CS 237: Probability in Computing

Wayne Snyder
Computer Science Department
Boston University

Lecture 16:

- Joint Random Variables: Basic Notions
- JRVs: Conditional JRVs and Independence

Random Variables: Review

Recall: A Random Variable X is a function from a sample space S into the reals:

$$X : S \to \mathcal{R}$$

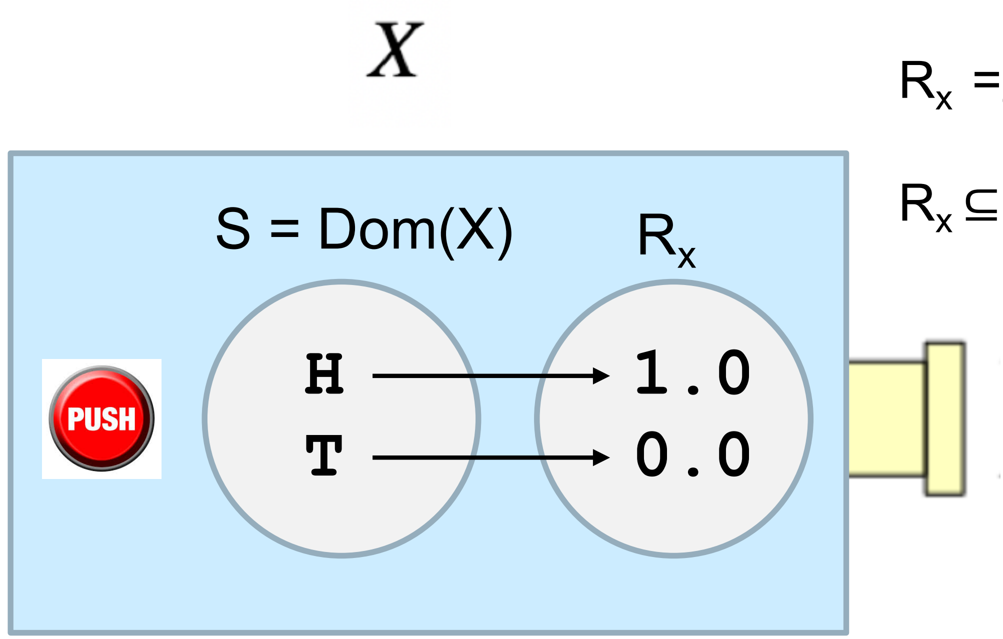
When an outcome is requested, the sample point is translated into a real number:

```
def X():
  return randint(0,2)
```

S = Domain(X)

$R_x =_{def} \text{Range}(X)$

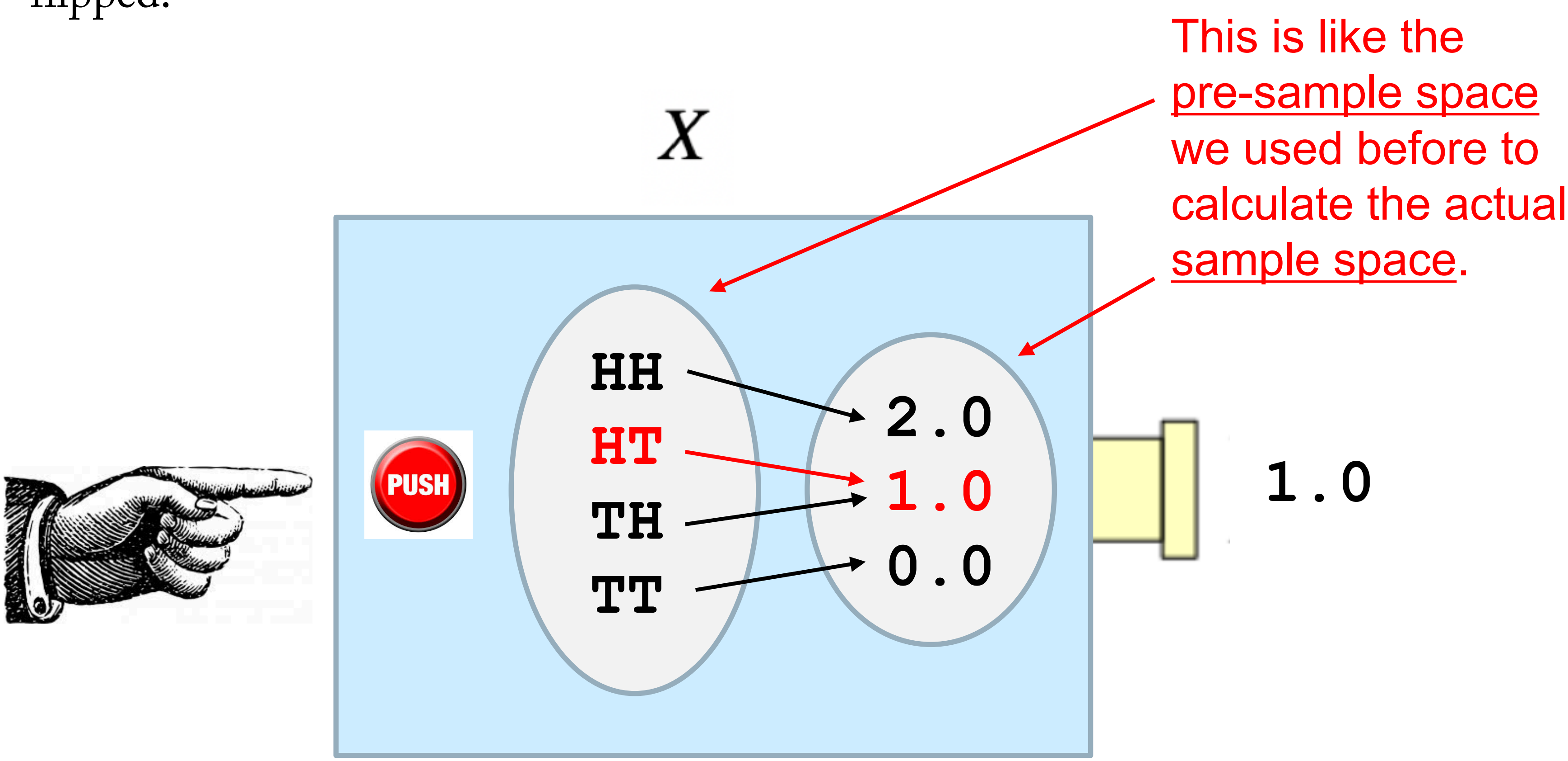
$R_x \subseteq \mathcal{R}$



Random Variables

This may seem awkward, but it helps to explain the difference between random experiments whose literal outcomes are not numbers, but which are translated into numbers for clarity.

Example: X = “the number of heads which appear when two fair coins are flipped.



Joint Random Variables

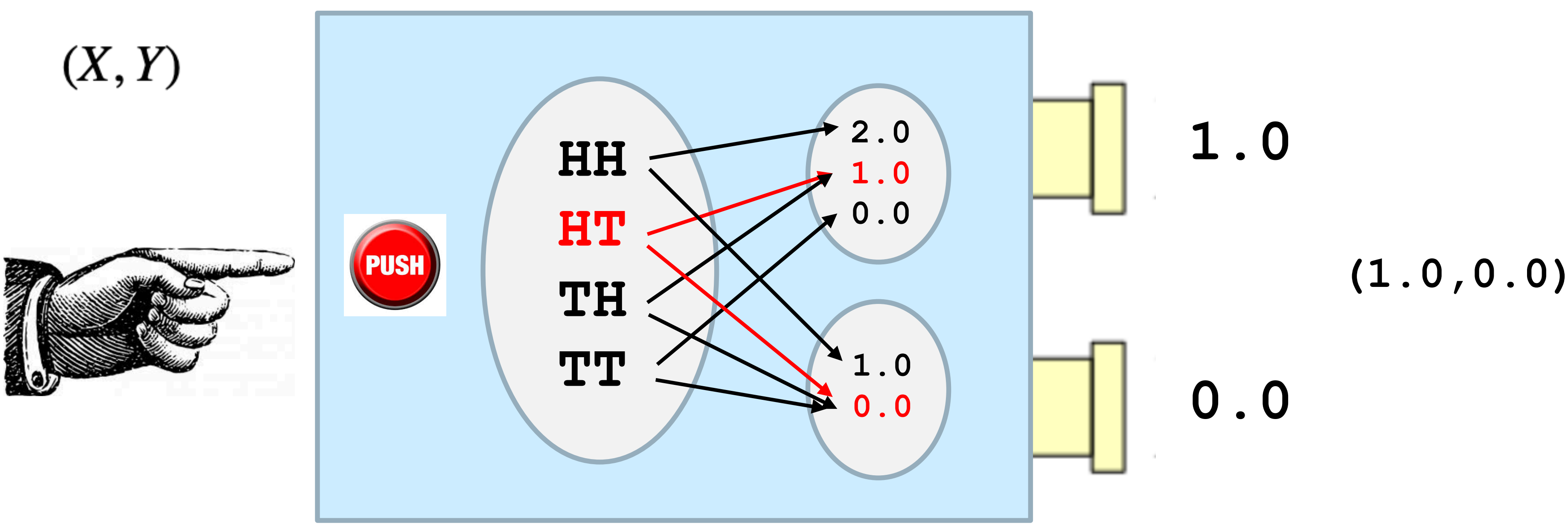
A Joint Random Variable is a pair of random variables:

$$(X, Y) \; : \; S \rightarrow \mathcal{R} \times \mathcal{R}$$

Now when an outcome is requested, the sample point is translated into two real numbers by the action of each random variable responding to the same experiment:

Throw two dice: X = "the number of heads showing," and
Y = "1 if both tosses are heads, 0 otherwise."

```
def XY():
  a = randint(0,2)
  b = randint(0,2)
  return (a+b,a*b)
```



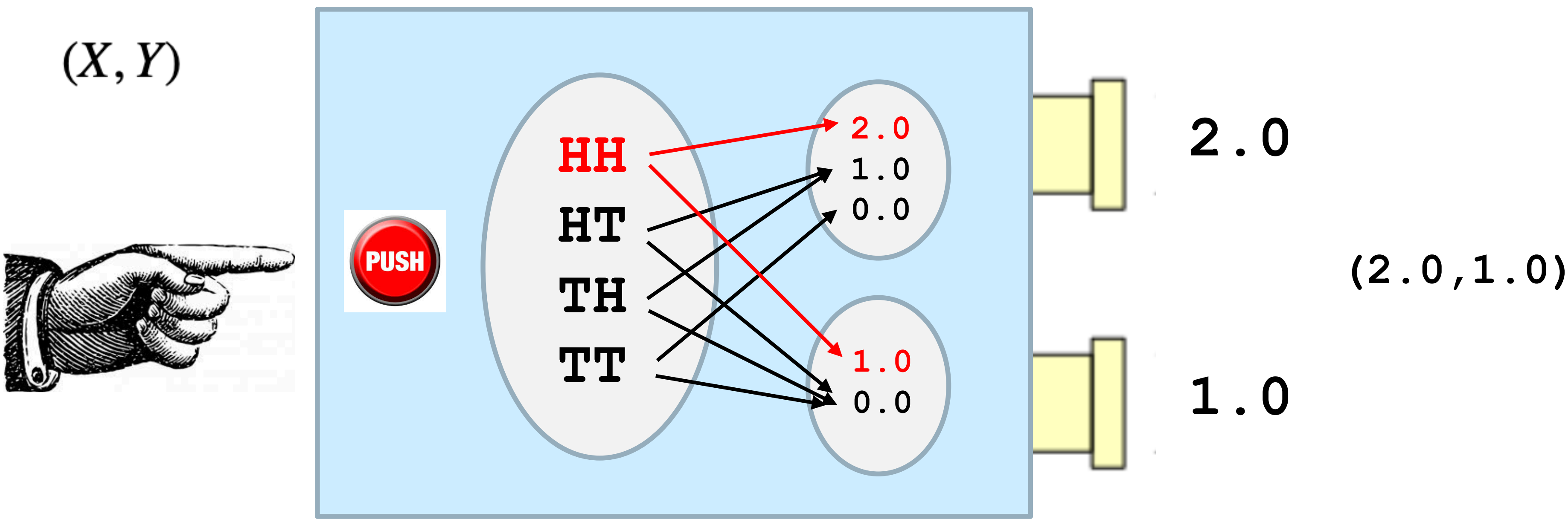
Joint Random Variables

A Joint Random Variable is a pair of random variables:

$$(X, Y) \; : \; S \rightarrow \mathcal{R} \times \mathcal{R}$$

Now when an outcome is requested, the sample point is translated into two real numbers by the action of each random variable responding to the same experiment:

Throw two dice: X = "the number of heads showing," and Y = "1 if both tosses are heads, 0 otherwise."

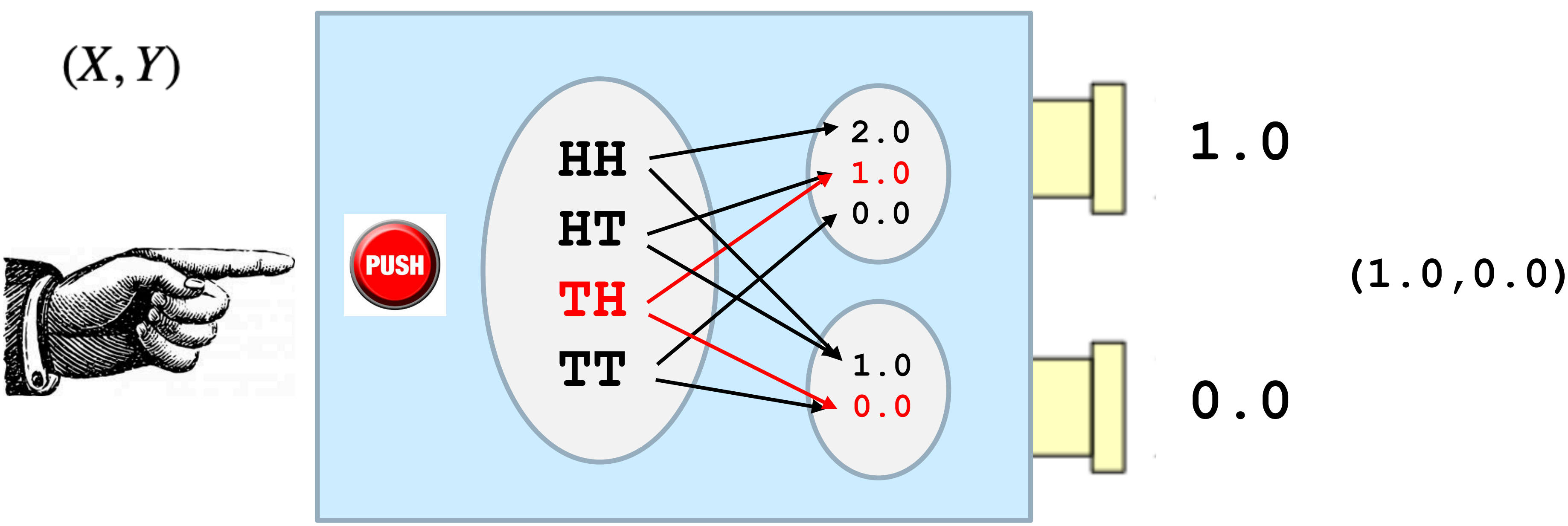


Joint Random Variables

Since the sample space can be just about anything, there is wide latitude in creating the random variables and their relationship (or lack thereof):

They may be obviously **dependent**:

Throw two dice: X = "the number of heads showing on both coins," and Y = "the number of heads showing on the first coin."

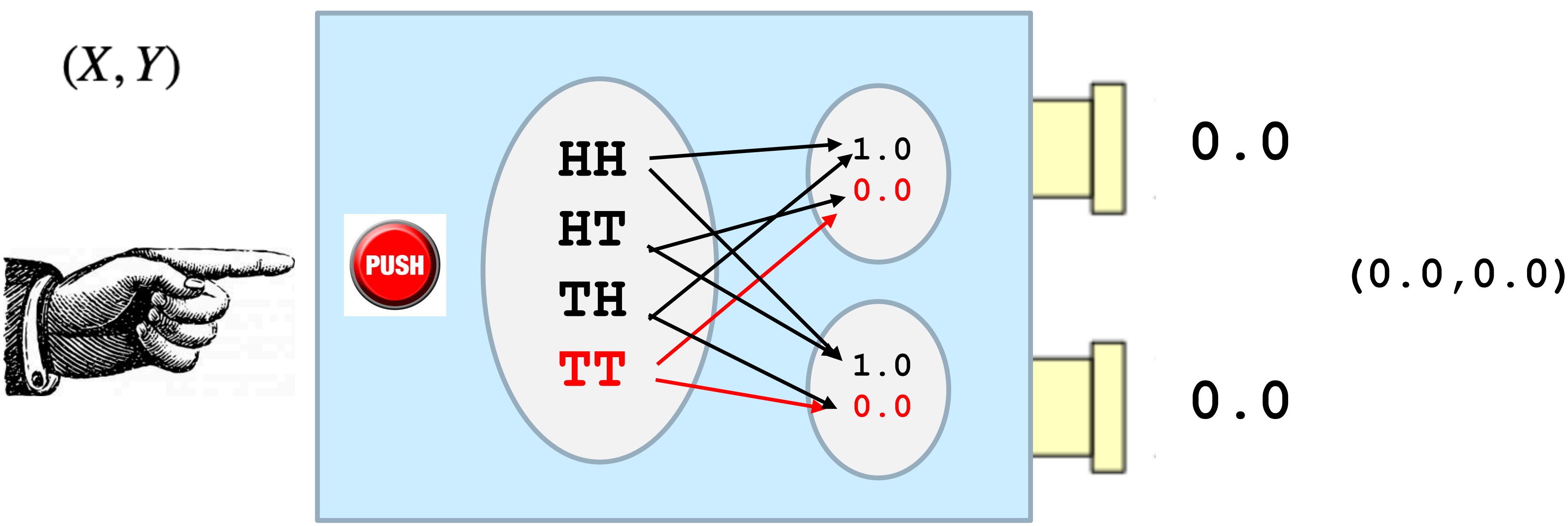


Joint Random Variables

Since the sample space can be just about anything, there is wide latitude in creating the random variables and their relationship (or lack thereof):

They may be obviously **independent**:

Throw two dice: X = "the number of heads showing on the second coin, and
Y = "the number of heads showing on the first coin."



Joint Random Variables

A joint random variable (X,Y) is called discrete if both X and Y are discrete, and continuous if both X and Y are continuous. Other combinations are possible, but we will only consider these two.

Probability Mass Function for a Discrete JRV (X,Y):

The probability that X produces value j and Y produces value k is:

$$f_{X,Y}(j,k) = P(X = j, Y = k)$$

Example 1: Toss 2 coins; X = # heads on first coin, Y = # heads on second

Sample Space	X	Y	
T T	0	0	$f_{X,Y}(0,0) = 0.25$
T H	0	1	$f_{X,Y}(0,1) = 0.25$
H T	1	0	$f_{X,Y}(1,0) = 0.25$
H H	1	1	$f_{X,Y}(1,1) = 0.25$

Joint Random Variables: Joint Probability Function

$$f_{X,Y}(j,k) \;=\; P(X=j, Y=k)$$

Example 1: Toss 2 coins; X = # heads on first coin, Y = # heads on second

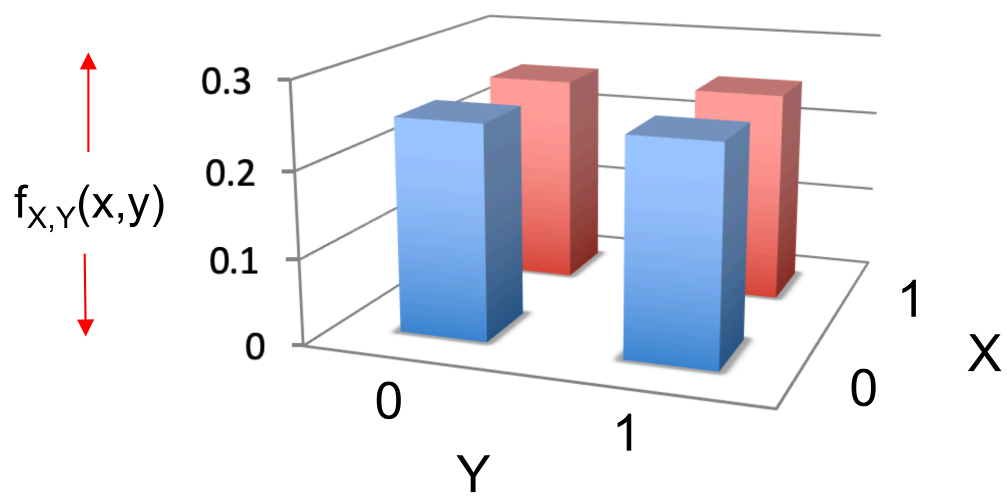
Sample Space	X	Y
T T	0	0
T H	0	1
H T	1	0
H H	1	1

$f_{X,Y}(0,0) = 0.25$
$f_{X,Y}(0,1) = 0.25$
$f_{X,Y}(1,0) = 0.25$
$f_{X,Y}(1,1) = 0.25$

Note:
Probabilities are volumes!

		Y: 0	Y: 1
X	**1**	0.25	0.25
	0	0.25	0.25

$f_{X,Y}(0,1)$



Joint Random Variables: Joint Probability Function

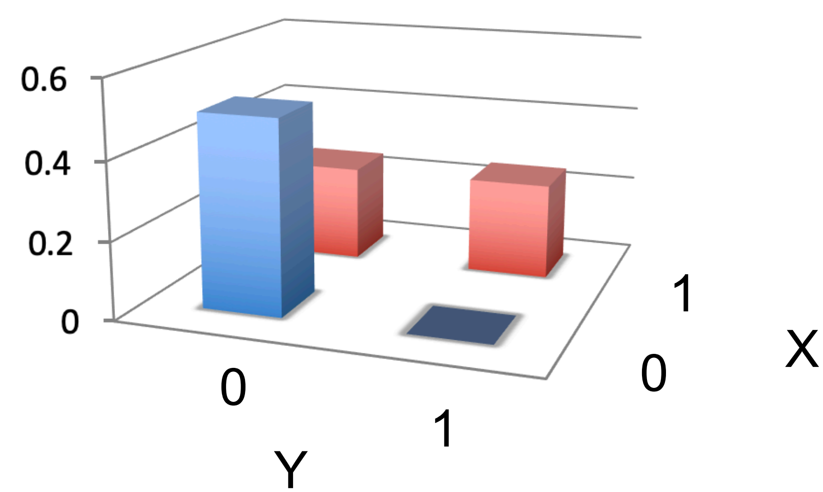
$$f_{X,Y}(j,k) = P(X = j, Y = k)$$

Example 2: Toss 2 coins; X = # heads on first coin, Y = 1 if 2 heads, 0 else

Sample Space	X	Y
T T	0	0
T H	0	0
H T	1	0
H H	1	1

$f_{X,Y}(0,0) = 0.5$
$f_{X,Y}(0,1) = 0.0$
$f_{X,Y}(1,0) = 0.25$
$f_{X,Y}(1,1) = 0.25$

		Y	
		0	**1**
X	**1**	0.25	0.25
	0	0.5	0

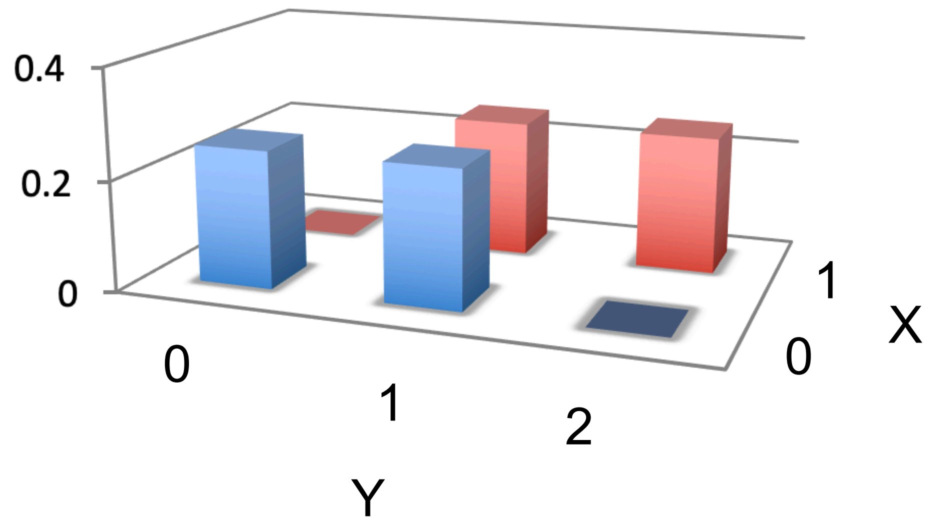


Joint Random Variables: Joint Probability Function

$$f_{X,Y}(j,k) = P(X = j, Y = k)$$

Example 3: Toss 2 coins; X = # heads on first coin, Y = total # of heads

<u>Sample Space</u>	<u>X</u>	<u>Y</u>
TT	0	0
TH	0	1
HT	1	1
HH	1	2



		Y		
		0	**1**	**2**
X	**1**	0	0.25	0.25
	0	0.25	0.25	0

Joint Random Variables: Joint Probability Function

Question: Toss 2 coins; X = 1 if 2 heads, 0 else, Y = total number of heads

What is the joint probability chart for this Joint Random Variables?

Toss 1	Toss 2	X	Y
H	H	1	2
H	T	0	1
T	H	0	1
T	T		

Y

X		**0**	**1**	**2**	$p(x_i)$
	1				
	0				
	$p(y_j)$:				

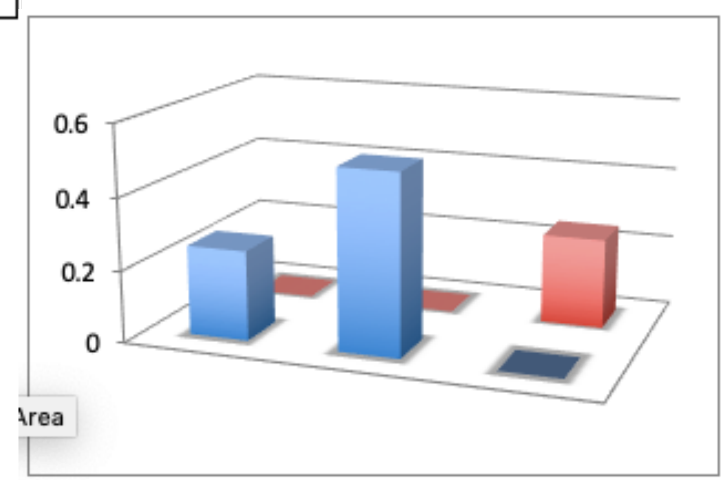
Joint Random Variables: Joint Probability Function

Question: Toss 2 coins; X = 1 if 2 heads, 0 else, Y = total number of heads

What is the joint probability chart for this Joint Random Variables?

Toss 1	Toss 2	X	Y
H	H	1	2
H	T	0	1
T	H	0	1
T	T	0	0

X \ Y	0	1	2	$p(x_i)$
1	0	0	0.25	0.25
0	0.25	0.5	0	0.75
$p(y_j)$:	0.25	0.5	0.25	1



Joint Random Variables: Marginal Distributions

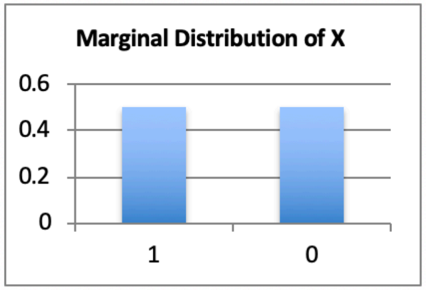
The Marginal Distributions of a Joint Random Variable are the individual random variables, considered separately from each other:

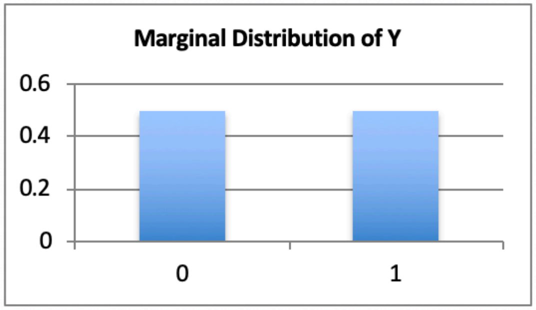
$$f_X(j) = P(X=j) = \sum_{k \in R_Y} f_{X,Y}(j,k) \qquad f_Y(k) = P(Y=k) = \sum_{j \in R_X} f_{X,Y}(j,k)$$

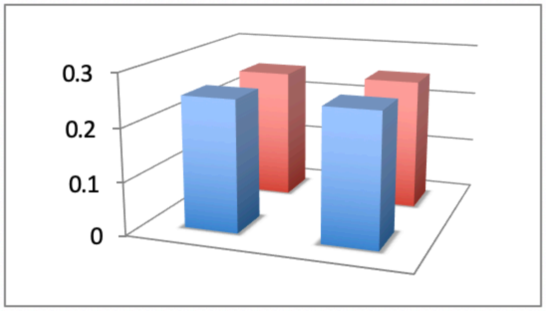
Example 1: Toss 2 coins;
X = # heads on first coin,
Y = # heads on second

Sample Space	X	Y
T T	0	0
T H	0	1
H T	1	0
H H	1	1

	Y = 0	Y = 1	
X = 1	0.25	0.25	0.5
X = 0	0.25	0.25	0.5
	0.5	0.5	



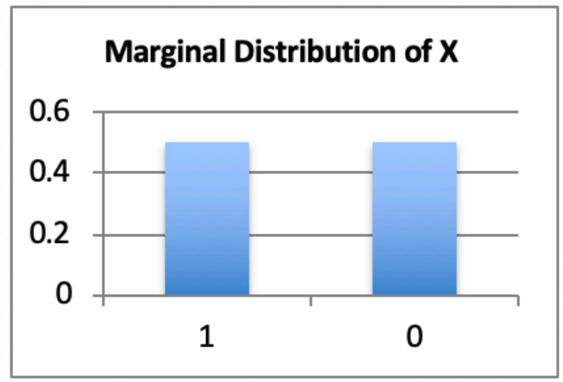


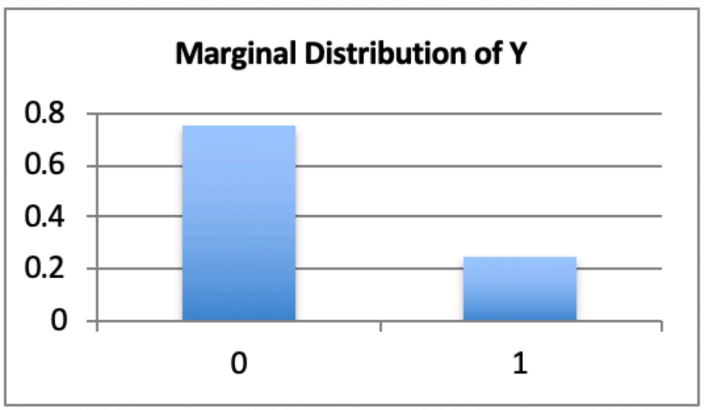


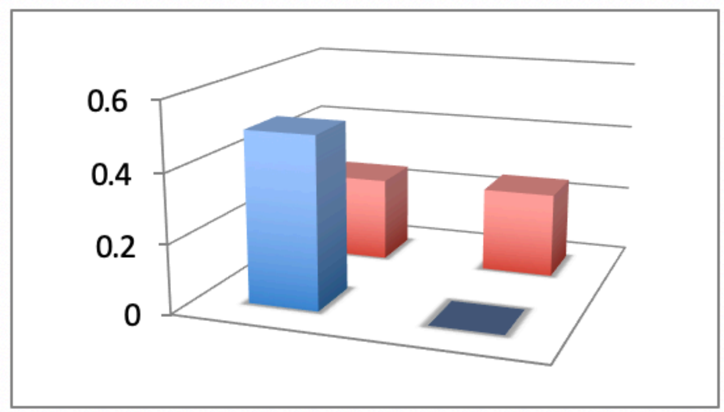
Joint Random Variables: Marginal Distributions

Example 2: Toss 2 coins; X = # heads on first coin, Y = 1 if 2 heads, 0 else

	Y		
X	0	1	
1	0.25	0.25	0.5
0	0.5	0	0.5
	0.75	0.25	



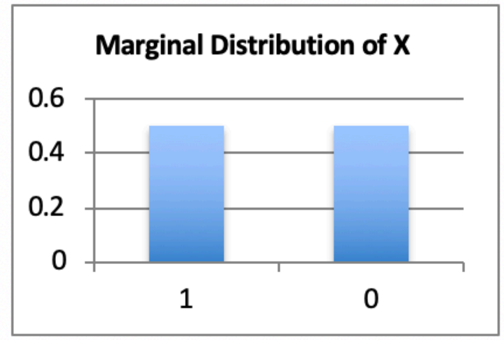


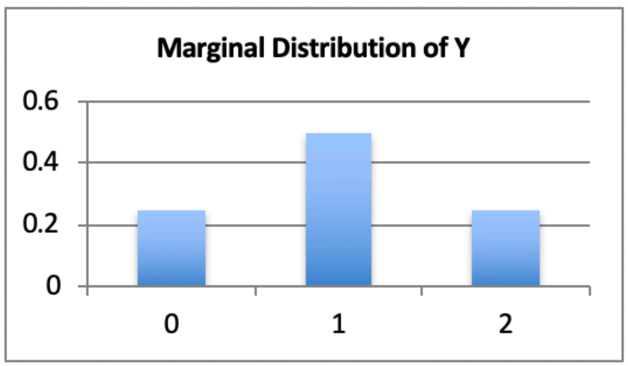


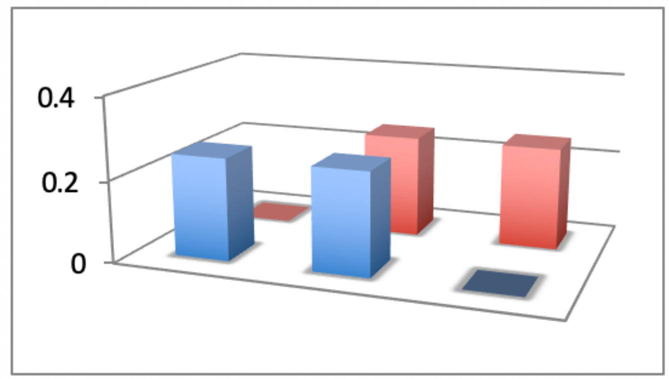
Joint Random Variables: Marginal Distributions

Example 3: Toss 2 coins; X = # heads on first coin, Y = total # of heads

		Y: 0	Y: 1	Y: 2	
X	1	0	0.25	0.25	0.5
	0	0.25	0.25	0	0.5
		0.25	0.5	0.25	







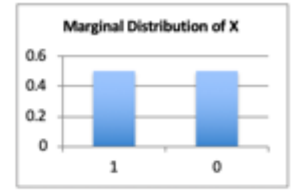
$X \sim B(2,0.5)$ $\quad E(X) = 1$ $\quad Var(X) = 2*0.5*0.5 = 0.5$ $\quad \sigma_X = 0.717$

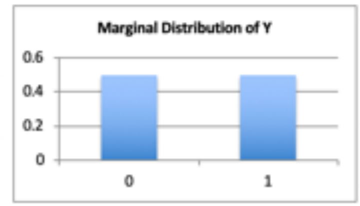
$Y \sim Bern(0.5)$ $\quad E(Y) = 1$ $\quad Var(Y) = 0.5*0.5 = 0.25$ $\quad \sigma_Y = 0.5$

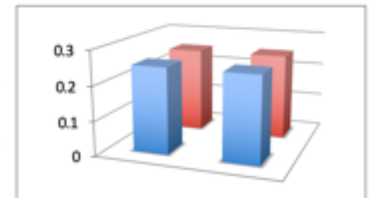
Joint Random Variables: Marginal Distributions

We will mostly concern ourselves with the bivariate case (two RVs), and in lab next week we will study ways of displaying 2D data.

X \ Y	0	1	
1	0.25	0.25	0.5
0	0.25	0.25	0.5
	0.5	0.5	

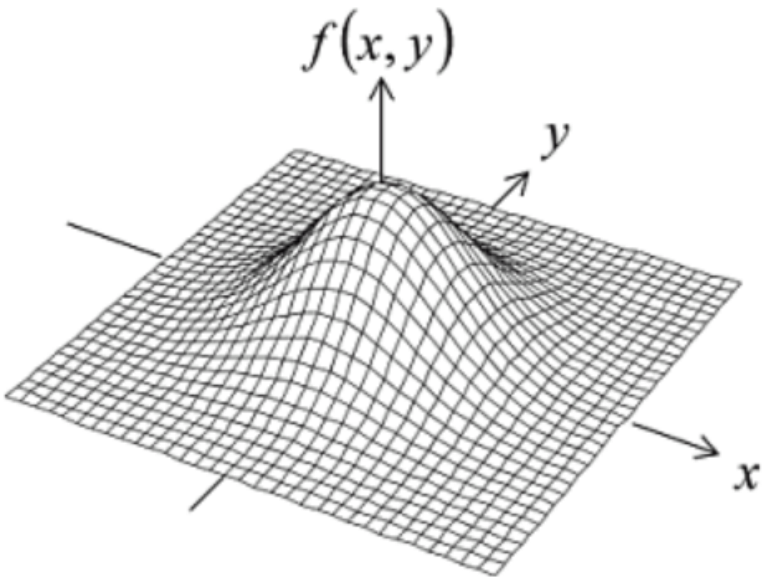


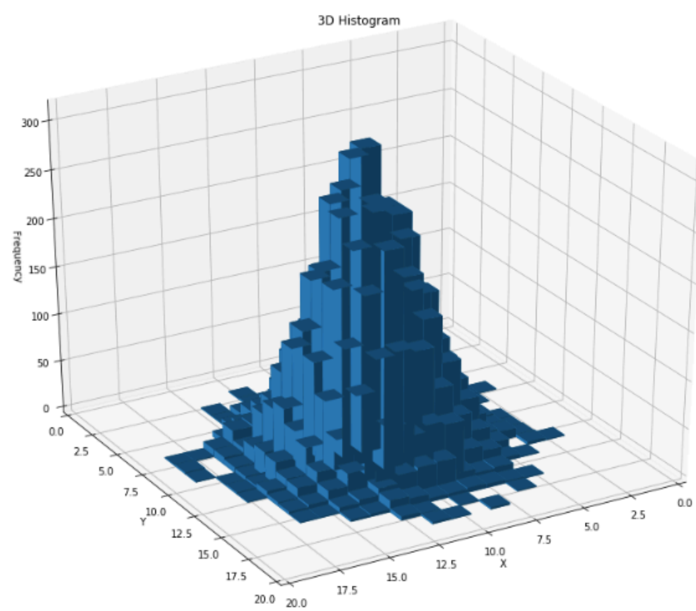


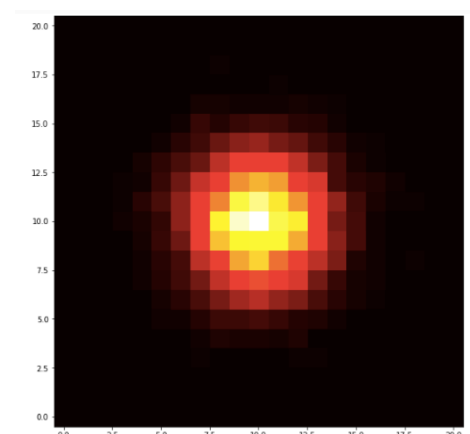


The main insight you need for the 2D case is that now,

- Probabilities are volumes; and
- The volume of a probability space must be 1.0.







Joint Random Variables: The Continuous Ca

We will not do much with the continuous case, but the modifications are straight-forward (must use 2D intervals, replace sums with integrals).

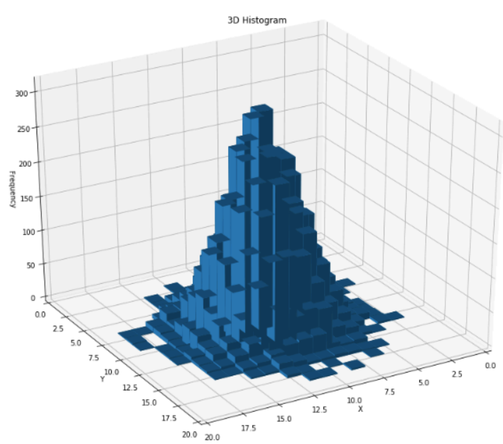
Discrete Case (can use PDF)

$$f_{X,Y}(x, y) = P(X = x, Y = y)$$

$$F_{X,Y}(x, y) = P(X \leq x, Y \leq y)$$

$$f_X(x) = \Sigma_{y \in R_Y} P(X = x, y)$$

$$f_Y(y) = \Sigma_{x \in R_X} P(x, Y = y)$$



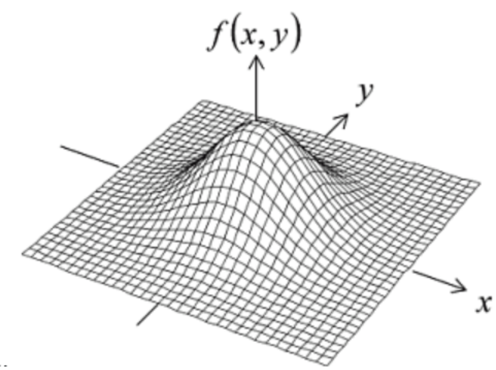
Continuous Case (must use CDF)

$$f_{X,Y}(x, y) = P(X = x, Y = y)$$

$$F_{X,Y}(x, y) = P(X \leq x, Y \leq y)$$

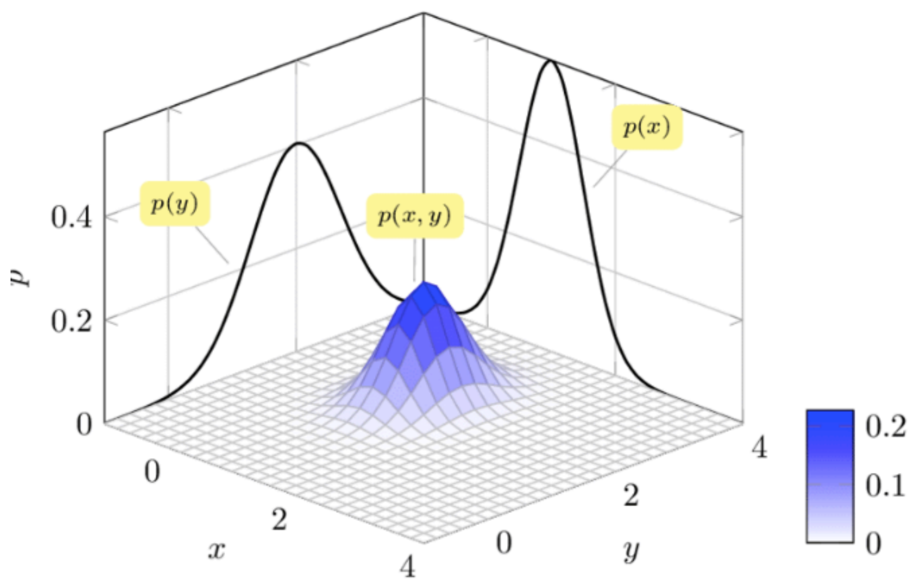
$$f_X(x) = \int_{y \in R_Y} P(X = x, y)$$

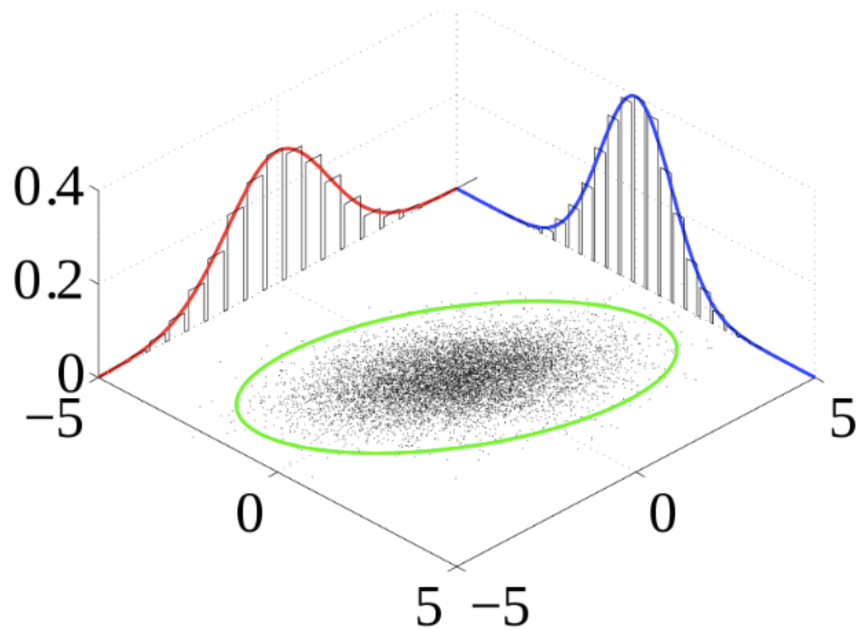
$$f_Y(y) = \int_{x \in R_X} P(x, Y = y)$$

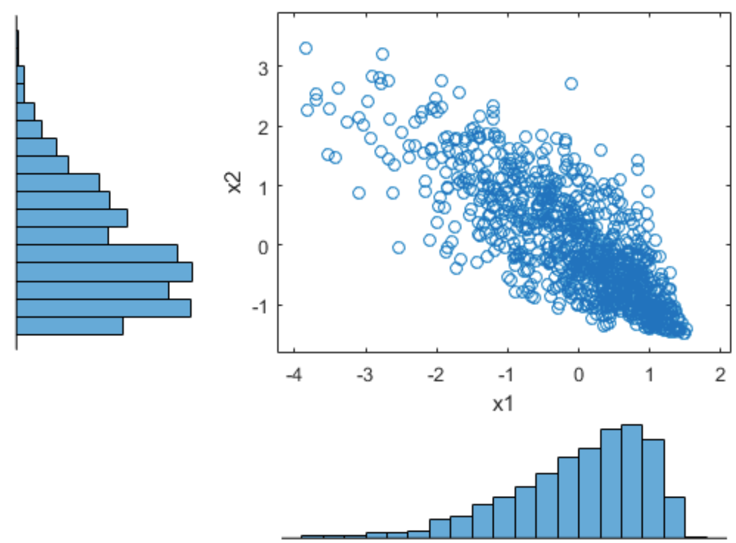


Joint Random Variables: The Continuous Case

It is often useful to display the marginal distributions along with the joint distribution:

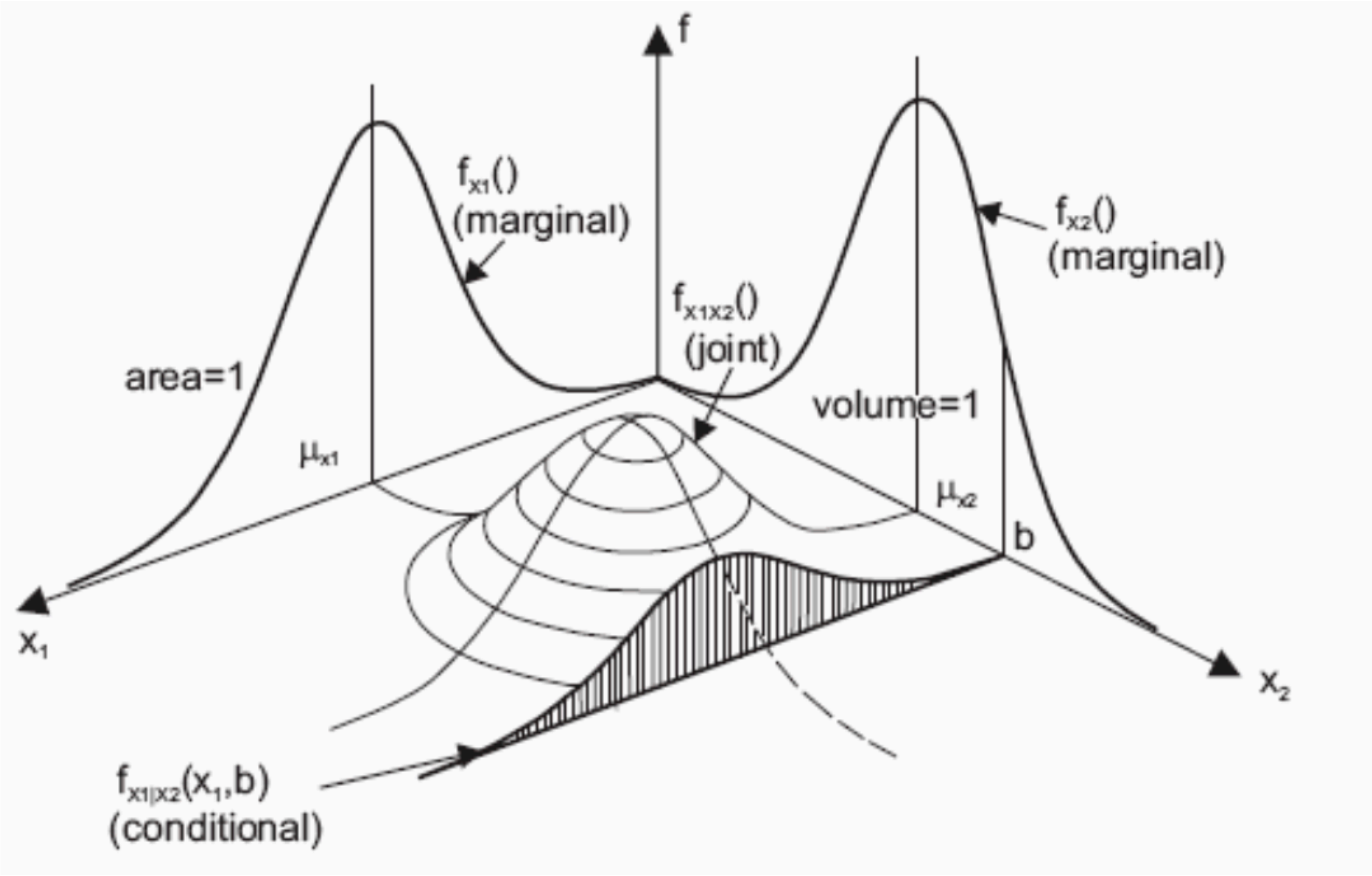






Joint Random Variables: The Continuous Case

It is often useful to display the marginal distributions along with the joint distribution:

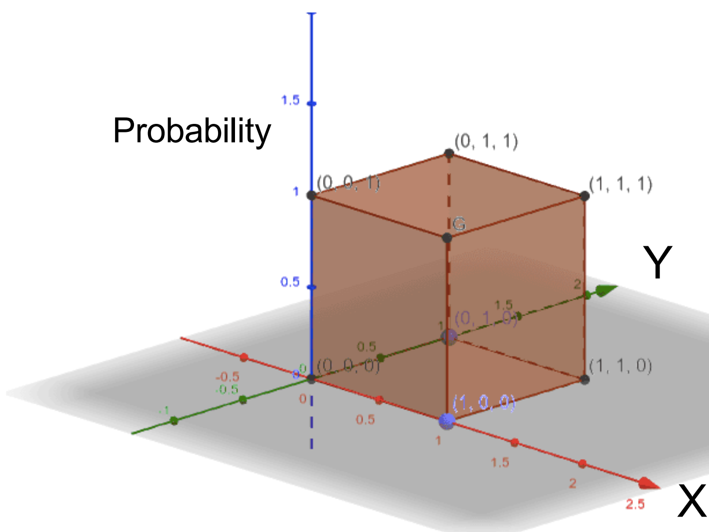


Joint Random Variables: The Continuous Case

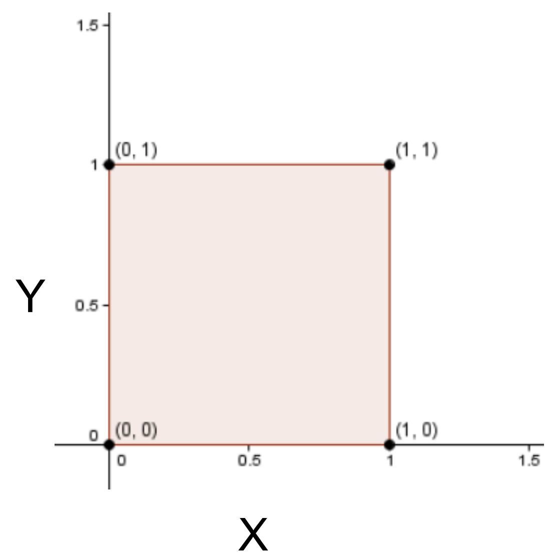
Example: Bivariate Uniform Distribution (X,Y)

```
def uniform2D():
     return  ( random(), random() )
```

PDF is a unit cube of volume 1.0:

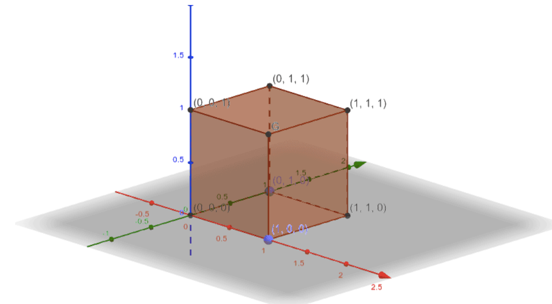


But in the uniform case it can be viewed from ABOVE as a unit square:

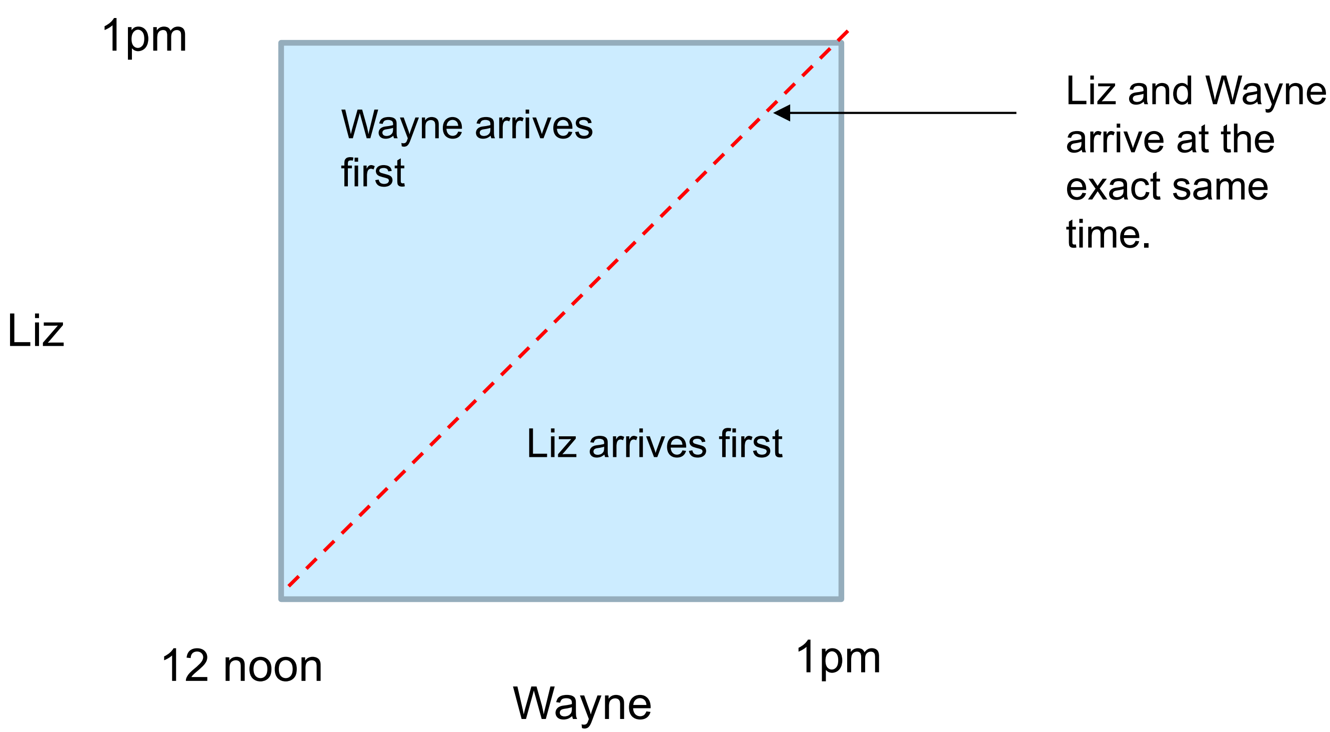


Joint Random Variables: The Continuous Case

Question: Suppose Liz and Wayne decide to meet at Starbucks sometime between 12 noon and 1pm, and each of them shows up uniformly at random in that interval.What is the probability that Wayne shows up before Liz?

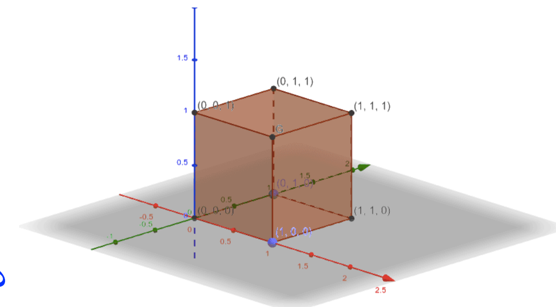


You might guess it is ½, but how to prove it?

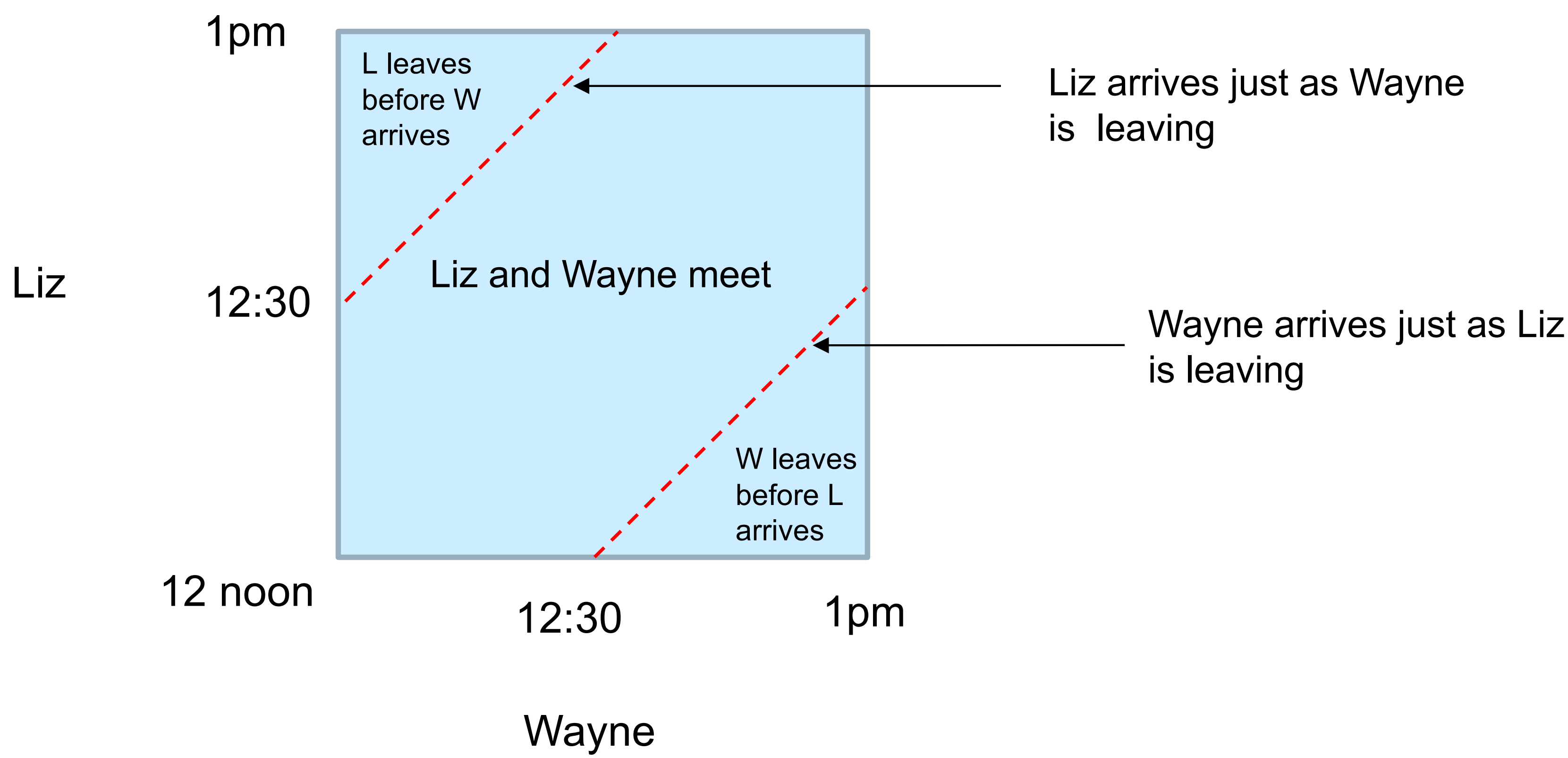


Joint Random Variables: The Continuous Case

Question: Suppose Liz and Wayne decide to meet at Starbucks sometime between 12 noon and 1pm, and each of them shows up uniformly at random in that interval. Each waits at most 30 minutes until the other one shows or it is 1pm, and then leaves. What is the probability that they do meet (don't miss each other)?



You might guess it is 3/4, but how to prove it?



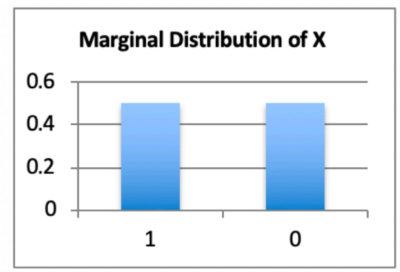
Joint Random Variables: Conditional Probability

We can easily define the notion of conditional probability, using the expected definition; e.g., the probability that X returns 1, given that we know that Y >= 1, is:

$$P(X = 1 \mid Y \geq 1) = \frac{P(X = 1 \text{ and } Y \geq 1)}{P(Y \geq 1)} = \frac{0.5}{0.75} = \frac{2}{3}$$

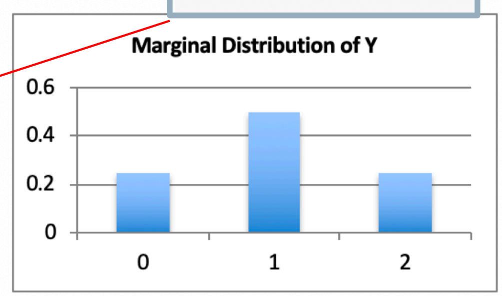
Example 3: Toss 2 coins; X = # heads on first coin, Y = total # of heads

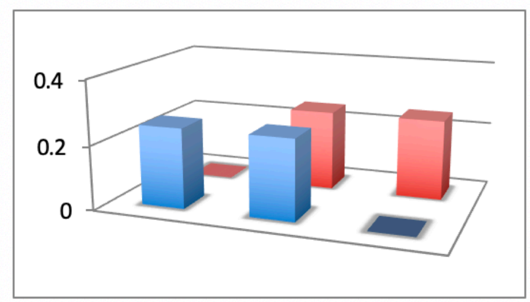
X \ Y	0	1	2	
1	0	0.25	0.25	0.5
0	0.25	0.25	0	0.5
	0.25	0.5	0.25	



Conditioned by $P(Y \geq 1)$:

X' \ Y'	1	2	
	1/3	1/3	2/3
	1/3	0	1/3
	2/3	1/3	





$X \sim B(2,0.5)$ $E(X) = 1$ $Var(X) = 2*0.5*0.5 = 0.5$ $\sigma_X = 0.717$

$Y \sim Bern(0.5)$ $E(Y) = 1$ $Var(Y) = 0.5*0.5 = 0.25$ $\sigma_Y = 0.5$

Joint Random Variables: Independence

Again, we can easily define the notion of independence, using the expected definition; e.g., two random variables are independent if and only if

$$f_{X,Y}(j,k) = f_X(j) * f_Y(k)$$

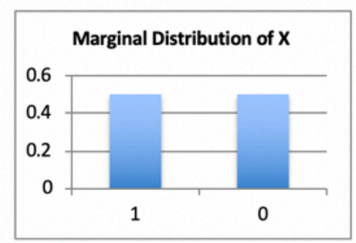
That is, each joint probability is the product of the marginal probabilities.

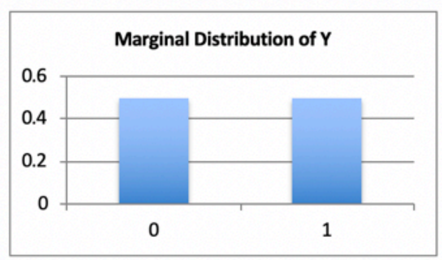
INDEPENDENT:

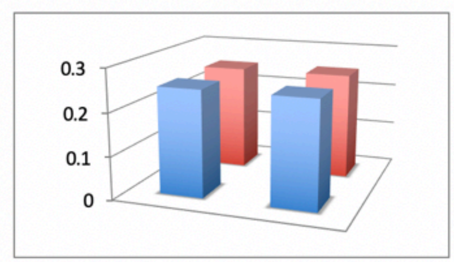
Example 1: Toss 2 coins;
X = # heads on first coin,
Y = # heads on second

Sample Space	X	Y
T T	0	0
T H	0	1
H T	1	0
H H	1	1

		Y = 0	Y = 1	
X	1	0.25	0.25	0.5
	0	0.25	0.25	0.5
		0.5	0.5	







Joint Random Variables: Independence

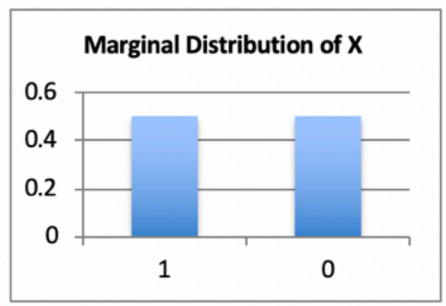
Again, we can easily define the notion of independence, using the expected definition; e.g., two random variables are independent if and only if

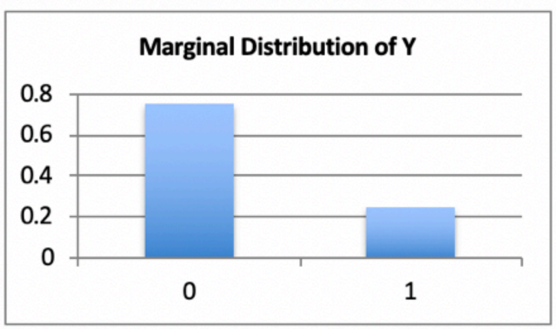
$$f_{X,Y}(j,k) = f_X(j) * f_Y(k)$$

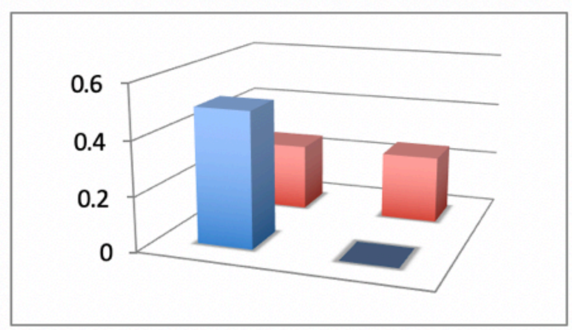
That is, each joint probability is the product of the marginal probabilities.

DEPENDENT:

	Y = 0	Y = 1	
X = 1	0.25	0.25	0.5
X = 0	0.5	0	0.5
	0.75	0.25	







Joint Random Variables

Again, we can easily define the notion of independence, using the expected definition; e.g., two random variables are independent if and only if

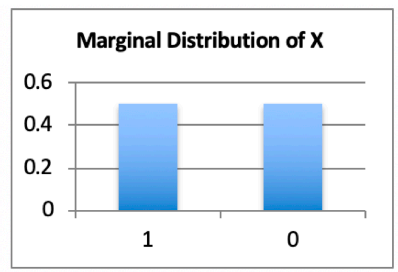
$$f_{X,Y}(j,k) = f_X(j) * f_Y(k)$$

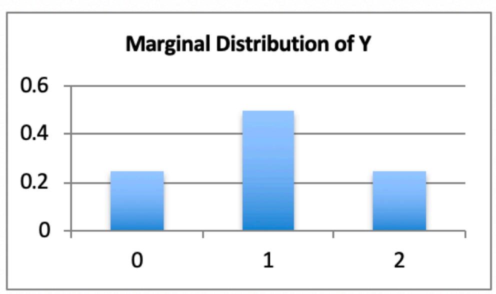
That is, each joint probability is the product of the marginal probabilities.

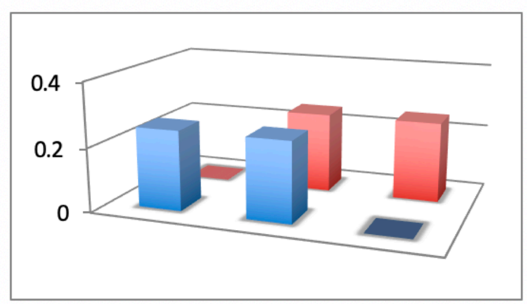
DEPENDENT

Example 3: Toss 2 coins; X = # heads on first coin, Y = total # of heads

		Y = 0	Y = 1	Y = 2	
X	1	0	0.25	0.25	0.5
	0	0.25	0.25	0	0.5
		0.25	0.5	0.25	







Joint Random Variables

Again, we can easily define the notion of independence, using the expected definition; e.g., two random variables are independent if and only if

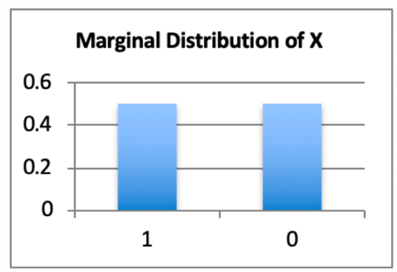
$$f_{X,Y}(j,k) = f_X(j) * f_Y(k)$$

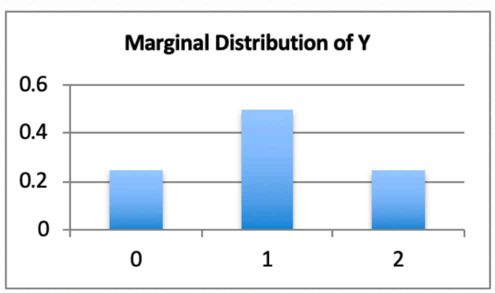
That is, each joint probability is the product of the marginal probabilities.

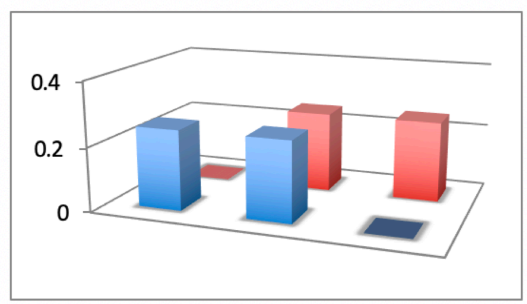
Dependent!

Example 3: Toss 2 coins; X = # heads on first coin, Y = total # of heads

		Y = 0	Y = 1	Y = 2	
X	1	0	0.25	0.25	0.5
	0	0.25	0.25	0	0.5
		0.25	0.5	0.25	







Joint Random Variables: Covariance

The notion of variance must be changed, however, and this will lead to another important topic!

Recall: The Variance of X is defined b y:

$$Var(X) = E[\,(X - \mu_X\,)^2\,]$$

$$= E(X^2) - \mu_X^2$$

The Covariance of two RV X and Y is defined as follows:

$$Cov(X, Y) = E[\,(X - \mu_X\,) * (Y - \mu_Y\,)\,]$$

$$= E(X * Y) - \mu_X * \mu_Y$$

Joint Random Variables: Correlation Coefficient

The Covariance of two RV X and Y has the same defects as the variance of a single RV:

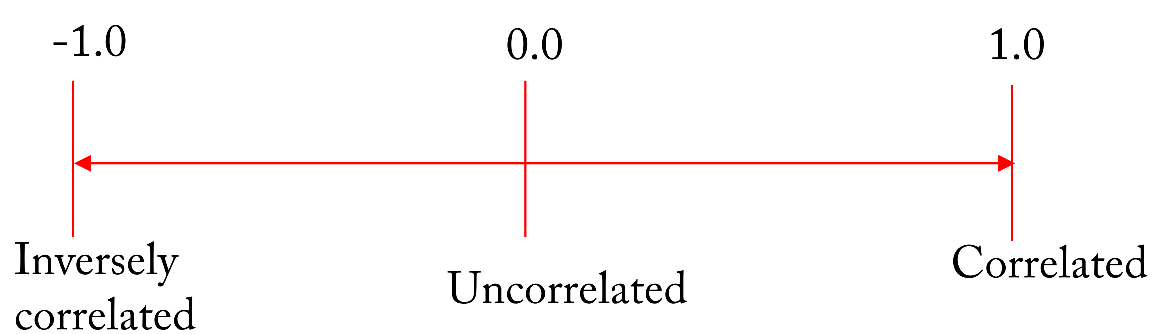
- The units are the product of the units of X and Y: if X = height and Y = weight, then the units might be foot-pounds!
- The scale is hard to work with: What does a coveriance of -123.4 mean?

Therefore we standardize the covariance to the interval [-1 .. 1]:

The Correlation Coefficient of X and Y is defined as

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X * \sigma_Y}$$

The range is from -1 to 1:



Joint Random Variables: Correlation Coefficient

The Correlation Coefficient of X and Y is defined as

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X * \sigma_Y} = \frac{E[\,(X-\mu_X\,)*(Y-\mu_Y\,)}{\sigma_X * \sigma_Y} = E\left[\frac{X-\mu_X}{\sigma_X} * \frac{Y-\mu_Y}{\sigma_X}\right] = E[\,Z_X * Z_Y\,]$$

where Z_X and Z_Y are the standardized forms of X and Y.

Usually we do this using Python, but if you calculate by hand, you should use the formula:

$$Cov(X,Y) = E[\,(X-\mu_X\,)*(Y-\mu_Y\,)\,]$$

$$= E(X*Y) - \mu_X * \mu_Y$$

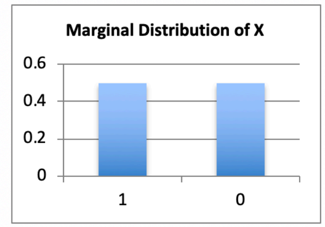
Joint Random Variables

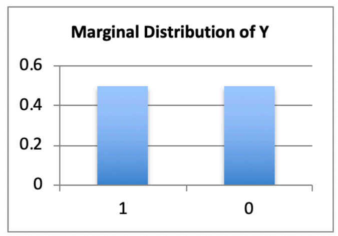
Example 1: Toss 2 coins; X = # heads on first, Y = # heads on second

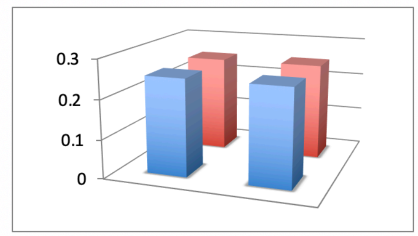
Joint Distribution for X and Y

Ex 1: toss 2 coins: X = # heads first, Y = # heads second

					Y		0.50	:σ_Y
				$(y_j\text{-}mu_Y)$^2 * $p(y_j)$:	0.125	0.125	0.25	:var(X)
				$(y_j\text{-}mu_Y)$:	-0.50	0.50		
			$p(x_i,y_j)$	y_j*$p(y_j)$:	0.00	0.50	0.50	:μ_Y
	$(x_i\text{-}mu_X)^2$ * $p(x_i)$	$(x_i\text{-}mu_X)$	x_i*$p(x_i)$		**0**	**1**	$p(x_i)$	
X	0.125	0.50	0.50	**1**	0.25	0.25	0.5	
	0.125	-0.50	0.00	**0**	0.25	0.25	0.5	
var(X):	0.25	μ_X:	0.50	$p(y_j)$:	0.5	0.5	1	
σ_X:	0.50						0.25	:σ_X * σ_Y







$(x_i\text{-}mu_X)$*$(y_j\text{-}mu_Y)$*$p(x_i,y_j)$:

-0.0625	0.0625
0.0625	-0.0625

cov(X,Y): 0.00

ρ(X,Y): 0.00

	0	1
1	0 / 0.25	1 / 0.25
0	0 / 0.25	0 / 0.25

$$E(X * Y) = 0 * 0.25 + 1 * 0.25 + 0 * 0.25 + 0 * 0.25 = 0.25$$

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X * \sigma_Y} = \frac{E(X * Y) - \mu_X * \mu_Y}{\sigma_X * \sigma_Y} = \frac{0.25 - 0.5 * 0.5}{0.5 * 0.5} = 0.0$$

Joint Random Variables

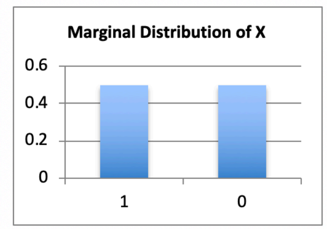
Calculating the Covariance and Correlation Coefficient is easier if you either (i) use Python, or (ii) use the short-hand formula:

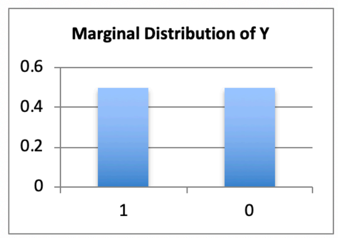
Example: Toss 2 coins; X = # heads on first, Y = # heads on second

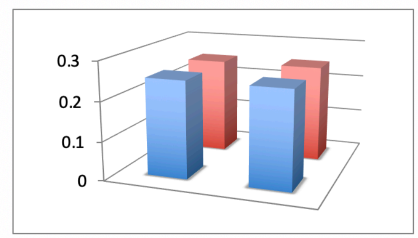
Joint Distribution for X and Y

Ex 1: toss 2 coins: X = # heads first, Y = # heads second

					Y			
							0.50	$:\sigma_Y$
				(y_j-mu_Y)^2 * $p(y_j)$:	0.125	0.125	0.25	:var(X)
				(y_j-mu_Y):	-0.50	0.50		
			p(x_i,y_j)	y_j*$p(y_j)$:	0.00	0.50	0.50	$:\mu_Y$
	$(x_i-mu_X)^2$ * $p(x_i)$	(x_i-mu_X)	x_i*$p(x_i)$		**0**	**1**	$p(x_i)$	
X	0.125	0.50	0.50	**1**	0.25	0.25	0.5	
	0.125	-0.50	0.00	**0**	0.25	0.25	0.5	
var(X):	0.25	μ_X:	0.50	$p(y_j)$:	0.5	0.5	1	
σ_X:	0.50						0.25	$:\sigma_X * \sigma_Y$







(x_i-mu_X)*(y_j-mu_Y)*$p(x_i,y_j)$:

-0.0625	0.0625
0.0625	-0.0625

cov(X,Y): 0.00

ρ(X,Y): 0.00

	0	1
1	0 / 0.25	1 / 0.25
0	0 / 0.25	0 / 0.25

$$E(X * Y) = 0 * 0.25 + 1 * 0.25 + 0 * 0.25 + 0 * 0.25 = 0.25$$

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X * \sigma_Y} = \frac{E(X * Y) - \mu_X * \mu_Y}{\sigma_X * \sigma_Y} = \frac{0.25 - 0.5 * 0.5}{0.5 * 0.5} = 0.0$$

Joint Random Variables

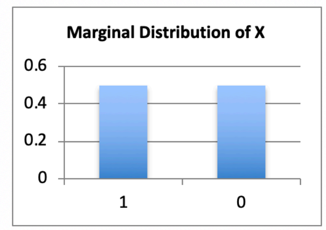
Calculating the Covariance and Correlation Coefficient is easier if you either (i) use Python, or (ii) use the short-hand formula:

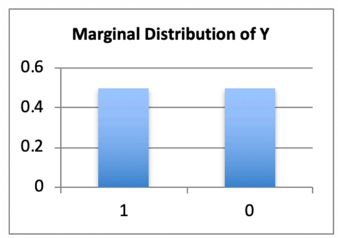
Example: Toss 1 coin; X = # heads, Y = # heads

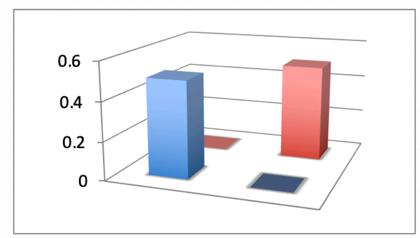
Joint Distribution for X and Y

Ex 2: toss 1 coins: X = # heads, Y = # heads

					Y			
							0.50	:σ_Y
				$(y_j\text{-mu}_Y)$^2 * $p(y_j)$:	0.125	0.125	0.25	:var(X)
				$(y_j\text{-mu}_Y)$:	-0.50	0.50		
			$p(x_i,y_j)$	y_j*$p(y_j)$:	0.00	0.50	0.50	:μ_Y
	$(x_i\text{-mu}_X)^2$ * $p(x_i)$	$(x_i\text{-mu}_X)$	x_i*$p(x_i)$		**0**	**1**	$p(x_i)$	
X	0.125	0.50	0.50	**1**	0	0.5	0.5	
	0.125	-0.50	0.00	**0**	0.5	0	0.5	
var(X):	0.25	μ_X:	0.50	$p(y_j)$:	0.5	0.5	1	
σ_X:	0.50						0.25	:σ_X * σ_Y







cov(X,Y): 0.25

ρ(X,Y): 1.00

$(x_i\text{-mu}_X)$*$(y_j\text{-mu}_Y)$*$p(x_i,y_j)$:

0.0000	0.1250
0.1250	0.0000

	0	1
1	0 0.0	1 0.5
0	0 0.5	0 0.0

$$E(X * Y) = 0 * 0.0 + 1 * 0.5 + 0 * 0.5 + 0 * 0.0 = 0.5$$

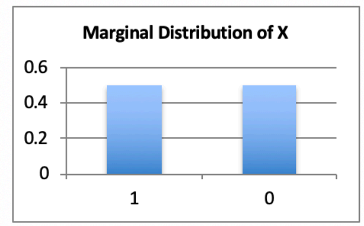
$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X * \sigma_Y} = \frac{E(X*Y) - \mu_X * \mu_Y}{\sigma_X * \sigma_Y} = \frac{0.5 - 0.5 * 0.5}{0.5 * 0.5} = \frac{0.25}{0.25} = 1.0$$

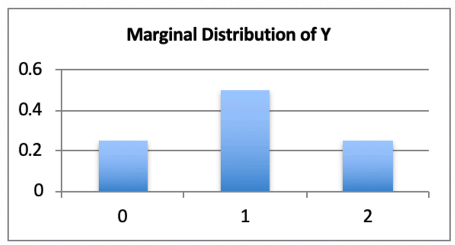
Joint Random Variables

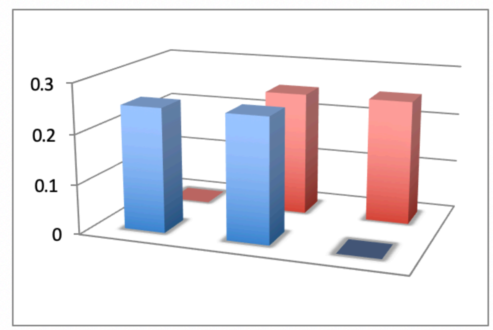
Calculating the Covariance and Correlation Coefficient is easier if you either
(i) use Python, or (ii) use the short-hand formula:

Example: Toss 2 coins; X = # heads on first coin, Y = total # of heads

						Y			
								0.71	:σ_Y
				$(y_j\text{-mu}_Y)$^2 * $p(y_j)$:	0.250	0.000	0.250	0.50	:var(X)
				$(y_j\text{-mu}_Y)$:	-1.00	0.00	1.00		
			$p(x_i,y_j)$	y_j*$p(y_j)$:	0.00	0.50	0.50	1.00	:μ_Y
	$(x_i\text{-mu}_X)^2$ * $p(x_i)$	$(x_i\text{-mu}_X)$	x_i*$p(x_i)$		**0**	**1**	**2**	$p(x_i)$	
X	0.125	0.50	0.50	**1**	0	0.25	0.25	0.5	
	0.125	-0.50	0.00	**0**	0.25	0.25	0	0.5	
var(X):	0.25	μ_X:	0.50	$p(y_j)$:	0.25	0.5	0.25	1	
σ_X:	0.50							0.35	:σ_X * σ_Y







				$(x_i\text{-mu}_X)$*$(y_j\text{-mu}_Y)$*$p(x_i,y_j)$:		
cov(X,Y):	0.25			0.0000	0.0000	0.1250
ρ(X,Y):	0.707			0.1250	0.0000	0.0000

	0	1	2
1	0 / 0.0	1 / 0.25	2 / 0.25
0	0 / 0.25	0 / 0.25	0 / 0.0

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X * \sigma_Y} = \frac{E(X*Y) - \mu_X * \mu_Y}{\sigma_X * \sigma_Y} = \frac{0.75 - 0.5 * 1.0}{0.5 * 0.707} = \frac{0.25}{0.354} = 0.707$$