Quantitative Genomics and Genetics

BTRY 4830/6830; PBSB.5201.03

Lecture 6: Introduction to Inference (Probability Models and Samples)

Jason Mezey

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

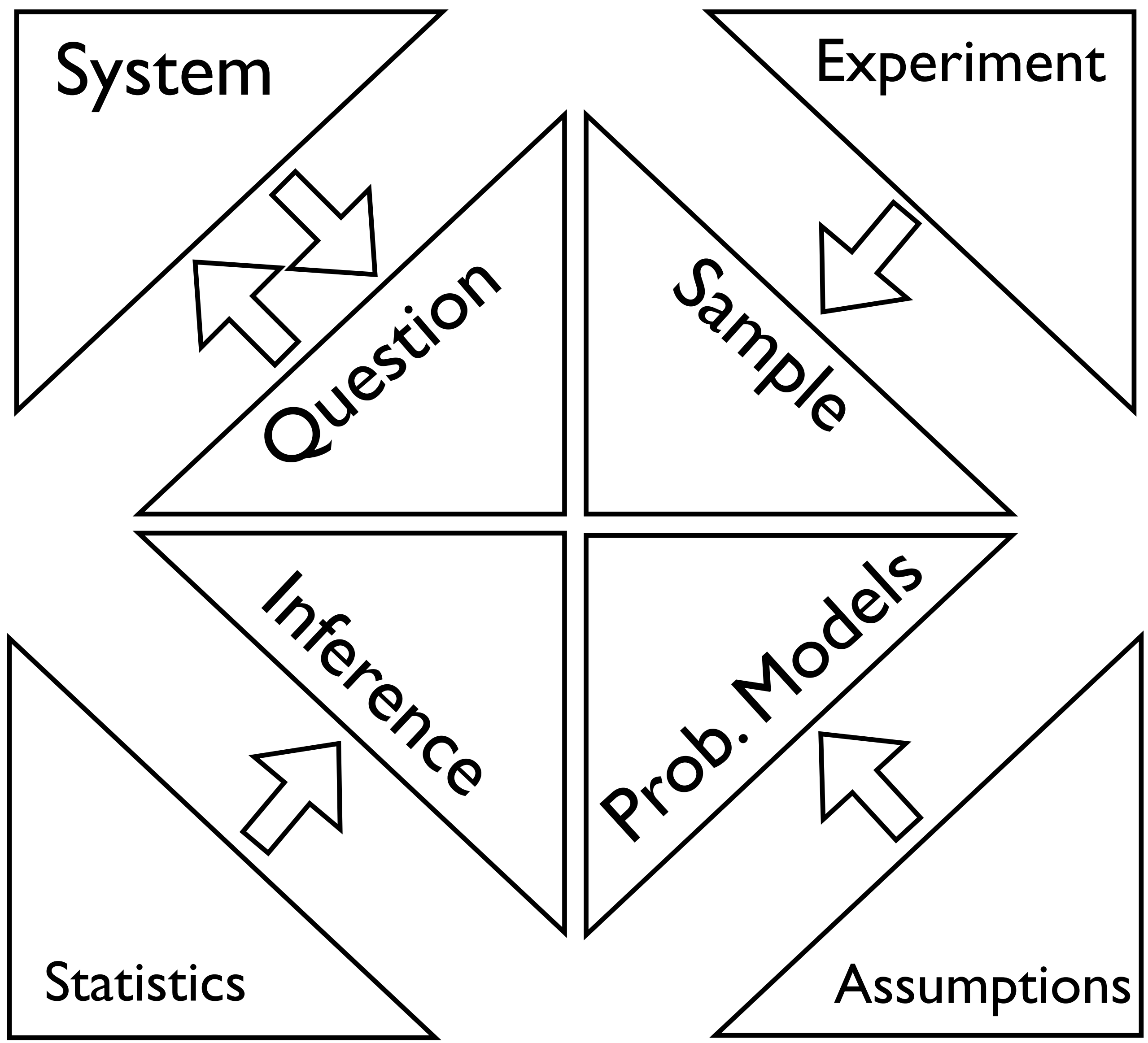
Announcements

- Almost there with CMS… I will send you a Piazza message about this later today so I can compile a complete list of those who need to get on (
- Homework #2: due 11:59PM, Fri., Feb 17 and must be uploaded CMS (!!)
- I will hold office hours this Mon (Feb. 13) 12:30-2:30 by zoom

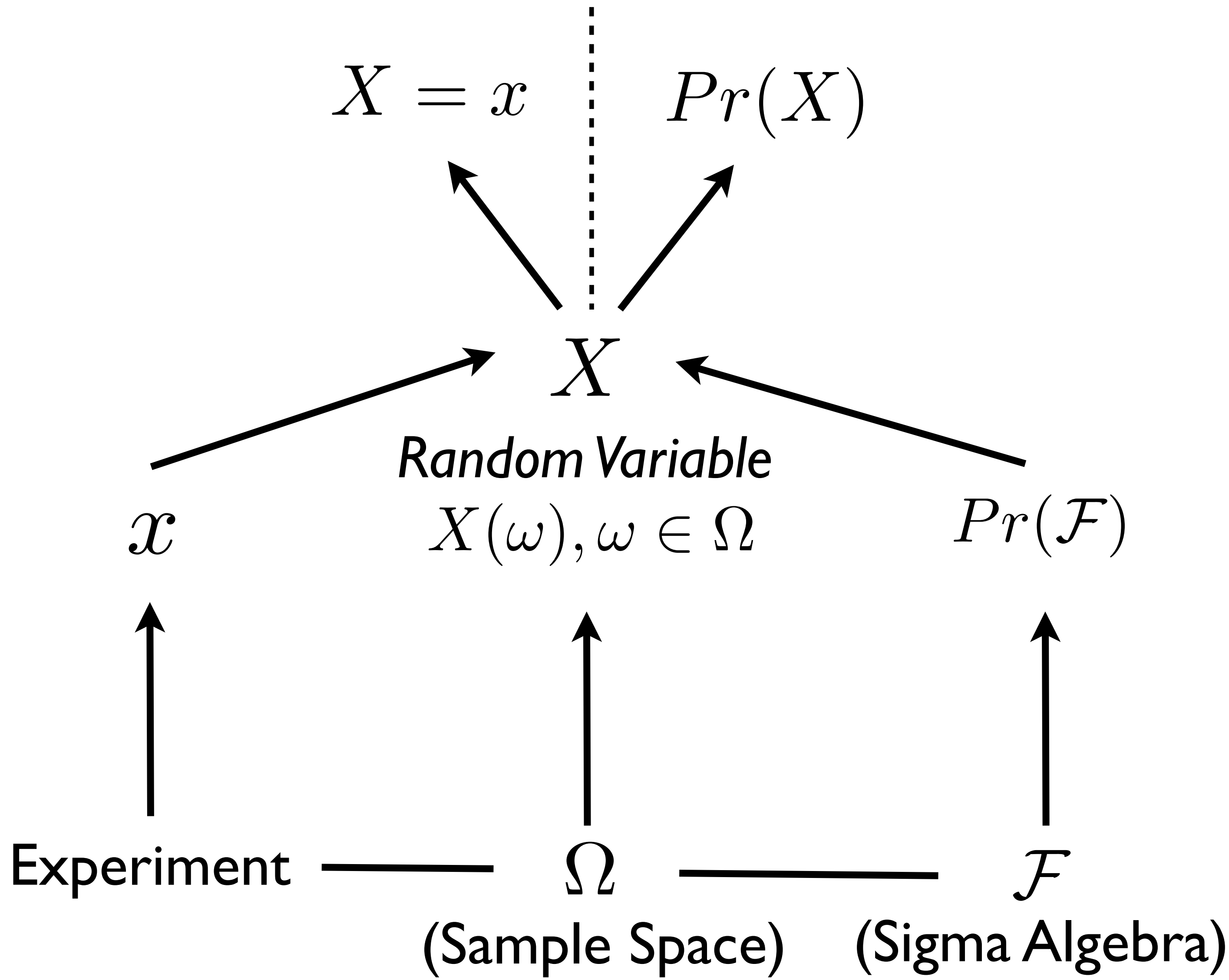
Summary of lecture 6: Introduction to inference

- Last lecture, we discussed expected values, variances and covariances
- Today we will begin our introduction to inference (!!) by introducing parameterized probability models, samples, and statistics!

Conceptual Overview



Review: Random Variables



Review: 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!?)

Review: Random vector conditional probability and independence

- 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: 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):

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

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

Review: 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*

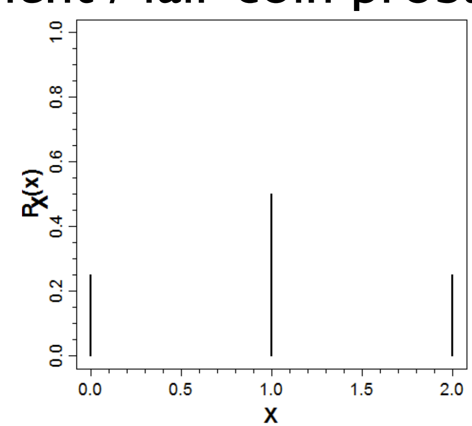
- Note that the interpretation of the expected value is the same in each case

- 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":

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

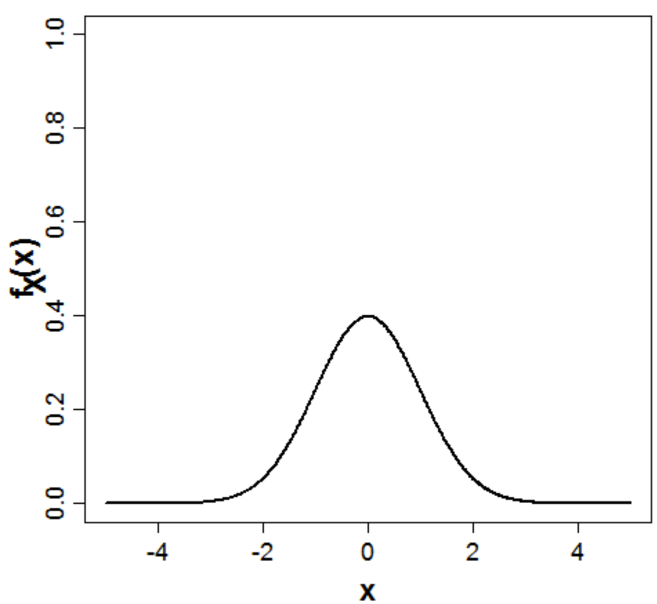


Review: Expectations II

- 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:



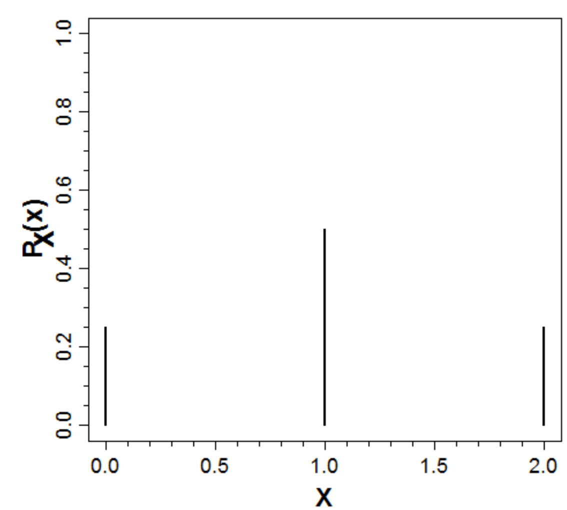
Review: Variances I

- We will define *variances* for *discrete* and *continuous* random variables, where again, the interpretation for both is the same
- The variance of a discrete random variable is defined as follows:

$$\mathrm{Var}(X) = \mathrm{V}(X) = \sum_{i=min(X)}^{max(X)} ((X = i) - \mathrm{E}X)^2 Pr(X = i)$$

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

$$Var(X) = (0-1)^2(0.25) + (1-1)^2(0.5) + (2-1)^2(0.25) = 0.5$$

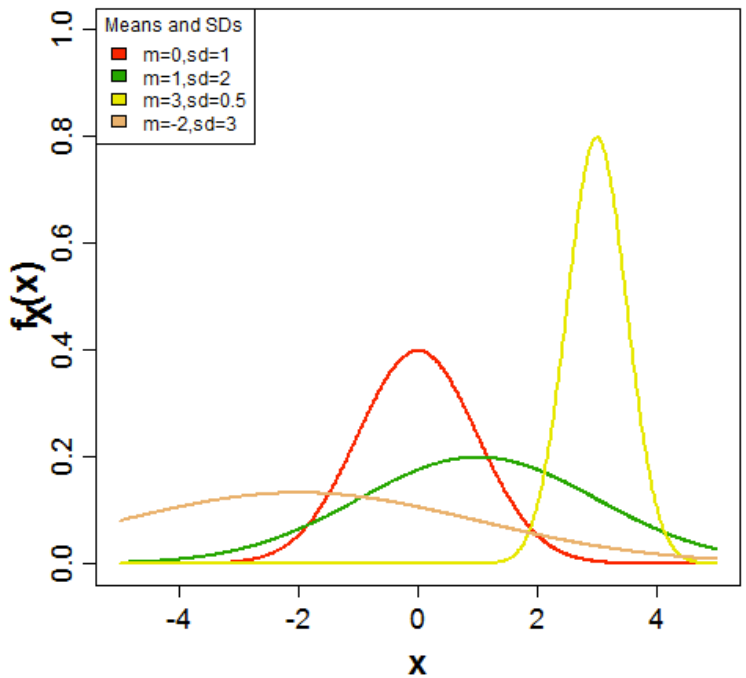


Review: Variances II

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

$$\mathrm{Var}(X) = \mathrm{V}X = \int_{-\infty}^{+\infty} (X - \mathrm{E}X)^2 f_X(x) dx$$

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



Review: Random vectors: expectations and variances

- Recall that a generalization of a random variable is a random vector, e.g.

$$\mathbf{X} = [X_1, X_2] \qquad P_{X_1,X_2}(x_1, x_2) \text{ or } f_{X_1,X_2}(x_1, x_2)$$

- The expectation (a function of a random vector and its distribution!) is a vector with the expected value of each element of the random vector, e.g.

$$\mathrm{E}\mathbf{X} = [\mathrm{E}X_1, \mathrm{E}X_2]$$

- Variances also result in variances of each element (and additional terms... see next slide!!)

- Note that we can determine the conditional expected value or variance of a random variable conditional on a value of another variable, e.g.

$$\mathrm{E}(X_1|X_2) = \sum_{i=min(X_1)}^{max(X_1)} (X_1 = i)Pr(X_i = i|X_2) \qquad \mathrm{V}(X_1|X_2) = \sum_{i=min(X_1)}^{max(X_1)} ((X_1 = i) - \mathrm{E}X_1)^2 Pr(X_i = i|X_2)$$

$$\mathrm{E}(X_1|X_2) = \int_{-\infty}^{+\infty} X_1 f_{X_1|X_2}(x_1|x_2)dx_1 \qquad \mathrm{V}(X_1|X_2) = \int_{-\infty}^{+\infty} (X_1 - \mathrm{E}X_1)^2 f_{X_1|X_2}(x_1|x_2)dx_1$$

Review: Random vectors: covariances

- Variances (again a function!) of a random vector are similar producing variances for each element, but they also produce **covariances**, which relate the relationships between random variables *of a random vector*!!

$$Cov(X_1, X_2) = \sum_{i=min(X_1)}^{i=max(X_1)} \sum_{j=min(X_2)}^{j=max(X_2)} ((X_1 = i) - \mathrm{E}X_1)((X_2 = j) - \mathrm{E}X_2)P_{X_1,X_2}(x_1, x_2)$$

$$\mathrm{Cov}(X_1, X_2) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (X_1 - \mathrm{E}X_1)(X_2 - \mathrm{E}X_2) f_{X_1,X_2}(x_1, x_2) dx_1 dx_2$$

- Intuitively, we can interpret a positive covariance as indicating "big values of X_1 tend to occur with big values of X_2 AND small values of X_1 tend to occur with small values of X_2"

- Negative covariance is the opposite (e.g. "big X_1 with small X_2 AND small X_1 with big X_2")

- Zero covariance indicates no relationship between big and small values of X_1 and X_2

Review: Covariance matrices

- Note that we have defined the "output" of applying an expectation function to a random vector but we have not yet defined the analogous output for applying a variance function to a random vector
- The output is a covariance matrix, which is square, symmetric matrix with variances on the diagonal and covariances on the off-diagonals
- For example, for two and three random variables:

$$\mathrm{Var}(\mathbf{X}) = \begin{bmatrix} \mathrm{Var} X_1 & \mathrm{Cov}(X_1, X_2) \\ \mathrm{Cov}(X_1, X_2) & \mathrm{Var} X_2 \end{bmatrix}$$

$$\mathrm{Var}(\mathbf{X}) = \begin{bmatrix} \mathrm{Var} X_1 & \mathrm{Cov}(X_1, X_2) & \mathrm{Cov}(X_1, X_3) \\ \mathrm{Cov}(X_1, X_2) & \mathrm{Var} X_2 & \mathrm{Cov}(X_2, X_3) \\ \mathrm{Cov}(X_1, X_3) & \mathrm{Cov}(X_2, X_3) & \mathrm{Var}(X_3) \end{bmatrix}$$

- Note that not all square, symmetric matrices are covariance matrices (!!), technically they must be positive (semi)-definite to be a covariance matrix

Review: Covariances and correlations

- Since the magnitude of covariances depends on the variances of X_1 and X_2, we often would like to scale these such that "1" indicates maximum "big with big / small with small" and "-1" indicates maximum "big with small" (and zero still indicates no relationship)

- A correlation captures this relationship:

$$\mathrm{Corr}(X_1, X_2) = \frac{\mathrm{Cov}(X_1, X_2)}{\sqrt{\mathrm{Var}(X_1)}\sqrt{\mathrm{Var}(X_2)}}$$

- Where we can similarly calculate a correlation matrix, e.g. for three random variables:

$$\mathrm{Corr}(\mathbf{X}) = \begin{bmatrix} 1 & \mathrm{Corr}(X_1, X_2) & \mathrm{Corr}(X_1, X_3) \\ \mathrm{Corr}(X_1, X_2) & 1 & \mathrm{Corr}(X_2, X_3) \\ \mathrm{Corr}(X_1, X_3) & \mathrm{Corr}(X_2, X_3) & 1 \end{bmatrix}$$

Algebra of expectations and variances

- If we consider a function (e.g., a transformation) on X (a function on the random variable but not on the probabilities directly!), recall that this can result in a different probability distribution for Y and therefore different expectations, variances, etc. for Y as well

- We will consider two types of functions on random variables and the result on expectation and variances: sums $Y = X_1 + X_2 + ...$ and $Y = a + bX_1$ where a and b are constants

- For example, for sums, $Y = X_1 + X_2$ we have the following relationships:

$$\mathrm{E}(Y) = \mathrm{E}(X_1 + X_2) = \mathrm{E}X_1 + \mathrm{E}X_2$$

$$\mathrm{Var}(Y) = \mathrm{Var}(X_1 + X_2) = \mathrm{Var}X_1 + \mathrm{Var}X_2 + 2\mathrm{Cov}(X_1, X_2)$$

- As another example, for $Y = X_1 + X_2 + X_3$ we have:

$$\mathrm{E}(Y) = \mathrm{E}(X_1 + X_2 + X_3) = \mathrm{E}X_1 + \mathrm{E}X_2 + \mathrm{E}X_3$$

$$\mathrm{Var}(Y) = \mathrm{Var}(X_1 + X_2 + X_3) = \mathrm{Var}X_1 + \mathrm{Var}X_2 + \mathrm{Var}X_3 + 2\mathrm{Cov}(X_1, X_2) + 2\mathrm{Cov}(X_1, X_3) + 2\mathrm{Cov}(X_2, X_3)$$

Algebra of expectations and variances

- For the function $Y = a + bX_1$ we obtain the following relationships:

$$\mathrm{E}Y = a + b\mathrm{E}X$$

$$\mathrm{Var}(Y) = b^2\mathrm{Var}(X)$$

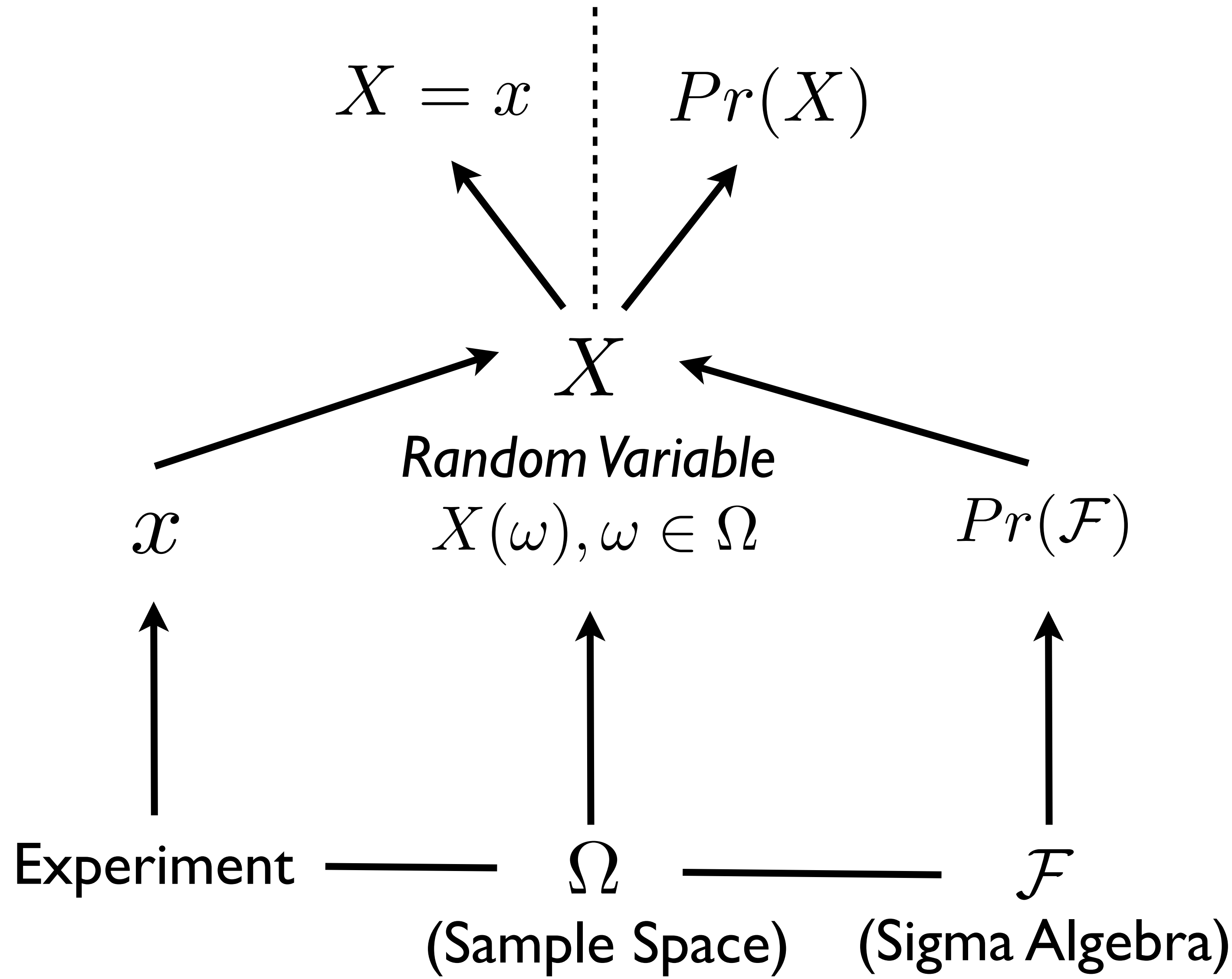
- Finally, note that if we were to take the covariance (or correlation) of two random variables Y_1 and Y_2 with the relationship:

$$Y_1 = a_1 + b_1 X_1,\ \ Y_2 = a_2 + b_2 X_2$$

$$\mathrm{Cov}(Y_1, Y_2) = b_1 b_2 \mathrm{Cov}(X_1, X_2)$$

$$\mathrm{Corr}(Y_1, Y_2) = \mathrm{Corr}(X_1, X_2)$$

So far


$X = x$
$Pr(X)$
X
Random Variable
x
$X(\omega), \omega \in \Omega$
$Pr(\mathcal{F})$
Experiment
Ω
(Sample Space)
$\mathcal{F}$
(Sigma Algebra)

Probability models I

- We have defined Pr(X), a probability model (=probability function!) on a random variable, which technically we produce by defining Pr function on the sigma algebra and the X (random variable function) on the sample space

- So far, we have generally considered such probability models / functions without defining them explicitly (except for a illustrative few examples)

- To define an explicit model for a given system / experiment we are going to assume that there is a "true" probability model, that is a consequence of the experiment that produces sample outcomes

- We place "true" in quotes since the defining a single true probability model for a given case could only really be accomplished if we knew every single detail about the system and experiment (would a probability model be useful in this case?)

- In practice, we therefore assume that the true probability distribution is within a restricted family of probability distributions, where we are satisfied if the true probability distribution in the family describes the results of our experiment pretty well / seems reasonable given our assumptions

Probability models II

- In short, we therefore start a statistical investigation *assuming* that there is a single true probability model that correctly describes the possible experiment outcomes given the uncertainty in our system
- In general, the starting point of a statistical investigation is to make *assumptions* about the form of this probability model
- More specifically, a convenient assumption is to assume our true probability model is specific model in a family of distributions that can be described with a compact equation
- This is often done by defining equations indexed by *parameters*

Probability models III

- **Parameter** - a constant(s) θ which indexes a probability model belonging to a family of models Θ such that $\theta \in \Theta$
- Each value of the parameter (or combination of values if there is more than on parameter) defines a different probability model: Pr(X)
- We assume one such parameter value(s) is the true model
- The advantage of this approach is this has reduced the problem of using results of experiments to answer a broad question to the problem of using a sample to make an educated guess at the value of the parameter(s)
- Remember that the foundation of such an approach is still an assumption about the properties of the sample outcomes, the experiment, and the system of interest (!!!)

Discrete parameterized examples

- Consider the probability model for the one coin flip experiment / number of tails.
- This is the Bernoulli distribution with parameter θ = *p* (what does *p* represent!?) where $\Theta = [0, 1]$
- We can write this *X* ~ Bern(*p*) and this family of probability models has the following form:

$$Pr(X = x|p) = P_X(x|p) = p^x(1-p)^{1-x}$$

- For the experiment of *n* coin flips / number of tails, *one possible* family Binomial distribution *X* ~ Bin(*n*, *p*):

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

$$Pr(X = x|n, p) = P_X(x|n, p) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$n! = n * (n-1) * (n-2) * ... * 1$$

- There are many other discrete examples: hypergeometric, Poisson, etc.

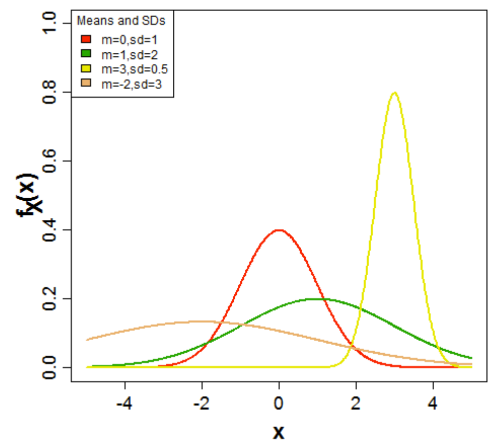
Continuous parameterized examples

- Consider the measure heights experiment (reals as approximation to the sample space) / identity random variable

- For this example we can use the family of normal distributions that are parameterized by $\theta = [\mu, \sigma^2]$ (what do these parameters represent!?) with the following possible values: $\Theta_{\mu} = (-\infty, \infty)$, $\Theta_{\sigma^2} = [0, \infty)$

- We often write this as $X \sim N(\mu, \sigma^2)$ and the equation has the following form:

$$Pr(X = x|\mu, \sigma^2) = f_X(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

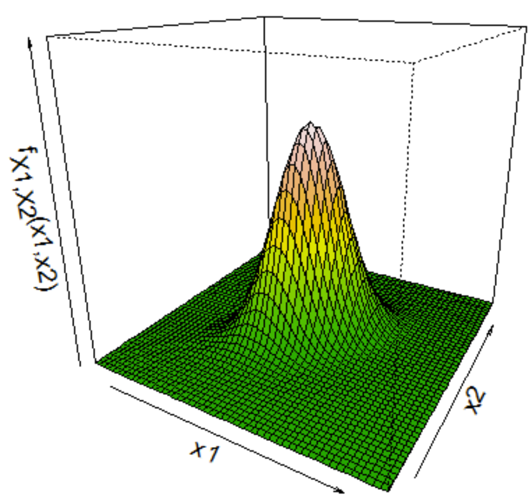


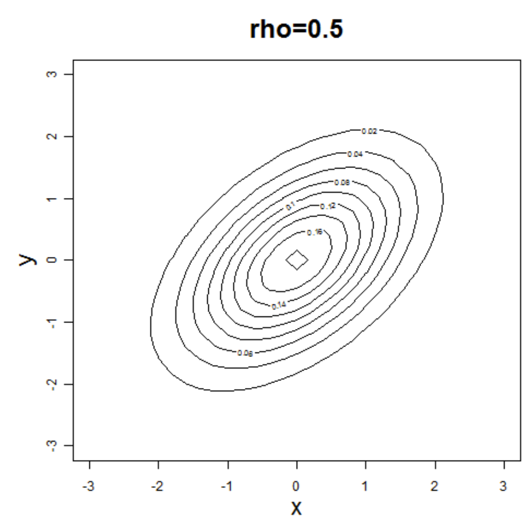
- There are many other continuous examples: uniform, exponential, etc.

Example for random vectors

- Since random vectors are the generalization of r.v.'s, we similarly can define parameterized probability models for random vectors
- As an example, if we consider an experiment where we measure "height" and "IQ" and we take the 2-D reals as the approximate sample space (vector identity function), we could assume the bivariate normal family of probability models:

$$f_{\mathbf{X}}(\mathbf{x}|\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho}} exp\left[-\frac{1}{2(1-\rho^2)}\left(\frac{(x_1-\mu_1)^2}{2\sigma_1^2} - \frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_1)^2}{2\sigma_2^2}\right)\right]$$





Introduction to inference I

- Recall that our eventual goal is to use a sample (generated by an experiment) to provide an answer to a question (about a system)
- So far, we have set up the mathematical foundation that we need to accomplish this goal in a probability / statistics setting (although note we have not yet provided formalism for a sample!!)
- Specifically, we have defined formal components of our framework and made assumptions that have reduced the scope of the problem
- With these components and assumptions in place, we are almost ready to perform *inference*, which will accomplish our goal

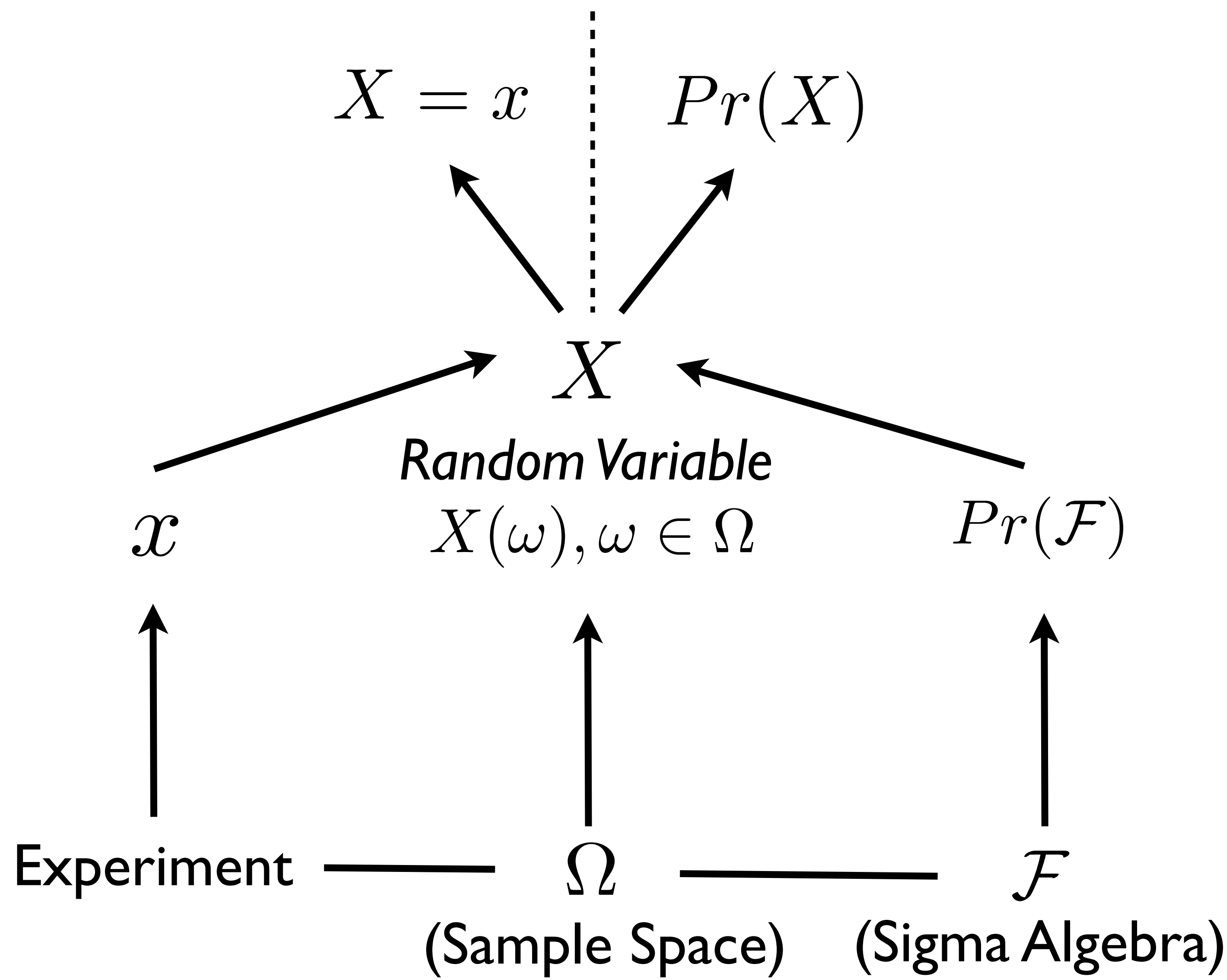
Introduction to inference II

- Our eventual goal is to use a sample (generated by an experiment) to provide an answer to a question (about a system)
- For our system and experiment, we are going to assume there is a single "correct" *probability function* (which in turn defines the probability of our possible random variable outcomes, the probability of possible random vectors that represent samples, and the probability of possible values of a statistic)
- For the purposes of inference, we often assume a *parameterized* family of *probability models* determine the possible cases that contain the "true" model that describes the result of the experiment
- This reduces the problem of inference to identifying the "single" value(s) of the parameter that describes this true model
- Inference (informally) is the process of using the output of an experiment to answer the question

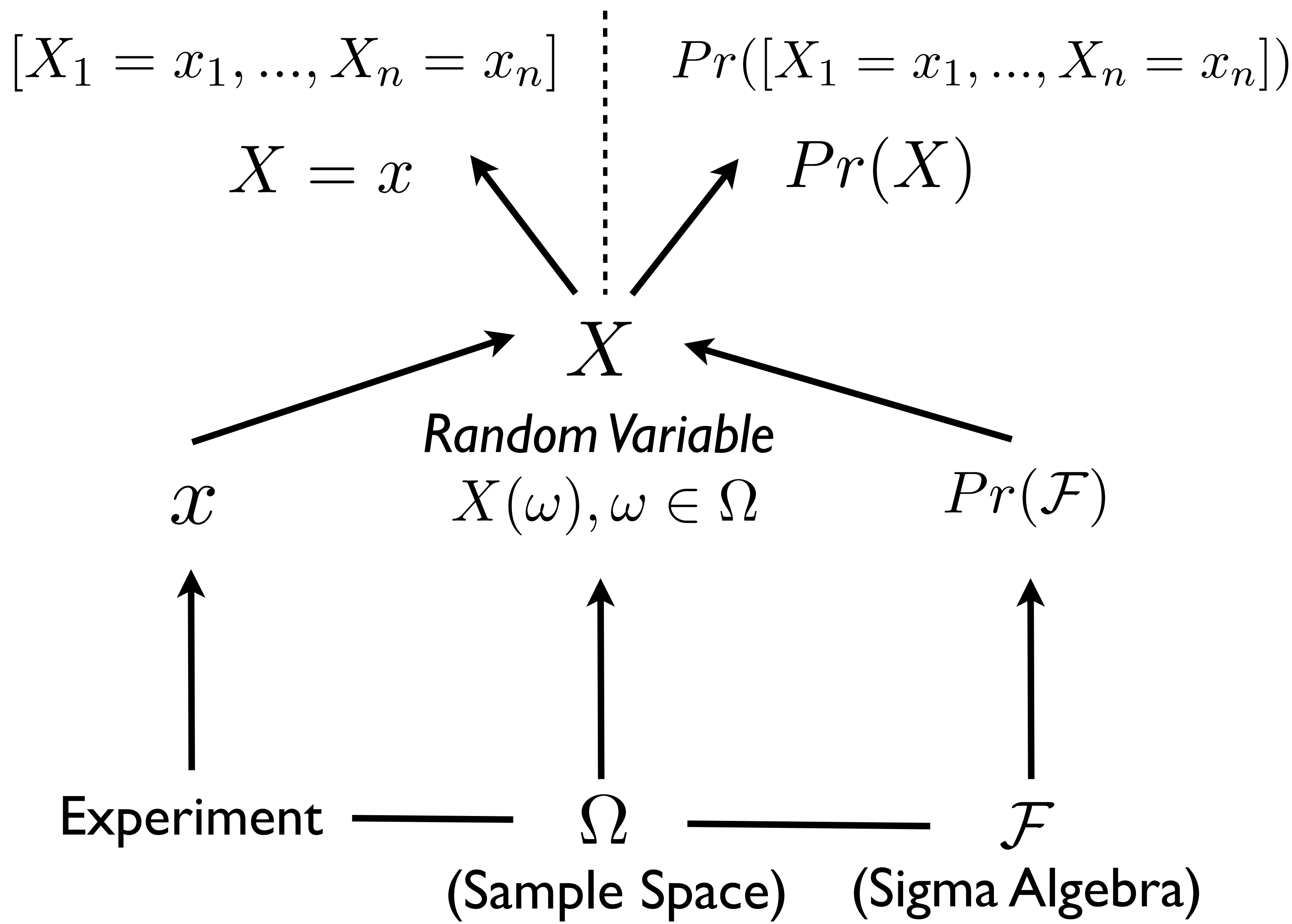
Introduction to inference III

- **Inference** - the process of reaching a conclusion about the true probability distribution (from an assumed family probability distributions, indexed by the value of parameter(s)) on the basis of a sample
- There are two major types of inference we will consider in this course: *estimation* and *hypothesis testing*
- Before we get to these specific forms of inference, we need to formally define: *experimental trials*, *samples*, *sample probability distributions* (or sampling distributions), *statistics*, *statistic probability distributions* (or statistic sampling distributions)

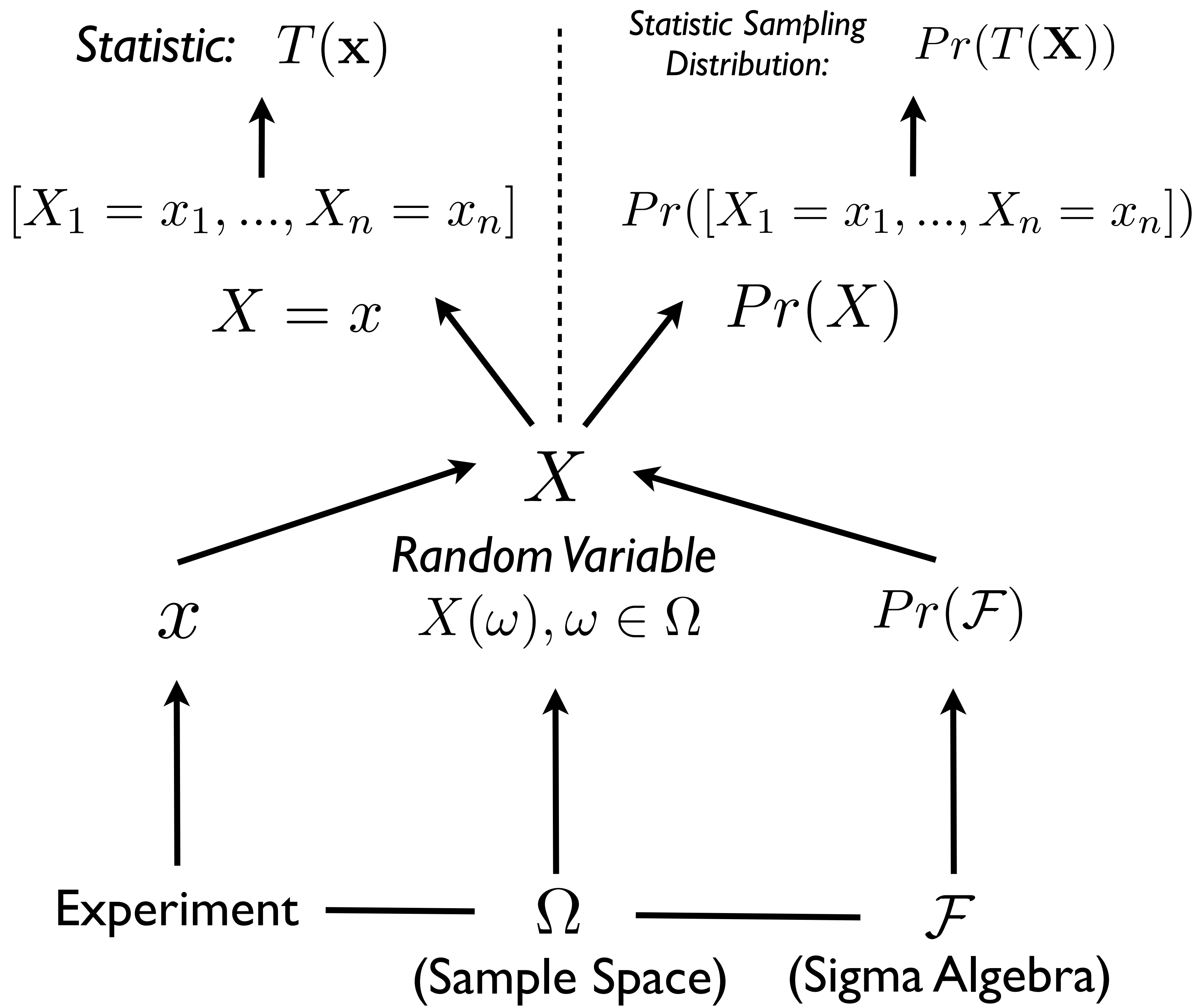
So far


$X = x$
$Pr(X)$
X
Random Variable
x
$X(\omega), \omega \in \Omega$
$Pr(\mathcal{F})$
Experiment
Ω
(Sample Space)
$\mathcal{F}$
(Sigma Algebra)

Where we're headed: Samples



Then: Statistics!



Experiments to Samples (what we observe!)

- **Experiment** - a manipulation or measurement of a system that produces an outcome we can observe
- **Experiment Outcome** - a possible outcome of the experiment
- **Sample Space** - set comprising all possible outcomes of an experiment
- **Experimental Trial** - one instance of an experiment
- **Sample** - (informal) results of one or more experimental trials
- Example (Experiment / Sample Space / Sample):
    - Coin flip / {H,T} / T, T, H, T, H
    - Two coin flips / {HH, HT, TH, TT} / HH, HT, HH, TH, HH
    - Measure heights in this class / Reals / 5’9”, 5’2”, 5’1”, 6’0”, 5’7”

Samples I

- **Sample** - repeated observations of a random variable *X*, generated by experimental trials
- We will consider samples that result from *n* experimental trials (what would be the ideal *n* = ideal experiment!?)
- Since a set of actual experimental outcomes may not be numbers (e.g., a set of H and T's) we want to map them to numbers…
- We already have the formalism to do this and represent a sample of size *n*, specifically this is a random vector:

$$[\mathbf{X} = \mathbf{x}] = [X_1 = x_1, ..., X_n = x_n]$$

- As an example, for our two coin flip experiment / number of tails r.v., we could perform *n*=2 experimental trials, which would produce a sample = random vector with two elements

Example: Observed Sample!

- For example, for our one coin flip experiment / number of tails r.v., we could produce a sample of n = 10 experimental trials, which might look like:

$$\mathbf{x} = [1, 1, 0, 1, 0, 0, 0, 1, 1, 0]$$

- As another example, for our measure heights / identity r.v., we could produce a sample of n=10 experimental trails, which might look like:

$$\mathbf{x} = [-2.3, 0.5, 3.7, 1.2, -2.1, 1.5, -0.2, -0.8, -1.3, -0.1]$$

Samples II

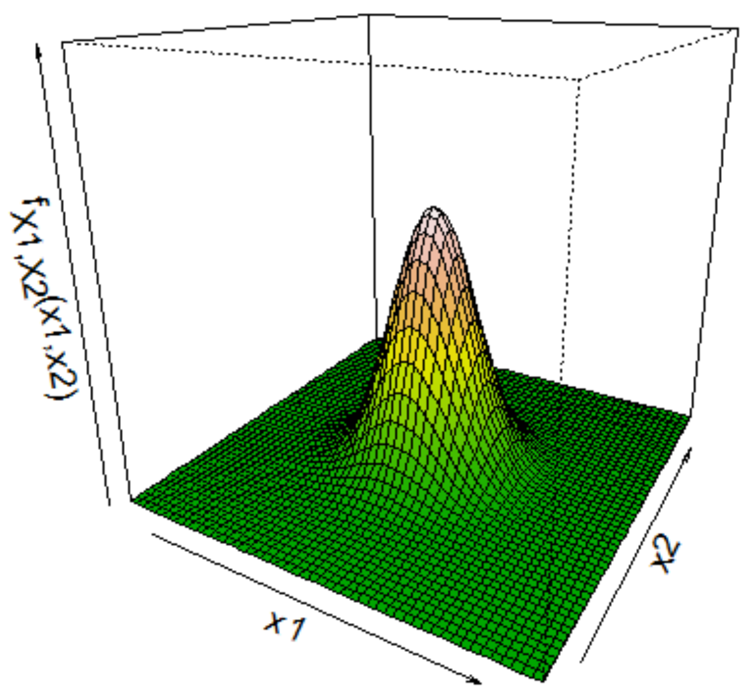
- Recall that we have defined experiments (= experimental trials) in a probability / statistics setting where these involve observing individuals from a population or the results of a manipulation
- We have defined the possible outcome of an experimental trial, i.e. the sample space Ω
- We have also defined a random variable X, where this can take values representing the outcomes of our experimental trials, i.e., $X = x$
- Since the random variable X also has an induced probability distribution associated with it, we can also consider Pr(X), i.e., the probability of each possible outcome of an experiment or the entire sample!
- Since this defines a probability model Pr(X), we have shifted our focus from the sample space to the random variable

Example of sampling distributions

- As an example, consider our height experiment (reals as approximate sample space) / normal probability model (with true but unknown parameters $\theta = [\mu, \sigma^2]$ / identity random variable

- If we assume an i.i.d sample, each sample $X_i = x_i$ has a normal distribution with parameters $\theta = [\mu, \sigma^2]$ and each is independent of all other $X_j = x_j$

- For example, the sampling distribution for an i.i.d sample of $n = 2$ is:



Sample Probability Distribution

- Note that since we have defined (or more accurately induced!) a probability distribution Pr(X) on our random variable, this means we have induced a probability distribution on the sample (!!):

$$Pr(\mathbf{X} = \mathbf{x}) = Pr(X_1 = x_1, X_2 = x_2, ..., X_n = x_n) = P_{\mathbf{X}}(\mathbf{x}) \text{ or } f_{\mathbf{X}}(\mathbf{x})$$

- This is the sample probability distribution or sampling distribution (often called the joint sampling distribution)

- While samples could take a variety of forms, we generally assume that each possible observation in the sample has the same form, such that they are identically distributed:

$$Pr(X_1 = x_1) = Pr(X_2 = x_2) = ... = Pr(X_n = x_n)$$

- We also generally assume that each observation is independent of all other observations:

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

- If both of these assumptions hold, than the sample is independent and identically distributed, which we abbreviate as i.i.d.

That's it for today

- Next lecture, we will begin our discussion of statistics (and estimators)!