Quantitative Genomics and Genetics

BTRY 4830/6830; PBSB.52201.03

Lecture 4: Random Variables and Random Vectors

Jason Mezey

Feb 2, 2023 (Th) 8:05-9:20

Announcements I

- **FIRST COMPUTER LAB IS TOMORROW / FRI (Thurs. Feb 2 / Fri. Feb 3) - more information to come this week!**
- PLEASE NOTE THE LAB TIMES (!!)
- For those IN ITHACA (= Labs with Mitch!):
  - Lab 1: 5:30-6:30PM on Thurs. (Weill Hall 226)
  - Lab 2: 8-9AM on Fri. (Weill Hall 226)
  - Please go to the Lab you registered for (!!)
- For those IN NYC (= Labs taught by Sam!):
  - Lab 1: 4-5PM on Thurs. (In WCMC1300 Classroom; G [B215], H [B217])
  - Lab 2: 9-10AM on Fri. (By zoom - Sam will distribute the invite)
  - PLEASE NOTE: if you are in HOUSTON or you are VERY EXPERIENCED with R, please join Fri (!!) - otherwise, join on Thurs!
- You may skip the first 2 labs without penalty BUT
  - If you are not VERY familiar with R programming you may want to go
  - If you do not already use Latex you may want to go (e.g., homeworks!)

Announcements II

- Everybody should be signed up on Piazza (!!) - note we may need to deal with advertisements…
- Check the lab website: https://mezeylab.biohpc.cornell.edu
- Class CMS is up (!!): https://cmsx.cs.cornell.edu/web/guest/
  - If you have a NetID you should be able to access
  - If you have a CWID (i.e., you are at Weill) we hope to have you on by this afternoon
  - We will post homework #1 on CMS today (!!)
  - If we cannot get Weill folks up on CMS by the time we post homework #1, we will distribute in another way…

Announcements III

- Homework #1 (PLEASE NOTE THE FOLLOWING):
  - Due 11:59PM, Weds., Feb 8 and MUST BE UPLOADED TO CMS (!!)
  - If you upload late (even by a minute…) you will get a penalty (note that no excuses will be accepted = you can always upload early…)
  - Homeworks are "open book" and you may work together but hand in your own work (!!)
  - Answers must be typed (!!) including all equations - if this is a problem go to computer lab this week (= intro to Latex!)
- Problems are divided into "easy", "medium, and "difficult"
  - You can complete the "easy" and "medium" (make sure you give yourself enough time!)
  - For the "difficult" at least attempt (but note that you can get an "A" in the class even if you do not / cannot complete these problems!)
- Please feel free to attend office hours for help (!!) see next slide

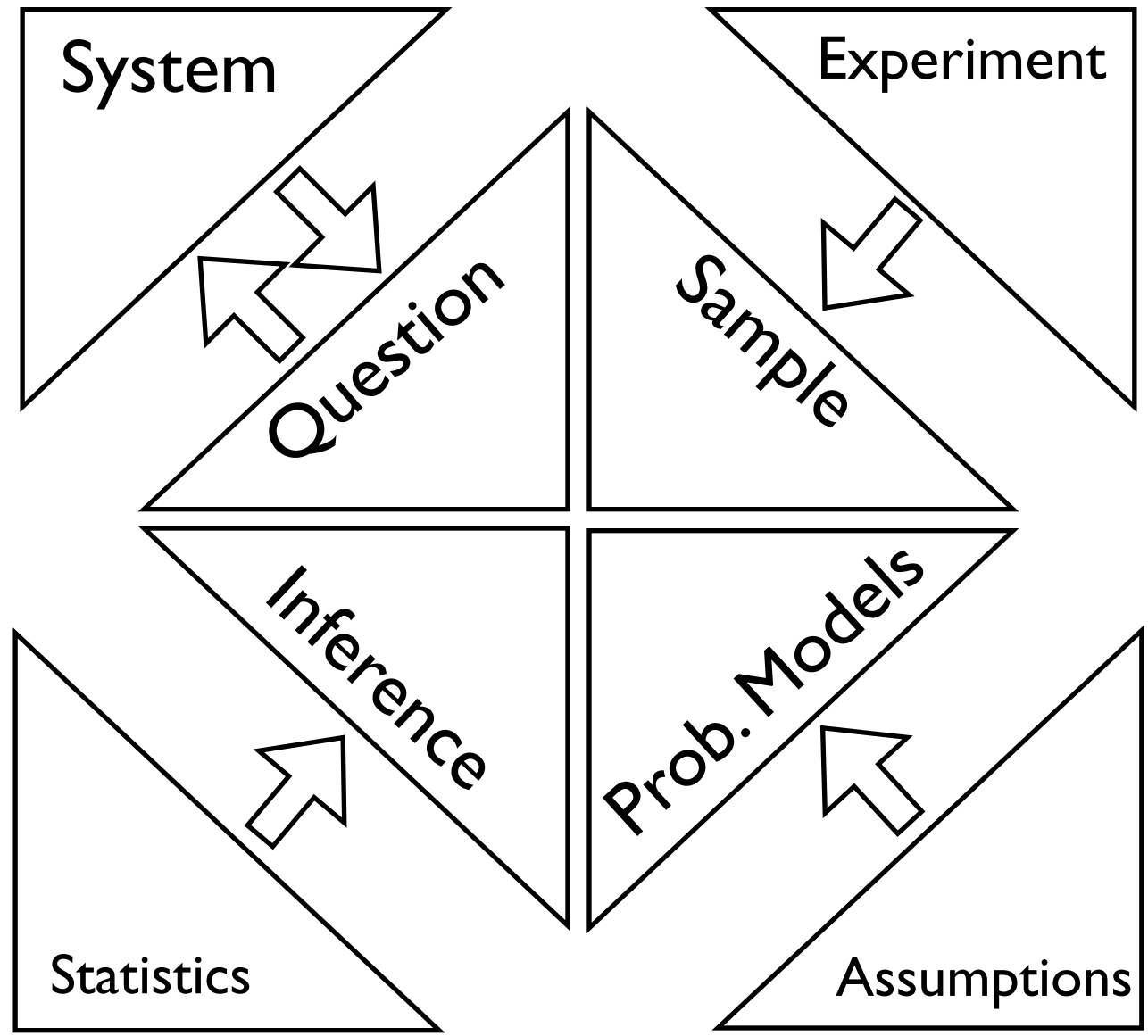
Announcements IV

- I will hold office hours on MONDAYS every week 12:30-2:30 by zoom (please note: if this day / time turns out to be inconvenient for many, we may change…)
- I will send out a Piazza message with the zoom link for office hours later today (please do not share beyond the class!)
- The first office hours will be this Mon, Feb 6 (I will send out a reminder)
- I will record office hours (and post them on CMS)
- You may also set up individual sessions with me (Jason) by appointment

Summary of lecture 4: Introduction to random variables and vectors

- Last class, we introduced conditional probability (and independence)
- Today we will introduce and discuss a critical concept (!!) random variables (and random vectors)

Conceptual Overview



Review: Probability functions 1

- **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 $\mathcal{A} \subset \Omega, Pr(\mathcal{A}) \geqslant 0$
2. $Pr(\Omega) = 1$
3. For $\mathcal{A}_1, \mathcal{A}_2, ... \in \Omega$, 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: 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):

	H_{2nd}	T_{2nd}
H_{1st}	HH	HT
T_{1st}	TH	TT

	H_{2nd}	T_{2nd}	
H_{1st}	0.25	0.25	0.5
T_{1st}	0.25	0.25	0.5
	0.5	0.5	

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

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: Example of independence

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

	H_{2nd}	T_{2nd}	
H_{1st}	$Pr(H_{1st} \cap H_{2nd})$	$Pr(H_{1st} \cap T_{2nd})$	$Pr(H_{1st})$
T_{1st}	$Pr(T_{1st} \cap H_{2nd})$	$Pr(T_{1st} \cap T_{2nd})$	$Pr(T_{1st})$
	$Pr(H_{2nd})$	$Pr(T_{2nd})$	

	H_{2nd}	T_{2nd}	
H_{1st}	0.25	0.25	0.5
T_{1st}	0.25	0.25	0.5
	0.5	0.5	

In this model, H_{1st} and H_{2nd} are independent, i.e. $Pr(H_{1st} \cap H_{2nd}) = Pr(H_{1st})Pr(H_{2nd})$

Review: Example of non-independence

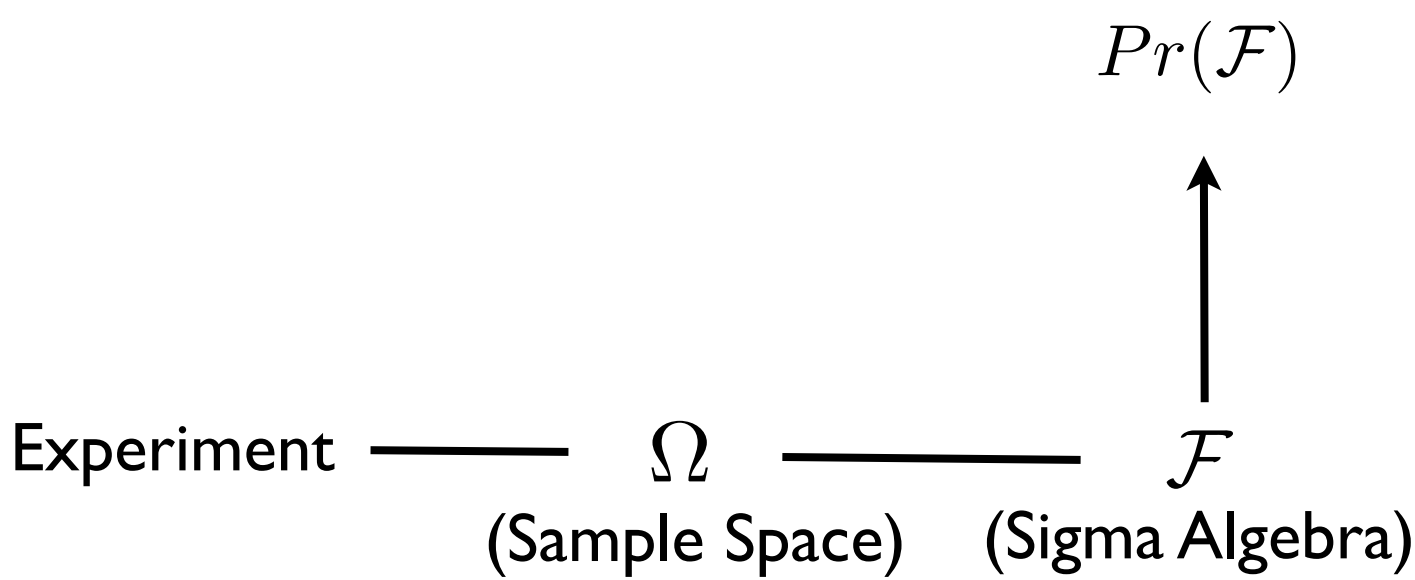
- Consider the sample space of "two coin flips" and the following probability model:

	H_{2nd}	T_{2nd}	
H_{1st}	$Pr(H_{1st} \cap H_{2nd})$	$Pr(H_{1st} \cap T_{2nd})$	$Pr(H_{1st})$
T_{1st}	$Pr(T_{1st} \cap H_{2nd})$	$Pr(T_{1st} \cap T_{2nd})$	$Pr(T_{1st})$
	$Pr(H_{2nd})$	$Pr(T_{2nd})$	

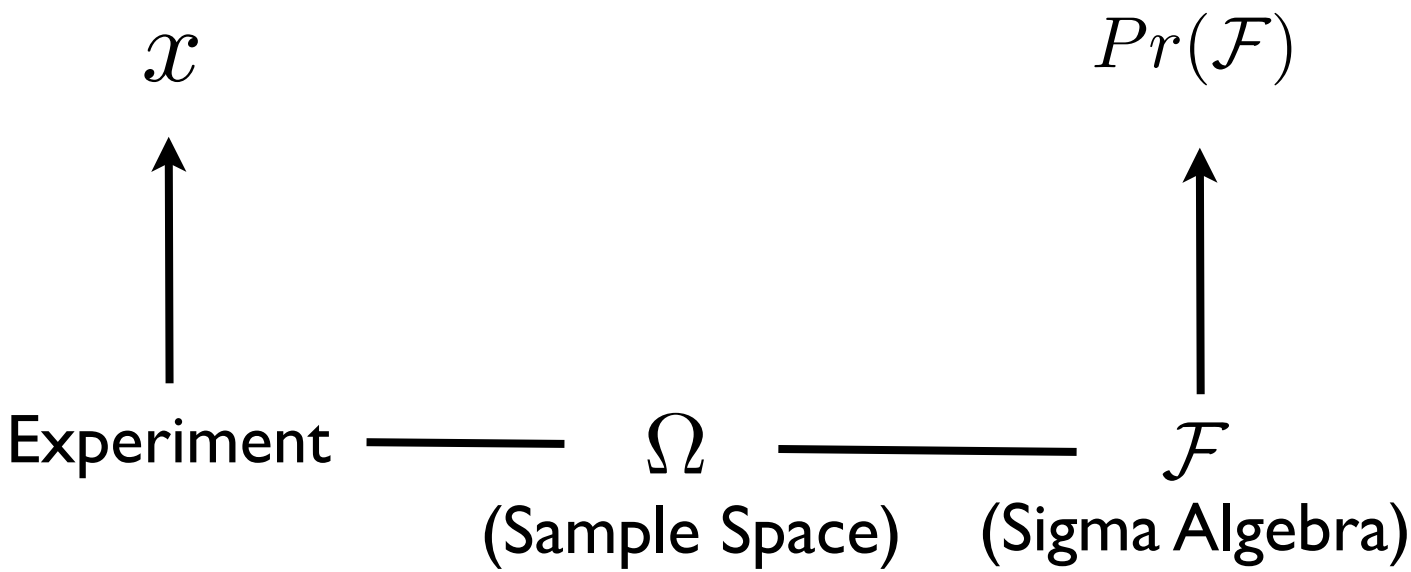
	H_{2nd}	T_{2nd}	
H_{1st}	0.4	0.1	0.5
T_{1st}	0.1	0.4	0.5
	0.5	0.5	

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

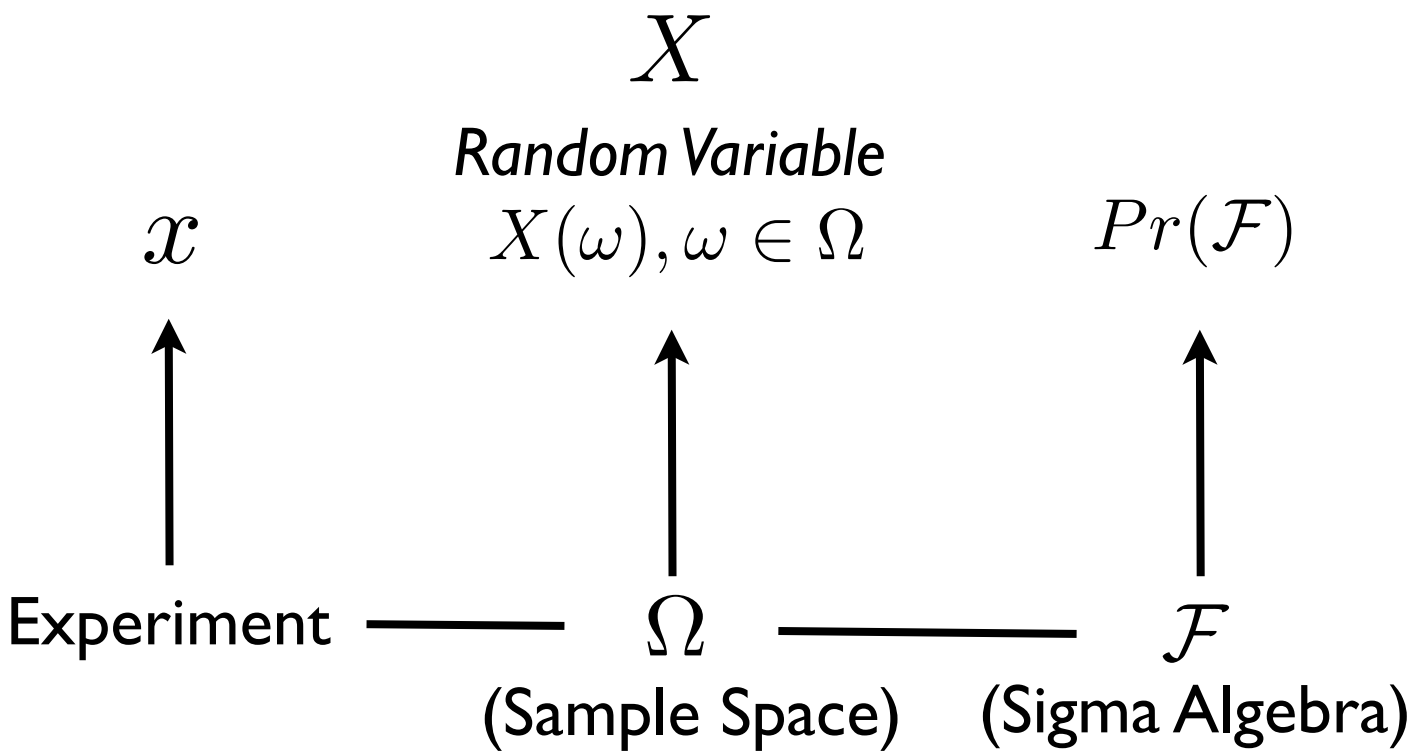
Next Essential Concept: Random Variables



Next Essential Concept: Random Variables



Next Essential Concept: Random Variables



Random variables I

- 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 / experimen

- We would like to have a concept that connects the *actual* outcomes of our experiment to this probability mode

- What's more, we are often in situations where we are interested in using numbers to represent the outcomes, e.g., , "Heads" and "Tails" accurately represent the outcomes of a coin flip example but they are not numbers (e.g., we may be interested in "number of heads")

- In addition, many of the mathematical tools we use in probability and statistics require the outcomes being represented within the reals

- We therefore are often interested in a function of the original sample space that maps this space to the reals

- 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. provides great versatility)

Random variables II

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

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

- 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

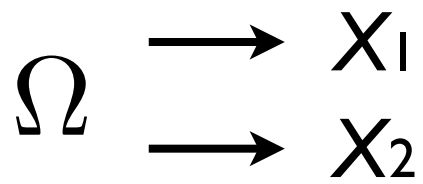
- As an example, let's define a random variable for the sample space of the "two coin flip" experiment that maps each sample outcome to the "number of Tails" of the outcome:

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

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"):

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



  - 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*:

$$Pr(\mathcal{F}) \longrightarrow Pr(X)$$

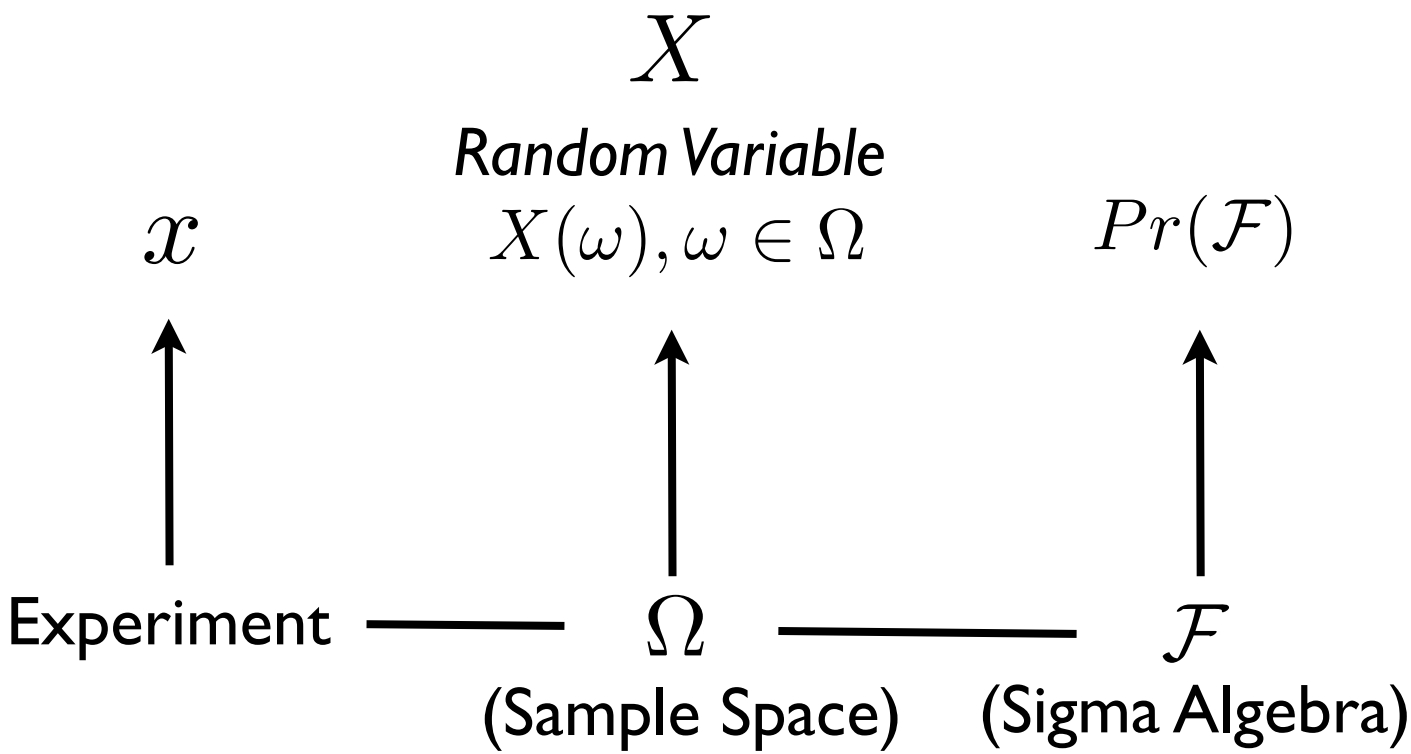
- In fact, this relationship allows us to "start" our modeling with the random variable and the probability on this random variable (i.e. the Sample Space, Sigma Algebra, and original probability function on random variable are implicit - but remember these foundations are always there!!)

- 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:

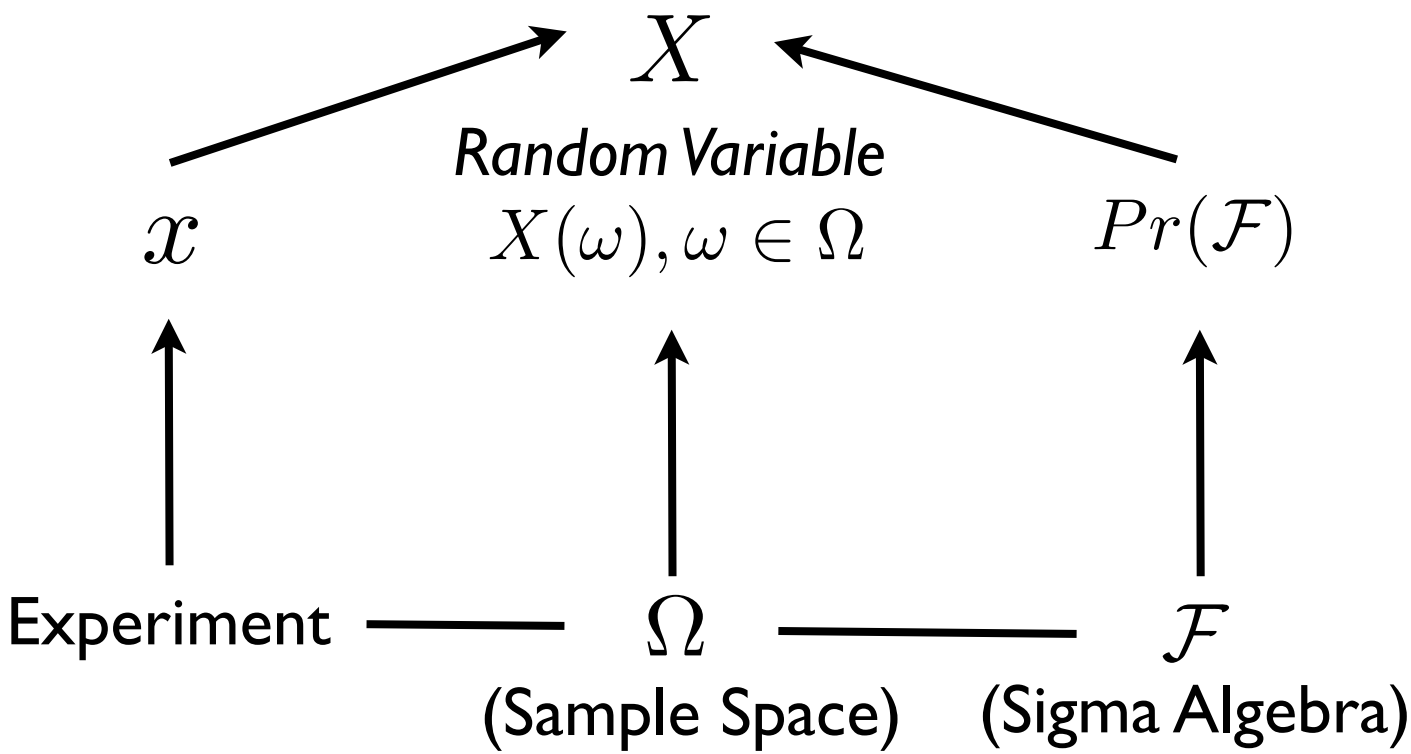
$$Pr(X = x)$$

- 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!!)

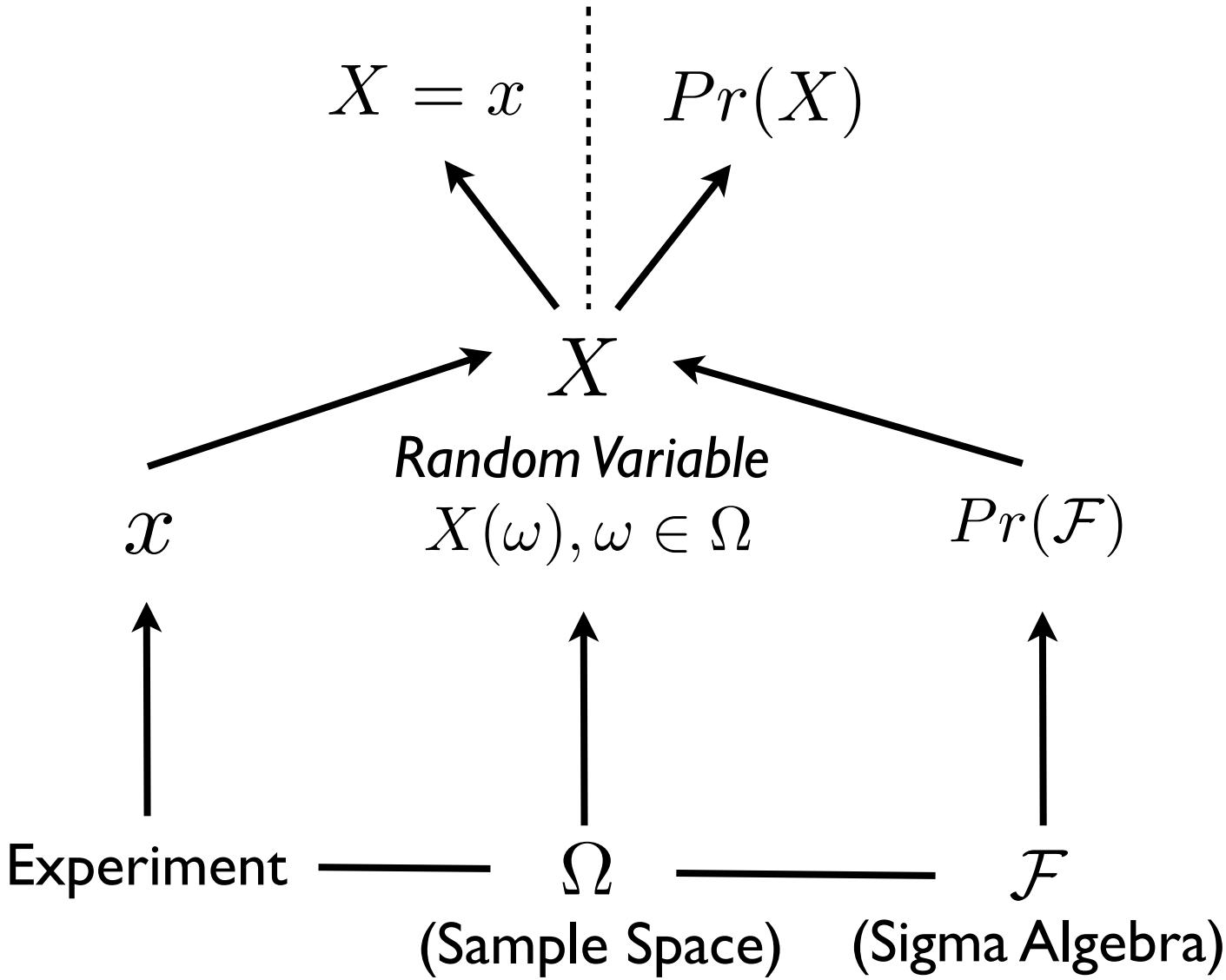
Next Essential Concept: Random Variables



Random Variables



Random Variables



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:

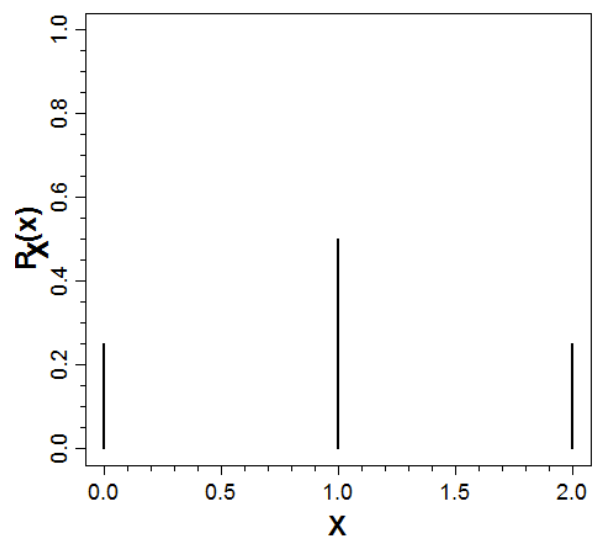
$$X(HH) = 0, X(HT) = 1, X(TH) = 1, X(TT) = 2$$

- 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:

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

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



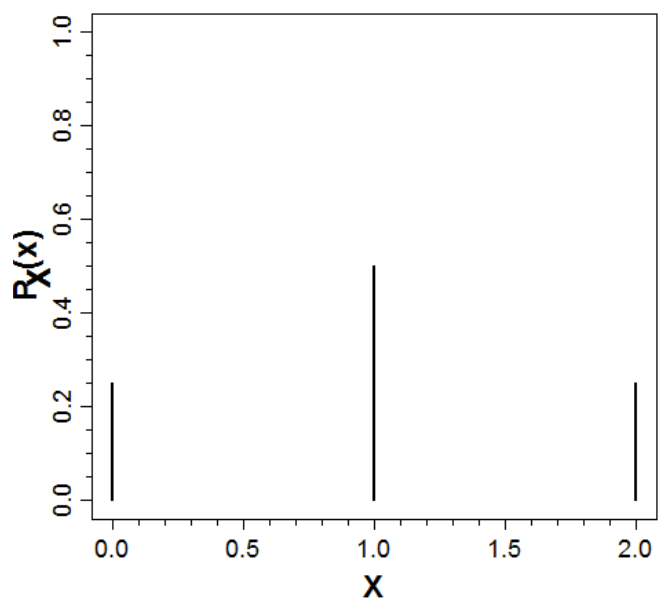
Discrete random variables / cumulative mass functions (cmf)

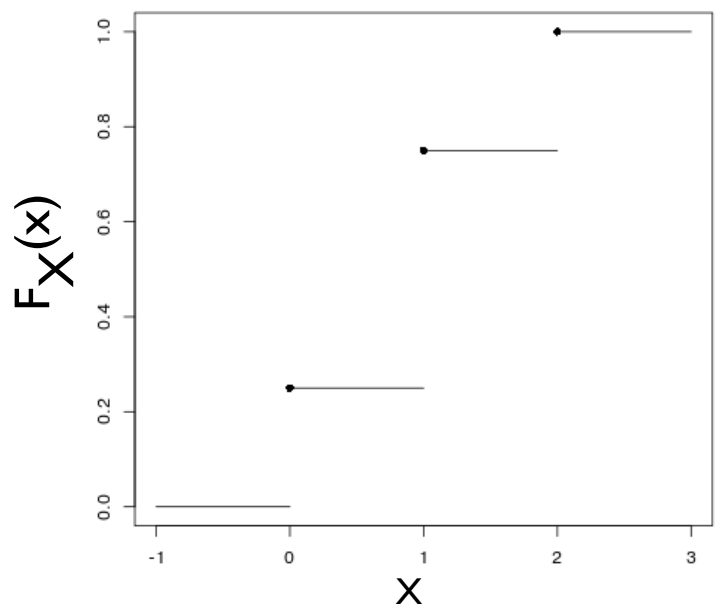
- An alternative (and important!) representation of a discrete probability model is a **cumulative mass function** which we will abbreviate (cmf):

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

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

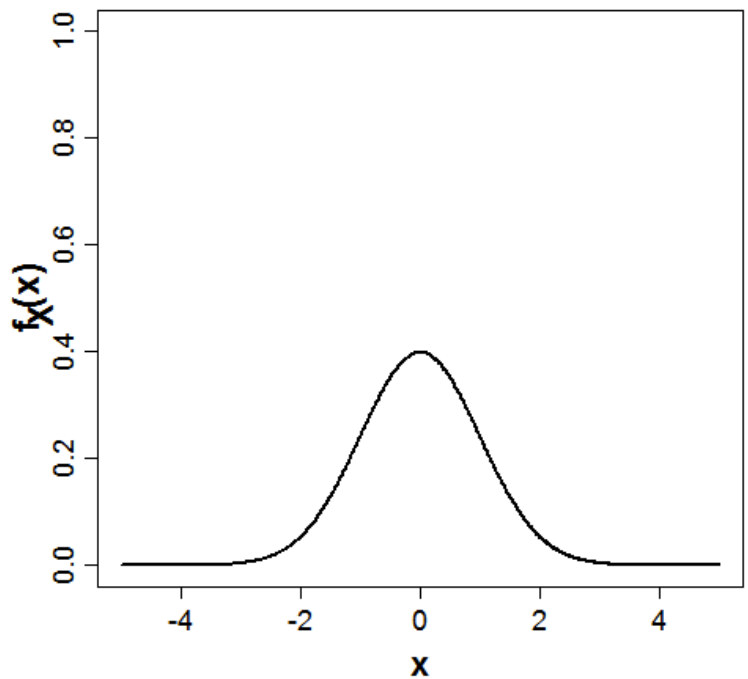
- For the reals, we define a probability density function (pdf): $f_X(x)$
- The pdf of X, a continuous r.v., does not represent the probability of a specific value of X, rather we can use it to find the probability that a value of X falls in an interval [a,b]:

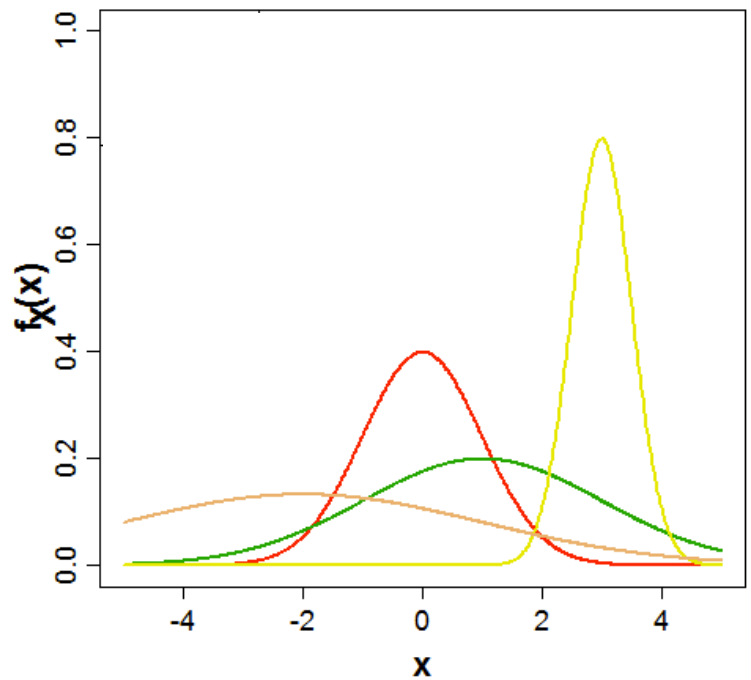
$$Pr(a \leqslant X \leqslant b) = \int_a^b f_X(x)dx$$

- 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!?

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: $f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$



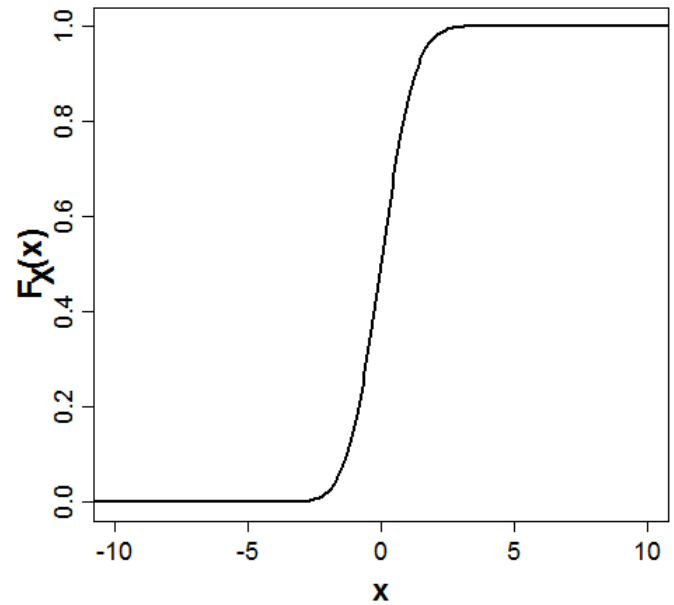


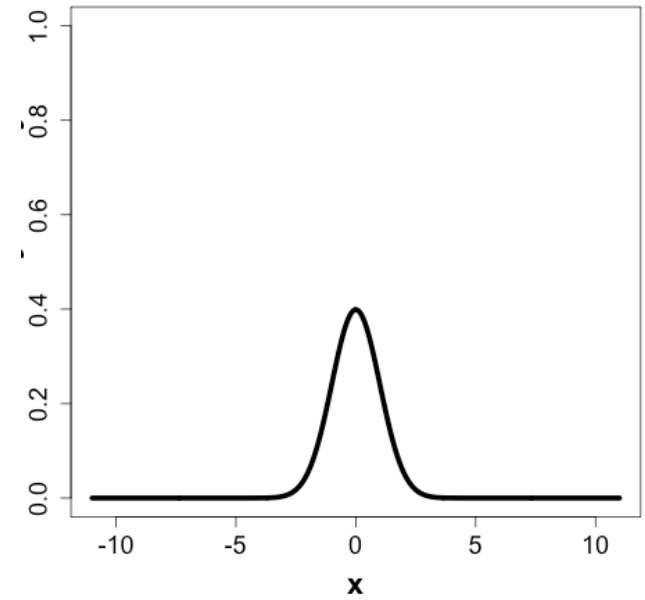
Continuous random variables / cumulative density functions (cdf)

- 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?)





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

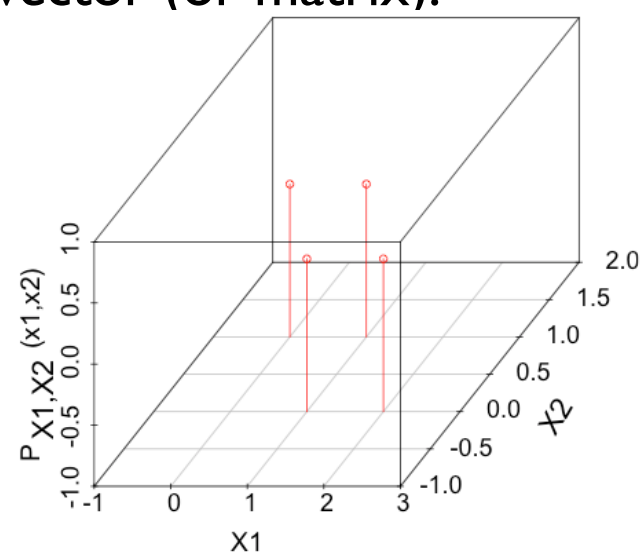
- Consider the two coin flip experiment and assume a probability function for a fair coin: $Pr(\{HH\}) = Pr(\{HT\}) = Pr(\{TH\}) = Pr(\{TT\}) = 0.25$
- Let's define two random variables: "number of Tails" and "first flip is Heads"

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

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

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

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



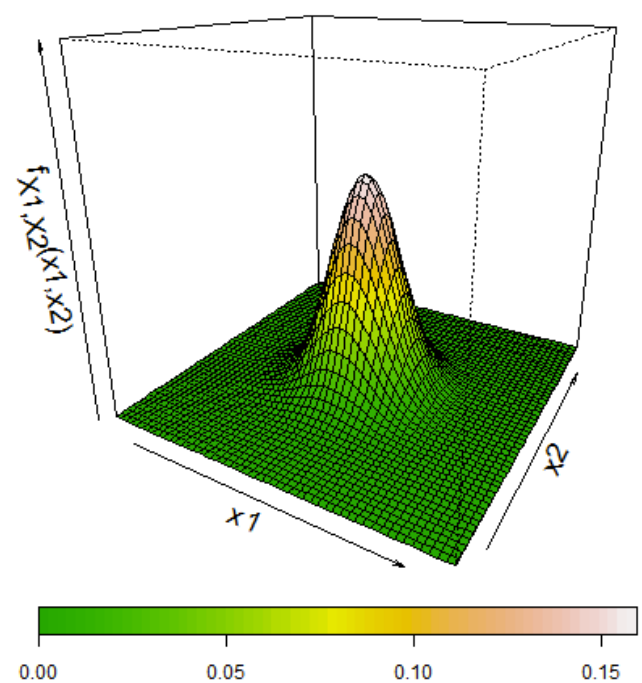
Example of a continuous random vector

- 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]

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



Review: 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 (!!)

Review: random vectors conditional probability and independence II

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

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

$$Pr(X_1 = 0) = Pr(\{HH\}) = 0.25$$
$$Pr(X_1 = 1) = Pr(\{HT, TH\}) = 0.5$$

$$Pr(X_1 = 0, X_2 = 0) = Pr(\{HH\} \cap \{TH, TT\}) = Pr(\emptyset) = 0$$
$$Pr(X_1 = 1, X_2 = 0) = Pr(\{HT, TH\} \cap \{TH, TT\}) = Pr(\{TH\}) = 0.25$$

	$X_2 = 0$	$X_2 = 1$	
$X_1 = 0$	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	

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

$$Pr(X_i \cap X_j) = Pr(X_i) Pr(X_j)$$

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

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*

That's it for today

- Next lecture, we will introduce expectations, variances, and related!