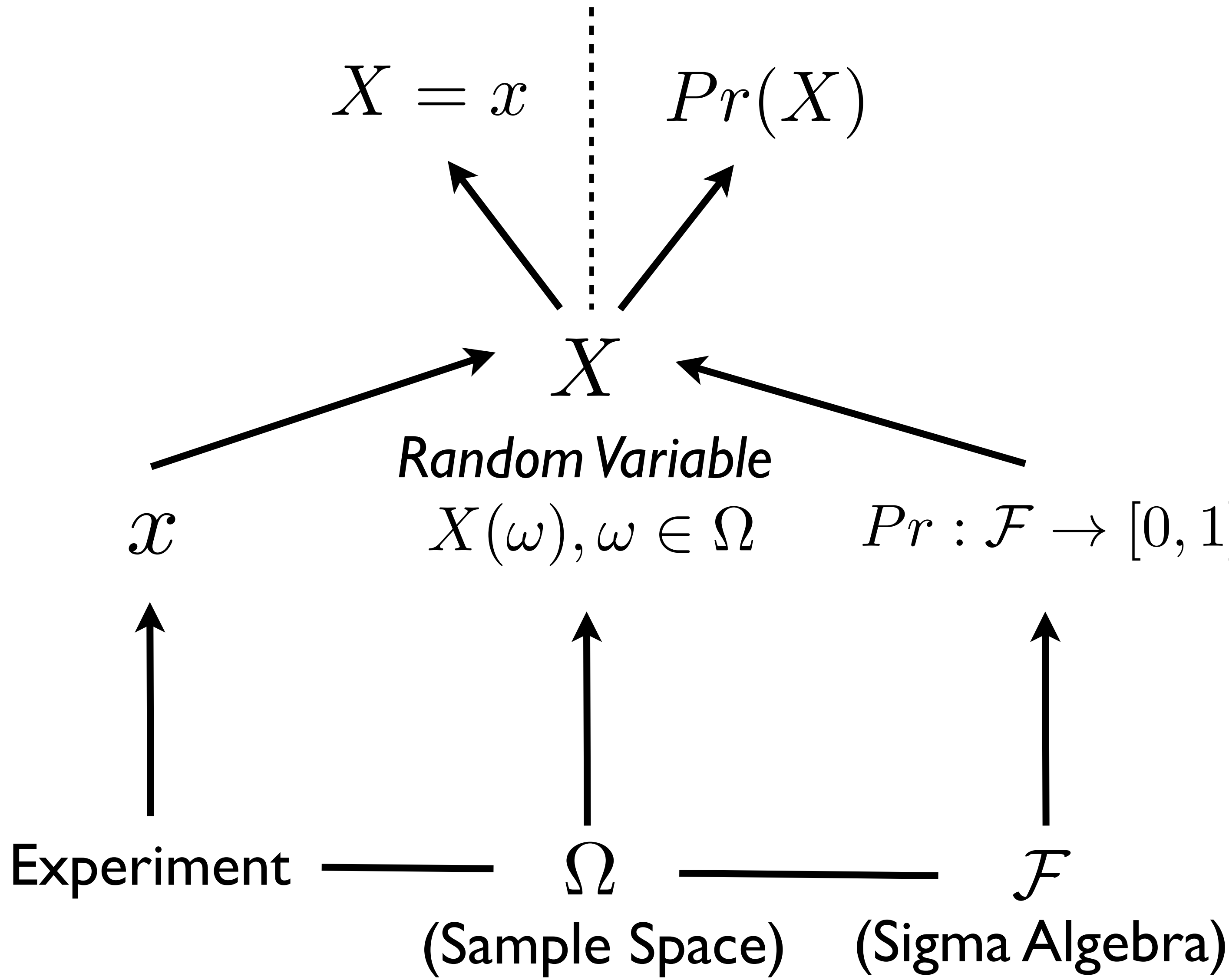
- In fact, this relationship allows us to "start" our modeling with the random variable and the probability on this random variable (i.e. the Sample Space, Sigma Algebra, and original probability function on random variable are implicit - but remember these foundations are always there!!)

- To bridge probability of an occurrence and what actually occurs in the experiment e 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!!)

Random Variables



Review: Discrete vs Continuous Random Variables

- There are TWO broad categories of random variables: "Discrete" and "Continuous"
  - If the values the random variable can take can be "counted" then the random variable is DISCRETE
  - If the values the random variable cannot be "counted" (e.g., the random variable can take any values on the REALs) then the random variable is CONTINUOUS
- We need to treat the (mathematical) mechanics of these two categories differently…
- *Technical points: (A) discrete random variables may be finite or infinite as long as they take "countable" states (e.g., the naturals are countable while the reals are uncountable), (B) a continuous random variable can only be defined on an uncountable sample space (usually the reals), but a discrete (or mixed) random variable may be defined in a continuous sample space*

Review: 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:

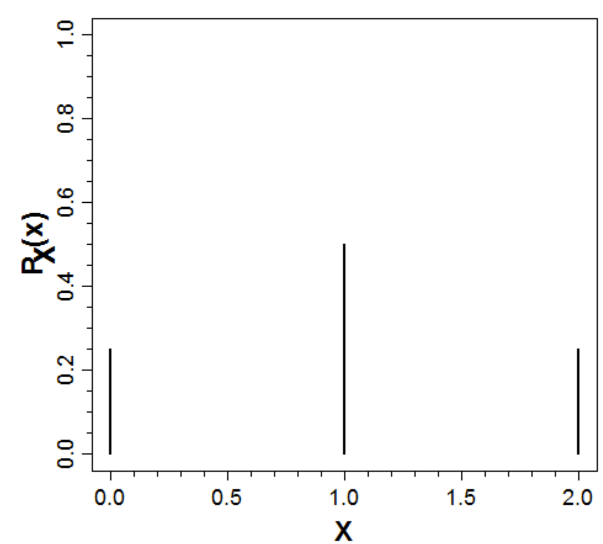
$$X(HH) = 0, X(HT) = 1, X(TH) = 1, X(TT) = 2$$

- 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:

$$Pr(\{HH\}) = Pr(\{HT\}) = Pr(\{TH\}) = Pr(\{TT\}) = 0.25$$

$$P_X(x) = Pr(X = x) = \begin{cases} Pr(X = 0) = 0.25 \\ Pr(X = 1) = 0.5 \\ Pr(X = 2) = 0.25 \end{cases}$$



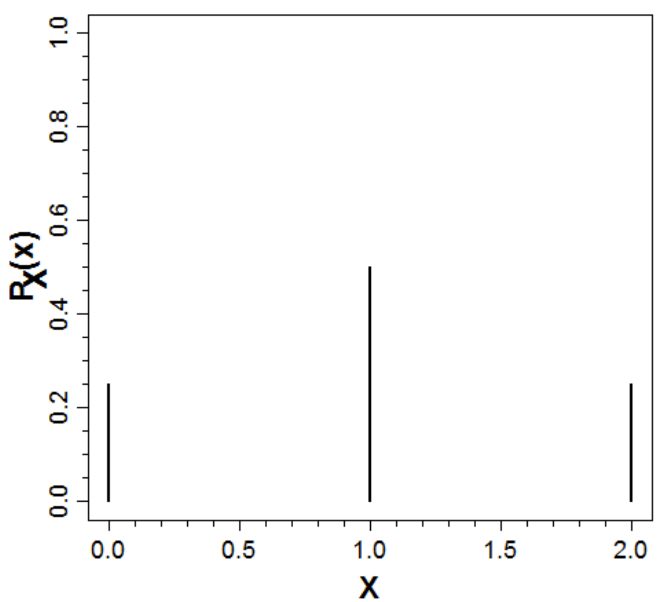
Review: Discrete random variables / cumulative mass functions (cmf)

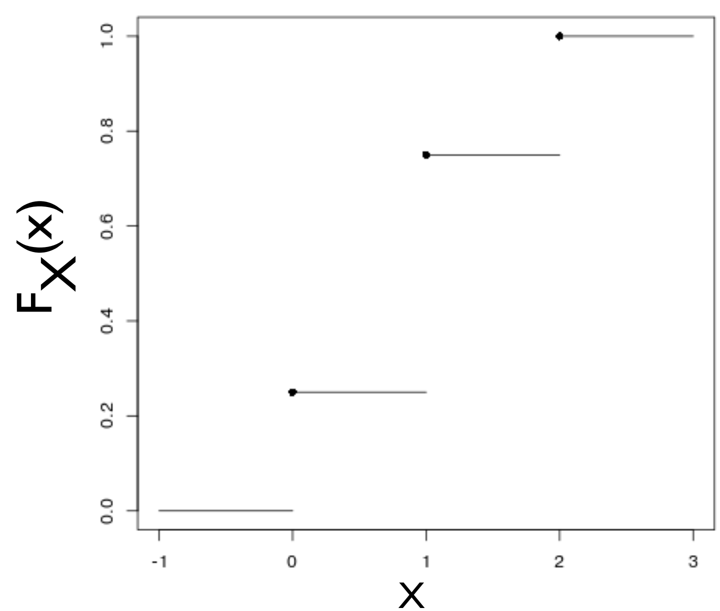
- An alternative (and important!) representation of a discrete probability model is a **cumulative mass function** which we will abbreviate (cmf):

$$F_X(x) = Pr(X \leqslant x)$$

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

Mathematical properties of continuous r.v.'s

- For the reals, we define a probability density function (pdf): $f_X(x)$
- 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]:

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

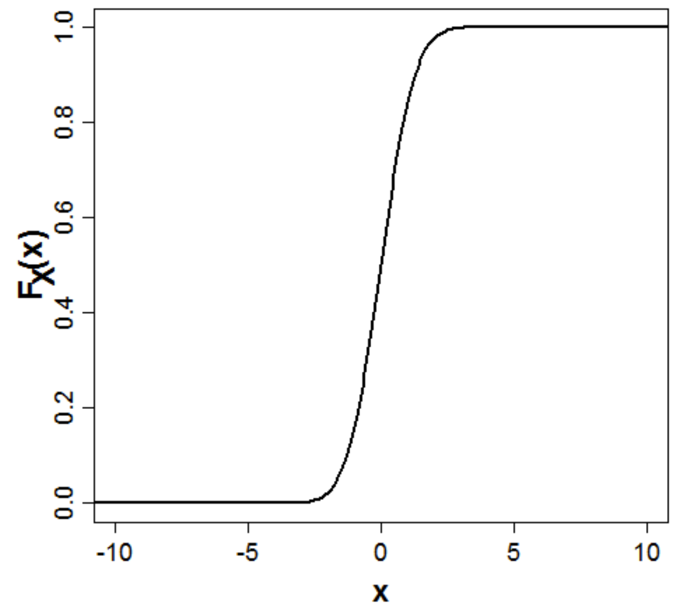
- Related to this concept, for a continuous random variable, the probability of specific value (or point) is zero (why is this!?)
- For a specific continuous distribution the cdf is unique but the pdf is not, since we can assign values to non-measurable sets
- If this is the case, how would we ever get a specific value when performing an experiment!?

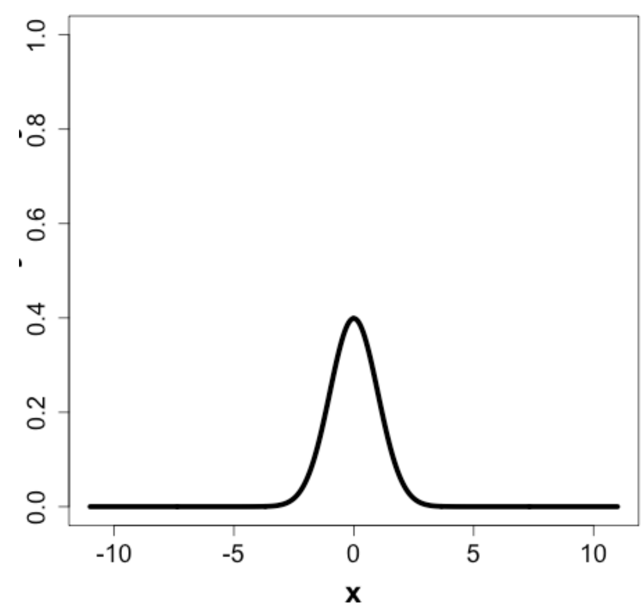
Continuous random variables / cumulative density functions (cdf)

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

$$F_X(x) = \int_{-\infty}^{x} f_X(x)dx$$

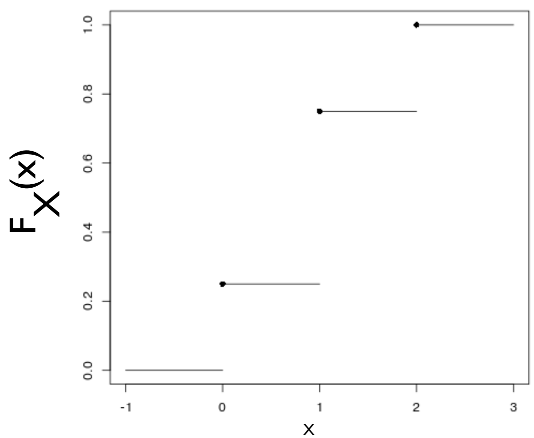
- Again, a graph illustration is instructive
- Note the cdf runs from zero to one (why is this?)

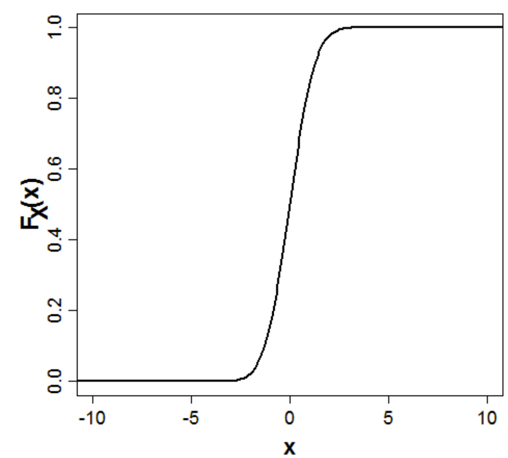




Random variables V (advanced topic)

- While it will not be clear from this class why cumulative (mass or density) distributions are important (e.g., we will barely use them…) in your advanced stats classes you will use them a lot





- The reason you will see them in advanced classes is these have nice properties for mathematical purposes
  - As an example, ALL cumulative distribution functions have the SAME domain (=the reals!) and image / codomain (from zero to one!) and all monotonic increasing functions with this property represent ALL possible probability models
  - They have other nice properties (e.g., they are uniquely defined as opposed to probability mass or density functions)
- This is another reason why random variables are the central focus in many classes

Reminder: 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"):

$$X_1 : \Omega \to \mathbb{R}$$
$$X_2 : \Omega \to \mathbb{R}$$

$$\Omega \longrightarrow X_1$$
$$\Omega \longrightarrow X_2$$

  - 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 vectors

- We are often in situations where we are interested in defining more than one r.v. on the same sample space
- When we do this, we define a **random vector**
- Note that a vector, in its simplest form, may be considered a set of numbers (e.g. [1.2, 2.0, 3.3] is a vector with three elements)
- Also note that vectors (when a vector space is defined) ARE NOT REALLY NUMBERS although we can define operations for them (e.g. addition, "multiplication"), which we will use later in this course
- Beyond keeping track of multiple r.v.'s, a *random vector* works just like a r.v., i.e. a probability function induces a probability function on the random vector and we may consider discrete or continuous (or mixed!) random vectors
- Note that we can define several r.v.'s on the same sample space (= a random vector), but this will result in one probability distribution function (why!?)

Example of a discrete random vector

- Consider the two coin flip experiment and assume a probability function for a fair coin: $Pr(\{HH\}) = Pr(\{HT\}) = Pr(\{TH\}) = Pr(\{TT\}) = 0.25$

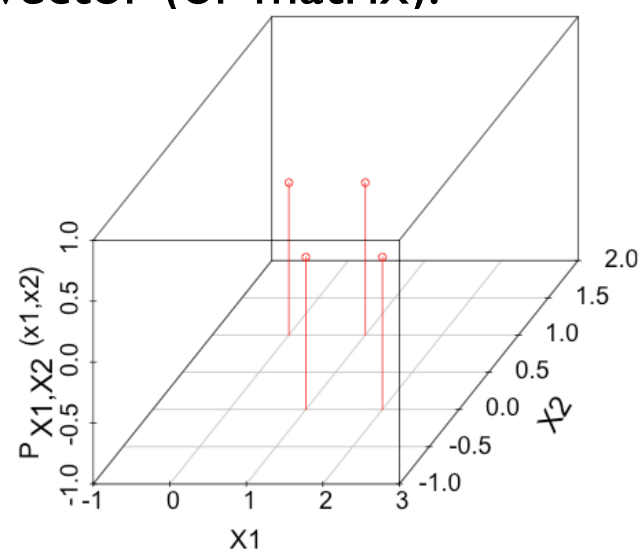
- Let's define two random variables: "number of Tails" and "first flip is Heads"

$$X_1 = \begin{cases} X_1(HH) = 0 \\ X_1(HT) = X_1(TH) = 1 \\ X_1(TT) = 2 \end{cases} \qquad X_2 = \begin{cases} X_2(TH) = X_2(TT) = 0 \\ X_2(HH) = X_2(HT) = 1 \end{cases}$$

- The probability function induces the following pmf for the random vector **X**=[X_1, X_2], where we use bold **X** do indicate a vector (or matrix):

$$Pr(\mathbf{X}) = Pr(X_1 = x_1, X_2 = x_2) = P_{\mathbf{X}}(\mathbf{x}) = P_{X_1,X_2}(x_1, x_2)$$

$$Pr(X_1 = 0, X_2 = 0) = 0.0, Pr(X_1 = 0, X_2 = 1) = 0.25$$
$$Pr(X_1 = 1, X_2 = 0) = 0.25, Pr(X_1 = 1, X_2 = 1) = 0.25$$
$$Pr(X_1 = 2, X_2 = 0) = 0.25, Pr(X_1 = 2, X_2 = 1) = 0.0$$



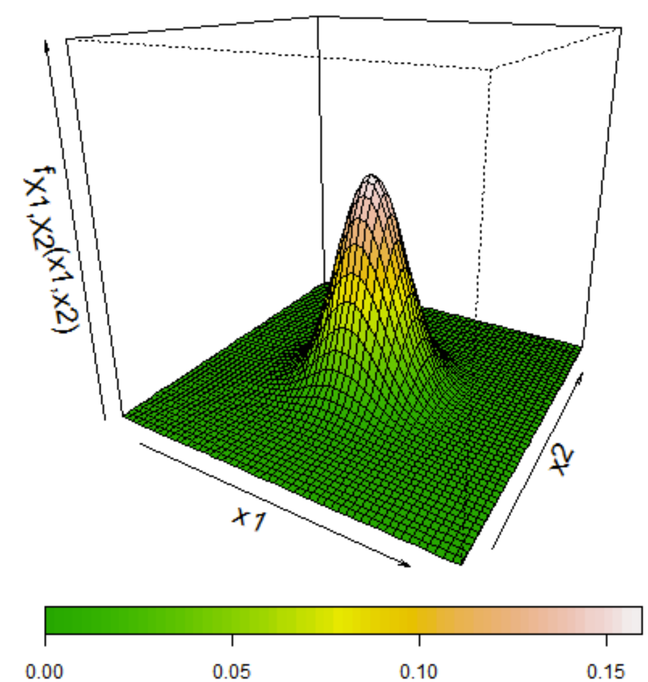
Example of a continuous random vector

- Consider an experiment where we define a two-dimensional *Reals* sample space for "height" and "IQ" for every individual in the US (as a reasonable approximation)
- Let's define a bivariate normal probability function for this sample space and random variables X_1 and X_2 that are identity functions for each of the two dimensions
- In this case, the pdf of $\mathbf{X}=[X_1, X_2]$ is a bivariate normal (we will not write out the formula for this distribution - yet):

$$Pr(\mathbf{X}) = Pr(X_1 = x_1, X_2 = x_2) = f_{\mathbf{X}}(\mathbf{x}) = f_{X_1,X_2}(x_1, x_2)$$

Again, note that we cannot use this probability function to define the probabilities of points (or lines!) but we can use it to define the probabilities that values of the random vector fall within (square) intervals of the two random variables (!) [a,b], [c,d]

$$Pr(a \leqslant X_1 \leqslant b, c \leqslant X_1 \leqslant d) = \int_a^b \int_c^d f_{X_1,X_2}(x_1, x_2) dx_1, dx_2$$



Random vector conditional probability and independence I

- Just as we have defined *conditional probability* (which are probabilities!) for sample spaces, we can define conditional probability for random vectors:

$$Pr(X_1|X_2) = \frac{Pr(X_1 \cap X_2)}{Pr(X_2)}$$

- As a simple example (discrete in this case - but continuous is analogous!), consider the two flip sample space, fair coin probability model, random variables: "number of tails" and "first flip is heads":

	$X_2 = 0$	$X_2 = 1$	
$X_1 = 0$	0.0	0.25	0.25
$X_1 = 1$	0.25	0.25	0.5
$X_1 = 2$	0.25	0.0	0.25
	0.5	0.5	

$$Pr(X_1 = 0|X_2 = 1) = \frac{Pr(X_1 = 0 \cap X_2 = 1)}{Pr(X_2 = 1)} = \frac{0.25}{0.5} = 0.5$$

- We can similarly consider whether r.v.'s of a random vector are independent, e.g.

$$Pr(X_1 = 0 \cap X_2 = 1) = 0.25 \neq Pr(X_1 = 0)Pr(X_2 = 1) = 0.25 * 0.5 = 0.125$$

- NOTE I: we can use either $Pr(X_i|X_j) = Pr(X_i)$ or $Pr(X_i \cap X_j) = Pr(X_i)Pr(X_j)$ to check independence!

- NOTE II: to establish Xi, Xj are independent you must check all possible relationships but the opposite is not true: if one does not show independence you've established they are not independent (!!)

Random vectors conditional probability and independence II

For random variables that are NOT independent…

	$X_2 = 0$	$X_2 = 1$	
$X_1 = 0$	0.0	0.25	0.25
$X_1 = 1$	0.25	0.25	0.5
$X_1 = 2$	0.25	0.0	0.25
	0.5	0.5	

To establish non-independence, just show ONE case that does not conform to the independence definition (e.g.,):

$$Pr(X_2 = 0|X_1 = 0) = \frac{Pr(X_2 = 0 \cap X_1 = 0)}{Pr(X_1 = 0)} = 0 \neq Pr(X_2 = 0) = 0.5$$

OR

$$Pr(X_2 = 0 \cap X_1 = 0) = 0 \neq Pr(X_2 = 0)Pr(X_1 = 0) = 0.5 * 0.25 = 0.125$$

And you're done!

For random variables that ARE independent…

	$X_2 = 0$	$X_2 = 1$	
$X_1 = 0$	0.125	0.125	0.25
$X_1 = 1$	0.25	0.25	0.5
$X_1 = 2$	0.125	0.125	0.25
	0.5	0.5	

To establish independence, you need to show ALL combinations of random variable states conform to the independence definition (!!):

$$Pr(X_2 = 0 \cap X_1 = 0) = Pr(X_2 = 0)Pr(X_1 = 0) = 0.5 * 0.25 = 0.125$$

$$Pr(X_2 = 0 \cap X_1 = 1) = Pr(X_2 = 0)Pr(X_1 = 1) = 0.5 * 0.5 = 0.25$$

$$Pr(X_2 = 0 \cap X_1 = 2) = Pr(X_2 = 0)Pr(X_1 = 2) = 0.5 * 0.25 = 0.125$$

$$Pr(X_2 = 1 \cap X_1 = 0) = Pr(X_2 = 1)Pr(X_1 = 0) = 0.5 * 0.25 = 0.125$$

$$Pr(X_2 = 1 \cap X_1 = 1) = Pr(X_2 = 1)Pr(X_1 = 1) = 0.5 * 0.5 = 0.25$$

$$Pr(X_2 = 1 \cap X_1 = 2) = Pr(X_2 = 1)Pr(X_1 = 2) = 0.5 * 0.25 = 0.125$$

Marginal distributions of random vectors

- Note that **marginal distributions** of random vectors are the probability of a r.v. of a random vector after summing (discrete) or integrating (continuous) over all the values of the other random variables:

$$P_{X_1}(x_1) = \sum_{x_2=min(X_2)}^{max(X_2)} Pr(X_1 = x_1 \cap X_2 = x_2) = \sum Pr(X_1 = x_1 | X_2 = x_2) Pr(X_2 = x_2)$$

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} Pr(X_1 = x_1 \cap X_2 = x_2) dx_2 = \int_{-\infty}^{\infty} Pr(X_1 = x_1 | X_2 = x_2) Pr(X_2 = x_2) dx_2$$

- Again, as a simple illustration, consider our two coin flip example:

	$X_2 = 0$	$X_2 = 1$	
$X_1 = 0$	0.0	0.25	0.25
$X_1 = 1$	0.25	0.25	0.5
$X_1 = 2$	0.25	0.0	0.25
	0.5	0.5	

Three last points about random vectors

- Just as we can define cmf's / cdf's for r.v.'s, we can do the same for random vectors:

$$F_{X_1,X_2}(x_1, x_2) = Pr(X_1 \leqslant x_1, X_2 \leqslant x_2)$$

$$F_{X_1,X_2}(x_1, x_2) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f_{X_1,X_2}(x_1, x_2) dx_1 dx_2$$

- We have been discussing random vectors with two r.v.'s, but we can consider any number *n* of r.v.'s:

$$Pr(\mathbf{X}) = Pr(X_1 = x_1, X_2 = x_2, ..., X_n = x_n)$$

- We refer to probability distributions defined over r.v. to be *univariate*, when defined over vectors with two r.v.'s they are *bivariate*, and when defined over three or more, they are *multivariate*

Expectations and variances

- We are now going to introduce fundamental functions of random variables / vectors: **expectations** and **variances**

- These are **functionals -** map a function to a scalar (number)

- These intuitively (but not rigorously!) these may be thought of as "a function on a function" with the following form:

$$f(\mathbf{X}(\Omega), Pr(\mathbf{X})) : \{\mathbf{X}, Pr(\mathbf{X})\} \rightarrow \mathbb{R}$$

- These are critical concepts for understanding the structure of probability models where the interpretation of the specific probability model under consideration

- They also have deep connections to many important concepts in probability and statistics

- Note that these are distinct from functions (*Transformations*) that are defined directly on *X* and not on Pr(*X*), i.e. *Y* = g(*X*):

$$g(\mathbf{X}) : X \rightarrow Y$$

$$g(\mathbf{X}) \rightarrow Y \Rightarrow Pr(X) \rightarrow Pr(Y)$$

Expectations I

- Following our analogous treatment of concepts for *discrete* and *continuous* random variables, we will do the same for *expectations* (and variances), which we also call *expected values*

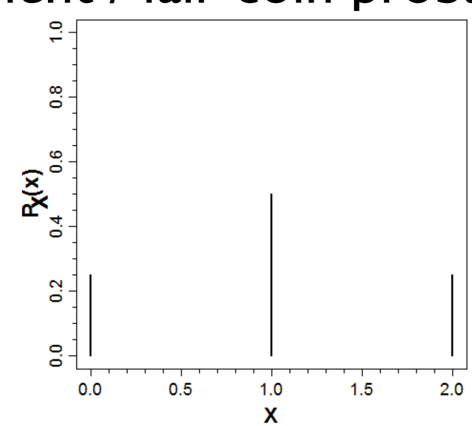
- Note that the interpretation of the expected value is the same in each case

- The expected value of a discrete random variable is defined as follows:

$$\mathrm{E}X = \sum_{i=min(X)}^{max(X)} (X = i)Pr(X = i)$$

- For example, consider our two-coin flip experiment / fair coin probability model / random variable "number of tails":

$$\mathrm{E}X = (0)(0.25) + (1)(0.5) + (2)(0.25) = 1$$

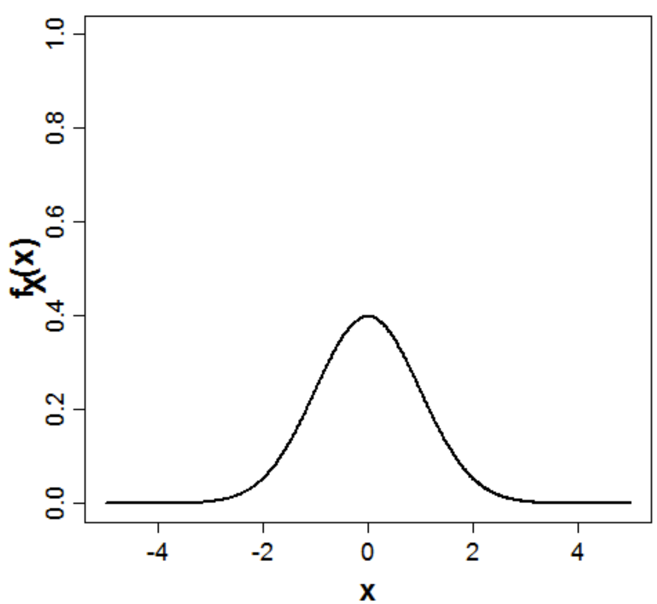


Expectations II

- The expected value of a continuous random variable is defined as follows:

$$\mathrm{E}X = \int_{-\infty}^{+\infty} X f_X(x) dx$$

- For example, consider our height measurement experiment / normal probability model / identity random variable:



Expectations III

- In the discrete case, this is the *same* as adding up all the possibilities that can occur and dividing by the total number, e.g. (0+1+1+2) / 4 = 1 (hence it is often referred to as the *mean* of the random variable
- An expected value may be thought of as the "center of gravity", where a median (defined as the number where half of the probability is on either side) is the "middle" of the distribution (note that for symmetric distributions, these two are the same!)
- The expectation of a random variable *X* is the value of *X* that minimizes the sum of the squared distance to each possibility
- For some distributions, the expectation of the random variable may be infinite. In such cases, the expectation does not exist

That's it for today

- Next lecture, we will continue our discussion of expectations, variances, and related!