Quantitative Genomics and Genetics

BTRY 4830/6830; PBSB.5201.03

Lecture 5: Expectations, Variances, and Covariances

Jason Mezey

Feb 7, 2023 (Th) 8:05-9:20

Announcements I

- Everyone should be all set with Piazza and the class website
- For CMS, everyone with a netID should now be good (please Piazza message me and let me know if you have a netID and are still not able to see the class CMS!) and we are working on CWID folks (note
-

Announcements II

- Homework 1, question 1 edit-correction

Problem 1 (Easy)

Consider a coin (system) that you would like to learn about and two types of experiments: 1. One flip of the coin (Experiment 1), and Two flips of the coin (Experiment 2). Consider the case where you are going to perform Experiment 1 TWICE and Experiment 2 ONCE.

a. Write out BOTH the sample spaces AND the Sigma Algebras for BOTH Experiments 1 and 2.

   Note for Experiment 1 the answer is: $\Omega = \{H, T\}$, σ-algebra$= \emptyset, \{H\}, \{T\}, \{H, T\}$

b. Using one sentence at most, explain why the sets $\{H_1, T_2\}$ describing the result of 'heads' when running the Experiment 1 the first time and a 'tails' when running Experiment 1 the second time, is distinct from the set $\{HT\}$ describing the results of running Experiment 2 one time.

c. Define a probability model on Ω (i.e. assign specific probabilities to each outcome) for Experiment 1 such that $Pr(\{H\}) = 0.8, Pr(\{T\}) = 0.2$. What is the probability of each event of the Sigma-algebra that you defined for this Experiment in part [a]? Note: we are asking for the probabilities for the Sigma-algebra of Experiment 1 (i.e., Problem 1a) and NOT for what resulted from running this experiment two times (i.e., Problem 1b)? Could this be a legitimate probability model for a coin / this experiment? Explain your answer using no more than one sentence.

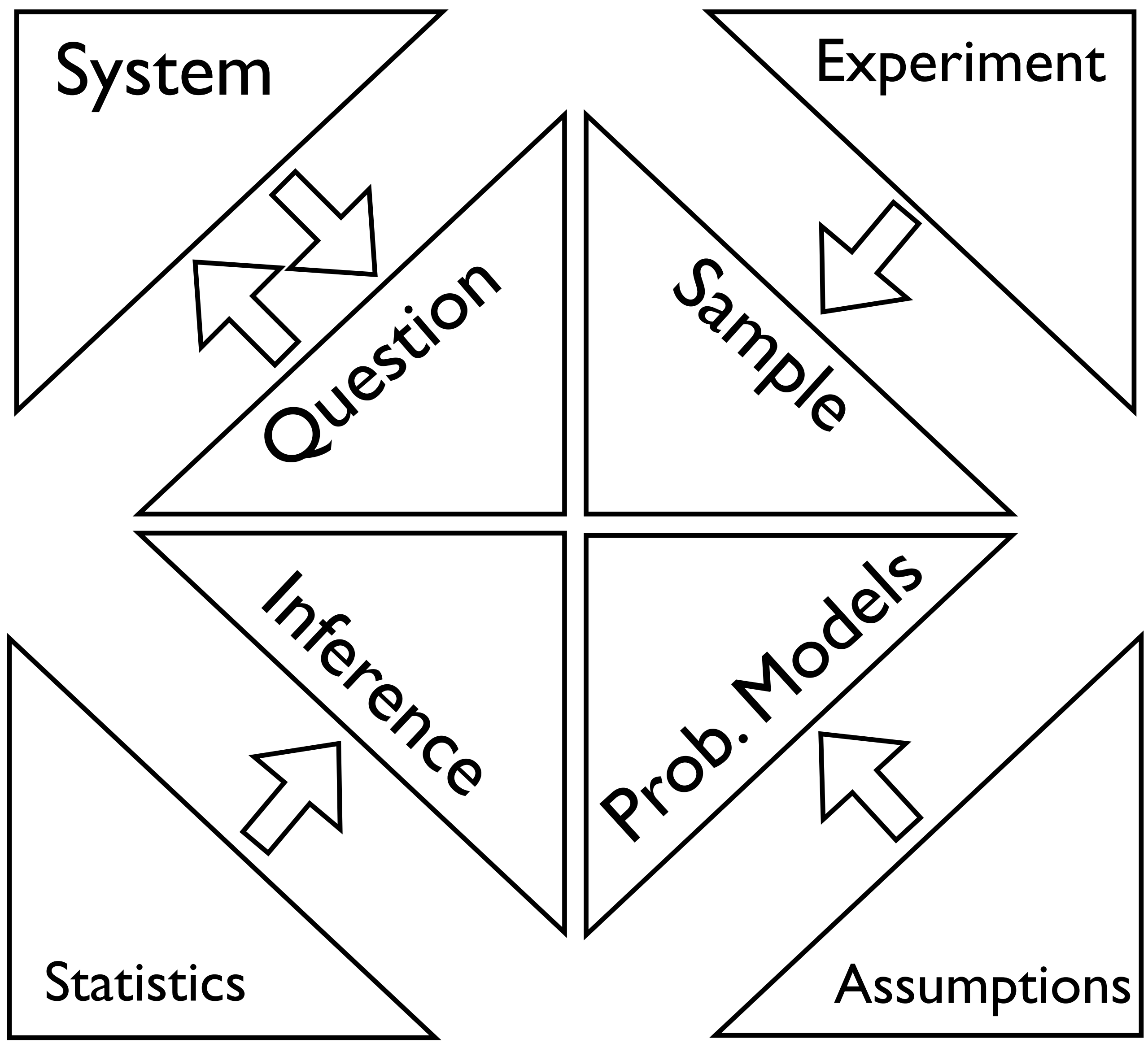
d. ~~Considering the probability model in part [c] calculate~~ $\cancel{Pr(\{H\} \cap \{T\})}$ ~~and explain why this demonstrates the events~~ $\cancel{\{H\}}$ ~~and~~ $\cancel{\{T\}}$ ~~are not independent.~~

- Note after class, I will post this version to CMS and send by Piazza message (and note: fine to hand in your work on the previous latex version!)

Summary of lecture 5: Expectations, Variances, Covariances

- Last class, we introduced probability distributions of random variables (including discrete and continuous forms)
- Today, we will continue our discussion by introducing random vectors ad their probability distributions (and conditional probability for random vectors)
- We will also discuss the extremely useful concept of expectations, variances, and covariances

Conceptual Overview



Review: Probability functions I

- **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 $\mathcal{A}_1, \mathcal{A}_2, ... \subset \Omega$, 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: Random variables I

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

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

- Intuitively:

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

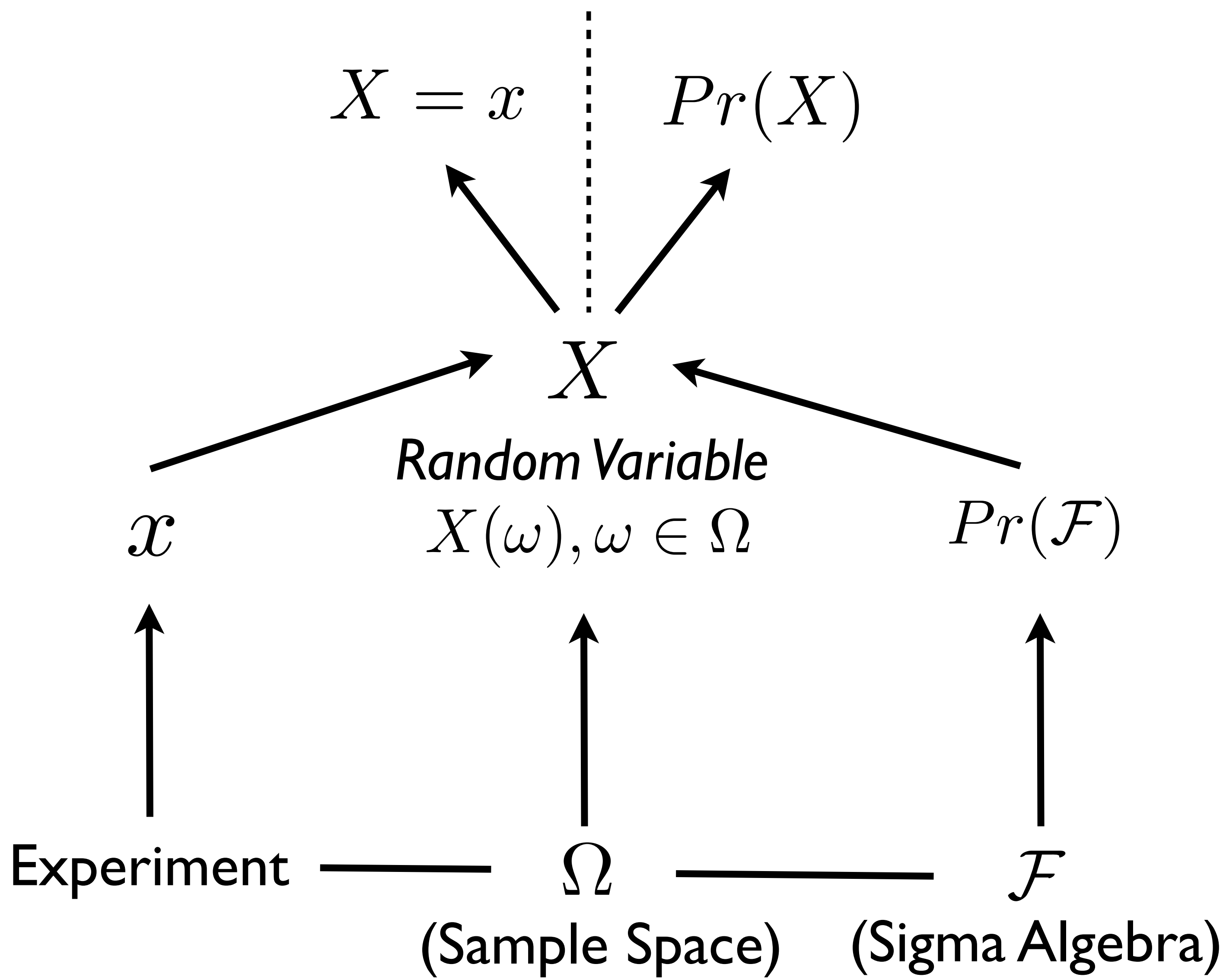
- Note that these functions are not constrained by the axioms of probability, e.g. not constrained to be between zero or one (although they must be measurable functions and admit a probability distribution on the random variable!!)

- 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:

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

Review: Random Variables II



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:

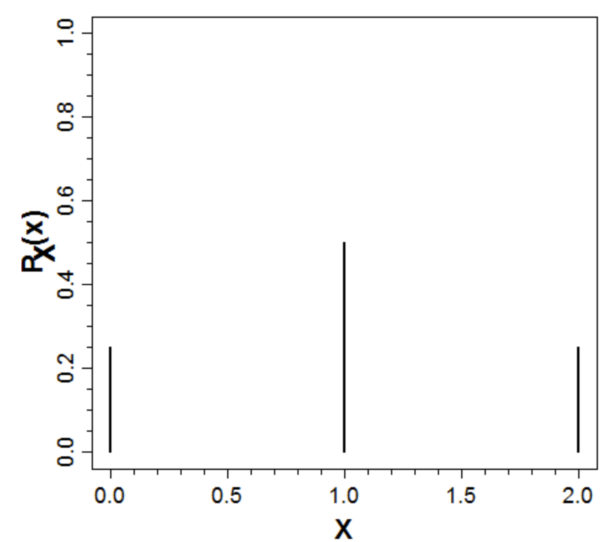
$$X(HH) = 0, X(HT) = 1, X(TH) = 1, X(TT) = 2$$

- 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:

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

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



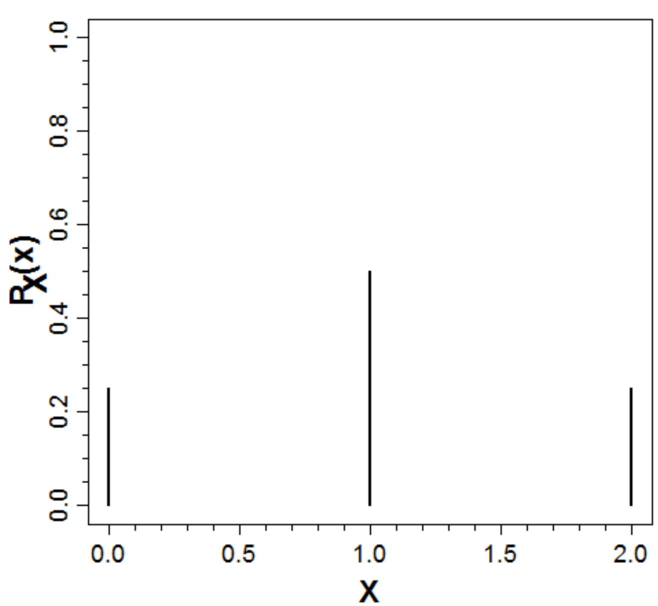
Review: Discrete random variables / cumulative mass functions (cmf)

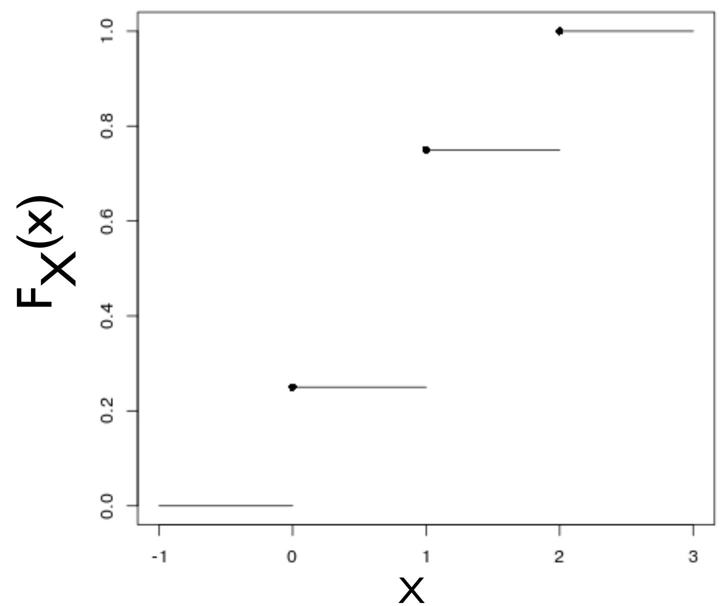
- An alternative (and important!) representation of a discrete probability model is a **cumulative mass function** which we will abbreviate (cmf):

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

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:





Review: 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

Review: 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]:

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

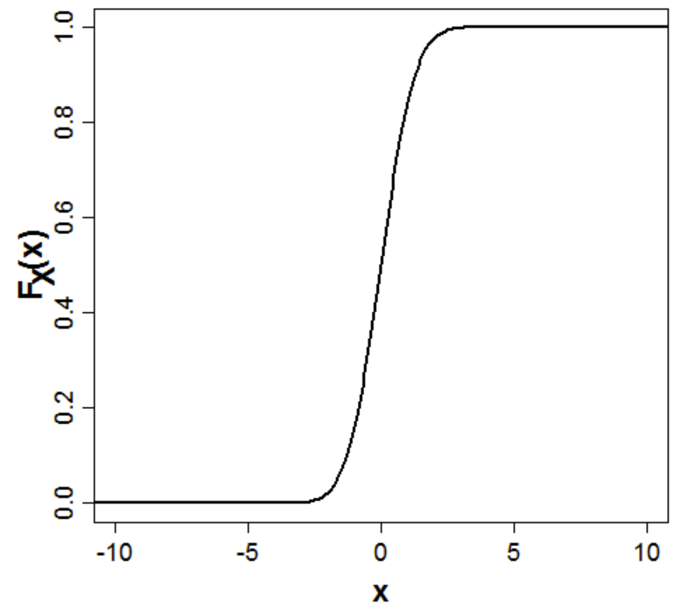
- Related to this concept, for a continuous random variable, the probability of specific value (or point) is zero (why is this!?)
- For a specific continuous distribution the cdf is unique but the pdf is not, since we can assign values to non-measurable sets
- If this is the case, how would we ever get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

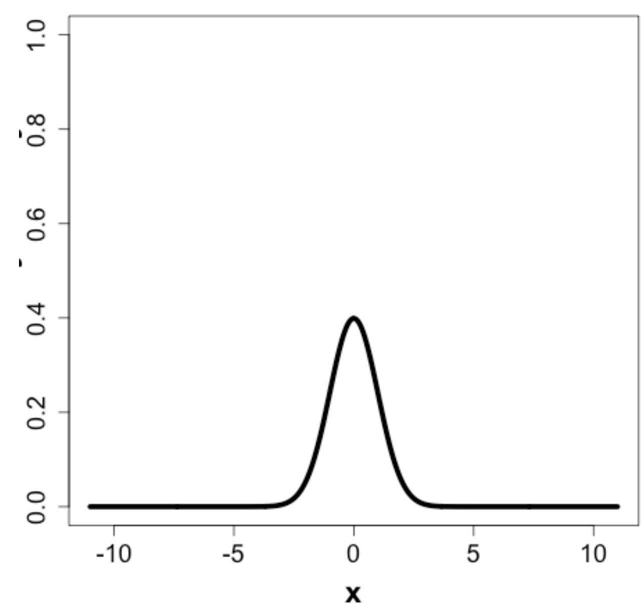
Review: Continuous random variable / cumulative density functions (cdf)

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

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

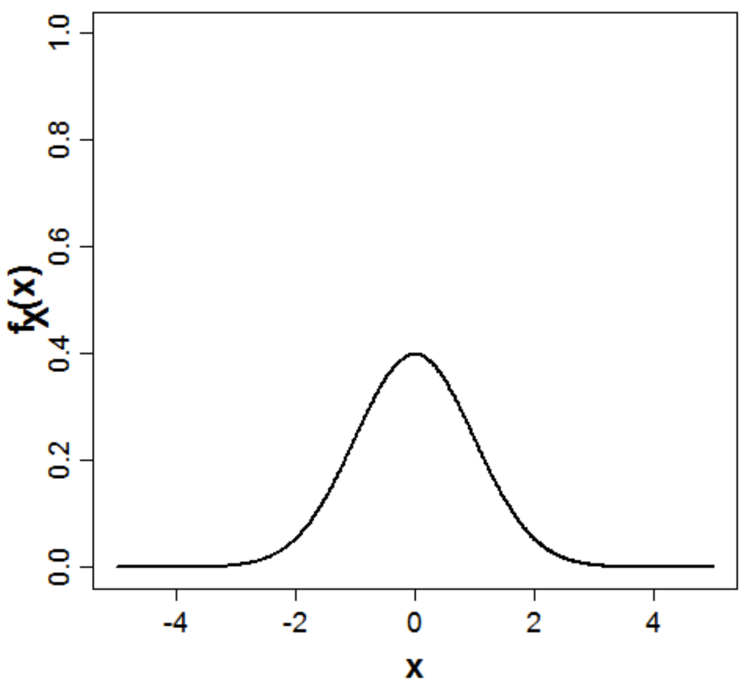
- Again, a graph illustration is instructive
- Note the cdf runs from zero to one (why is this?)

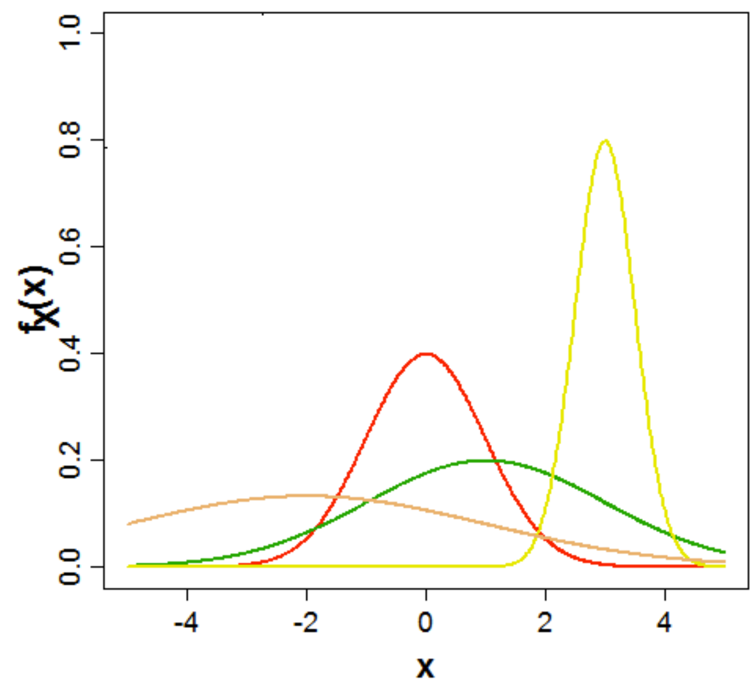




Probability density functions (pdf): normal example

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





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

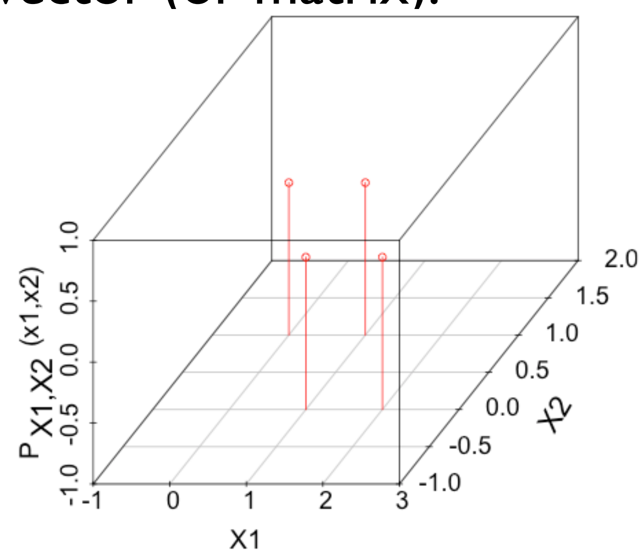
- Let's define two random variables: "number of Tails" and "first flip is Heads"

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

- 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):

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

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



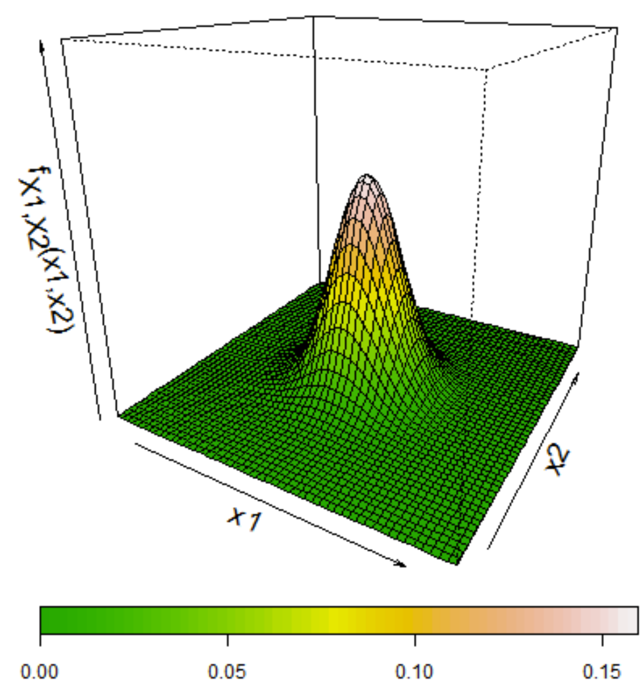
Example of a continuous random vector

- 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

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

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

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

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

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

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*

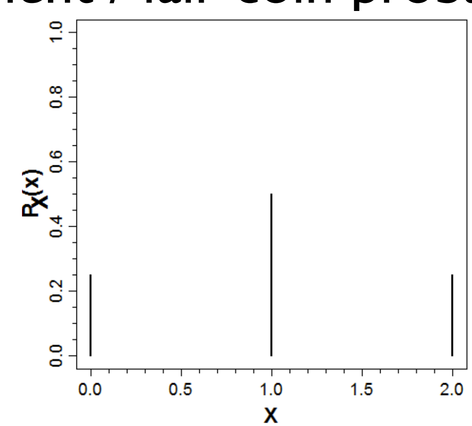
- Note that the interpretation of the expected value is the same in each case

- 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":

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

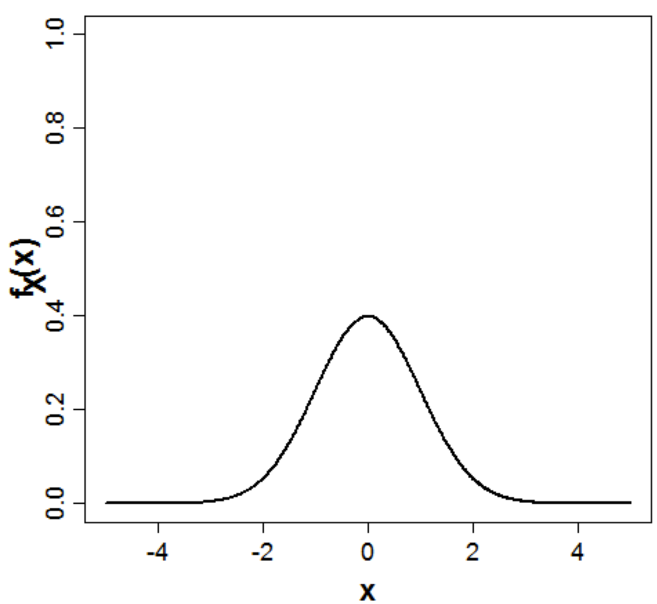


Expectations II

- 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

Variances I

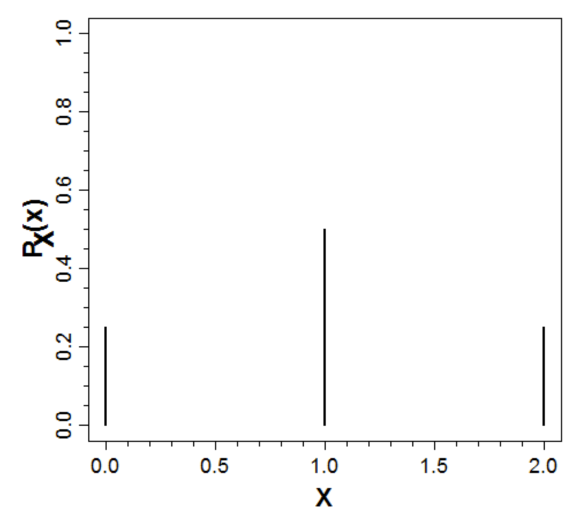
- We will define *variances* for *discrete* and *continuous* random variables, where again, the interpretation for both is the same

- 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":

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

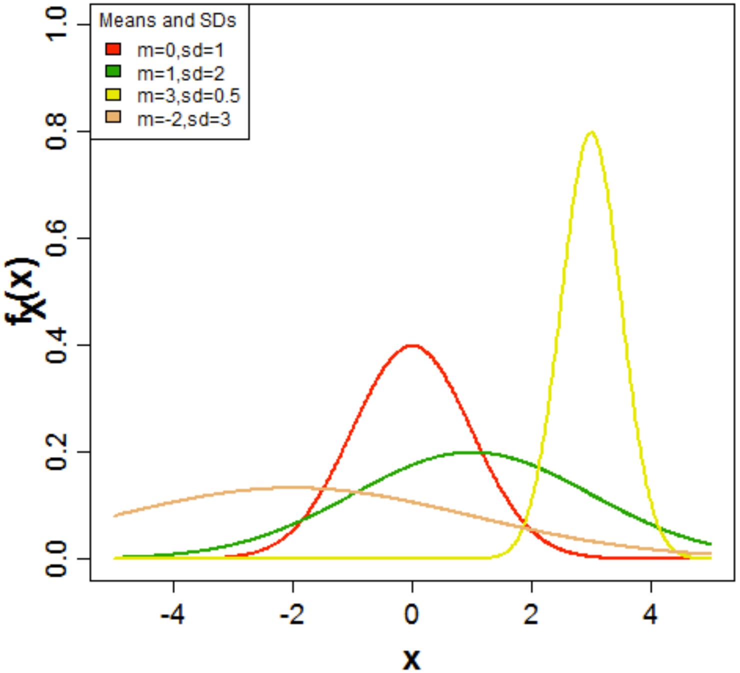


Variances II

- 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:



Variances III

- Intuitively, the variance quantifies the "spread" of a distribution
- The squared component of variance has convenient mathematical properties, e.g. we can view them as sides of triangles
- Other equivalent (and often used) formulations of variance:

$$\mathrm{Var}(X) = \mathrm{E}\left[(X - \mathrm{E}X)^2\right]$$
$$\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}X)^2$$

- Instead of viewing variance as including a squared term, we could view the relationship as follows:

$$\mathrm{Var}(X) \;=\; \mathrm{E}[(X - \mathrm{E}X)(X - \mathrm{E}X)]$$

Generalization: higher moments

- The expectation of a random variable is the "first" moment and we can generalize this concept to "higher" moments:

$$\mathrm{E}X^k = \sum X^k Pr(X)$$

$$\mathrm{E}X^k = \int X^k f_X(x) dx$$

- The variance is the second "central" moment (i.e. calculating a moment after subtracting off the mean) and we can generalize this concept to higher moments as well:

$$\mathrm{C}(X^k) = \sum (X - \mathrm{E}X)^k Pr(X)$$

$$\mathrm{C}(X^k) = \int (X - \mathrm{E}X)^k f_X(x) dx$$

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.

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

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

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*!!

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

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

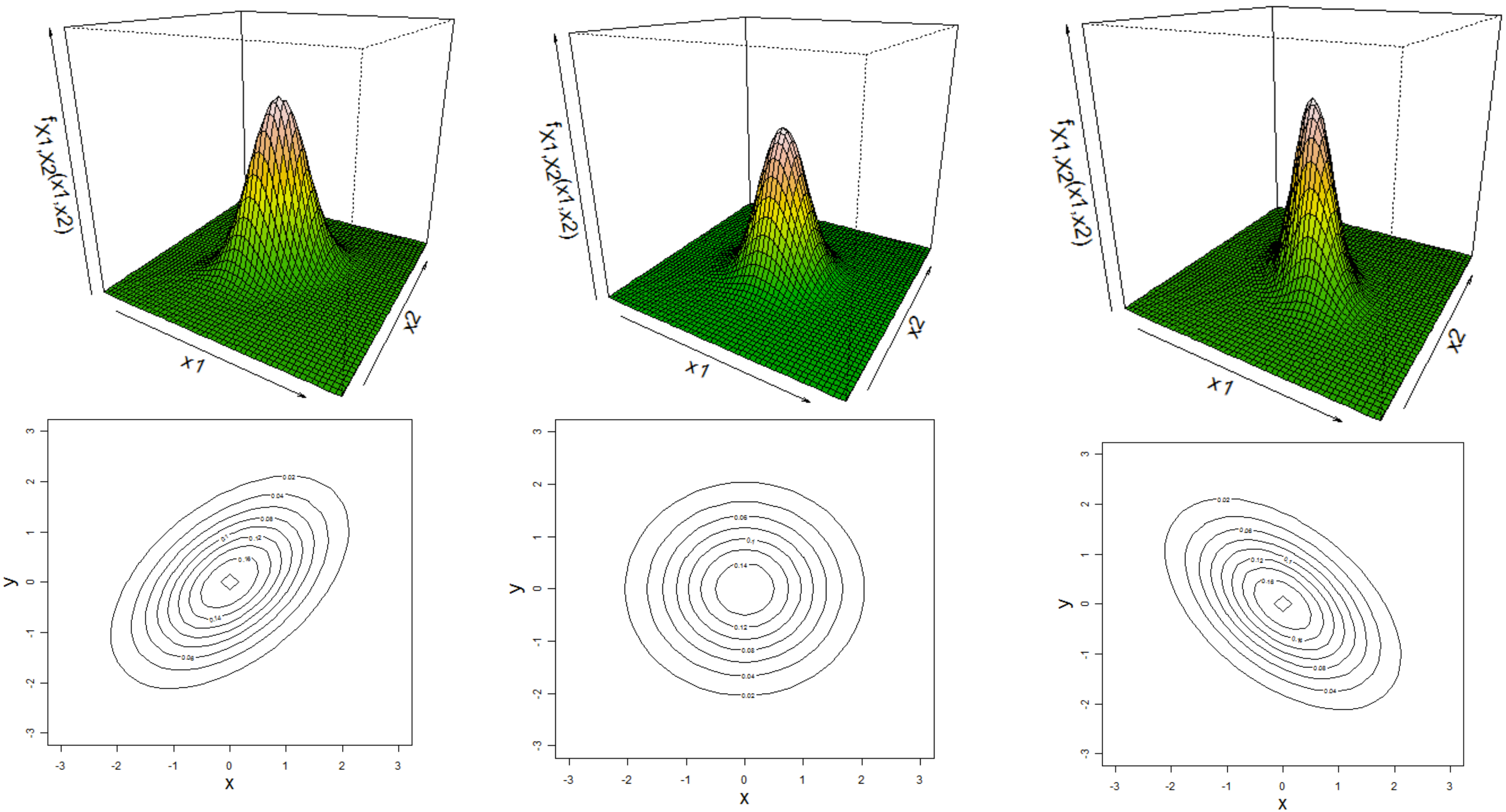
- Intuitively, we can interpret a positive covariance as indicating "big values of X_1 tend to occur with big values of X_2 AND small values of X_1 tend to occur with small values of X_2"

- 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

An illustrative example

- For example, consider our experiment where we have measured “height” and “IQ” / bivariate normal probability model / identity random variable:



Notes about covariances

- Covariance and independence, while related, are NOT synonymous (!!), although if random variables are independent, then their covariance is zero (but necessarily vice versa!)
- Covariances are symmetric: $\mathrm{Cov}(X_1, X_2) = \mathrm{Cov}(X_2, X_1)$
- Other equivalent (and often used) formulations of covariances:

$$\mathrm{Cov}(X_1, X_2) = \mathrm{E}\left[(X_1 - \mathrm{E}X_1)(X_2 - \mathrm{E}X_2)\right]$$

$$\mathrm{Cov}(X_1, X_2) = \mathrm{E}(X_1 X_2) - \mathrm{E}X_1 \mathrm{E}X_2$$

- From these formulas, it follows that the covariance of a random variable and itself is the variance:

$$\mathrm{Cov}(X_1, X_1) = \mathrm{E}(X_1 X_1) - \mathrm{E}X_1 \mathrm{E}X_1 = \mathrm{E}(X_1^2) - (\mathrm{E}X_1)^2 = \mathrm{Var}(X_1)$$

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

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 $Y = X_1 + X_2 + \ldots$ 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:

$$Y_1 = a_1 + b_1X_1,\ \ Y_2 = a_2 + b_2X_2$$

$$\mathrm{Cov}(Y_1, Y_2) = b_1b_2\mathrm{Cov}(X_1, X_2)$$

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

That's it for today

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