Probability Review and Statistical Estimation

Random Variables

- Informally, A is a random variable if
  - A denotes something about which we are uncertain
  - perhaps the outcome of a randomized experiment

Random Variables

- Informally, A is a <u>random variable</u> if
  - A denotes something about which we are uncertain
  - perhaps the outcome of a randomized experiment

- Examples

  A = True if a randomly drawn person from our class is female

  A = The hometown of a randomly drawn person from our class

  A = True if two randomly drawn persons from our class have same birthday

Random Variables

- Informally, A is a random variable if
  - A denotes something about which we are uncertain
  - perhaps the outcome of a randomized experiment

- Examples

  A = True if a randomly drawn person from our class is female

  A = The hometown of a randomly drawn person from our class

  A = True if two randomly drawn persons from our class have same birthday

- Define P(A) as
  - the fraction of possible worlds in which A is true

    or
  - the fraction of times A holds: in repeated runs of the random experiment

Random Variables

- Informally, A is a <u>random variable</u> if
  - A denotes something about which we are uncertain
  - perhaps the outcome of a randomized experiment

- Examples

  A = True if a randomly drawn person from our class is female

  A = The hometown of a randomly drawn person from our class

  A = True if two randomly drawn persons from our class have same birthday

- Define P(A) as
  - the fraction of possible worlds in which A is true

    or
  - the fraction of times A holds: in repeated runs of the random experiment
    - the set of possible worlds is called the sample space, S
    - A random variable A is a function defined over S

A: S ✐ {0,1}

A little formalism

More formally, we have

- a <u>sample space</u> S (e.g., set of students in our class)
  - aka the set of possible worlds

A little formalism

More formally, we have

- a sample space S (e.g., set of students in our class)
  - aka the set of possible worlds

- a random variable is a function defined over the sample space
  - Gender: S ✎ { m, f }
  - Height: S ✎ Reals

A little formalism

More formally, we have

- a sample space S (e.g., set of students in our class)
  - aka the set of possible worlds

- a random variable is a function defined over the sample space
  - Gender: S ✎ { m, f }
  - Height: S ✎ Reals
- an event is a subset of S
  - e.g., the subset of S for which Gender=f
  - e.g., the subset of S for which (Gender=m) AND (eyeColor=blue)

A little formalism

More formally, we have

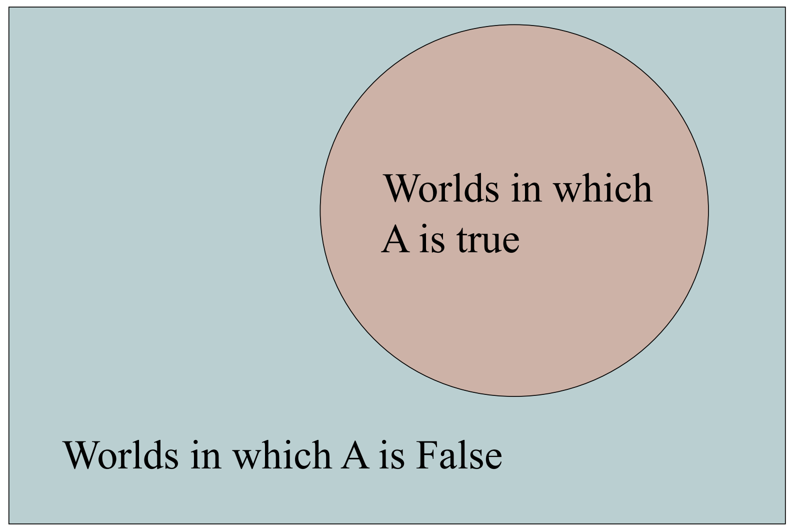
- a <u>sample space</u> S (e.g., set of students in our class)
  - aka the set of possible worlds

- a <u>random variable</u> is a function defined over the sample space
  - Gender: S ✎ { m, f }
  - Height: S ✎ Reals
- an <u>event</u> is a subset of S
  - e.g., the subset of S for which Gender=f
  - e.g., the subset of S for which (Gender=m) AND (eyeColor=blue)
- we re often interested in probabilities of specific events
- and of specific events conditioned on other specific events

Visualizing A

Sample space of all possible worlds

Its area is 1



P(A) = Area of reddish oval

The Axioms of Probability

- 0 <= P(A) <= 1
- P(True) = 1
- P(False) = 0
- P(A or B) = P(A) + P(B) - P(A and B)

The Axioms of Probability

- $0 <= P(A) <= 1$
- $P(True) = 1$
- $P(False) = 0$
- $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

[di Finetti 1931]:

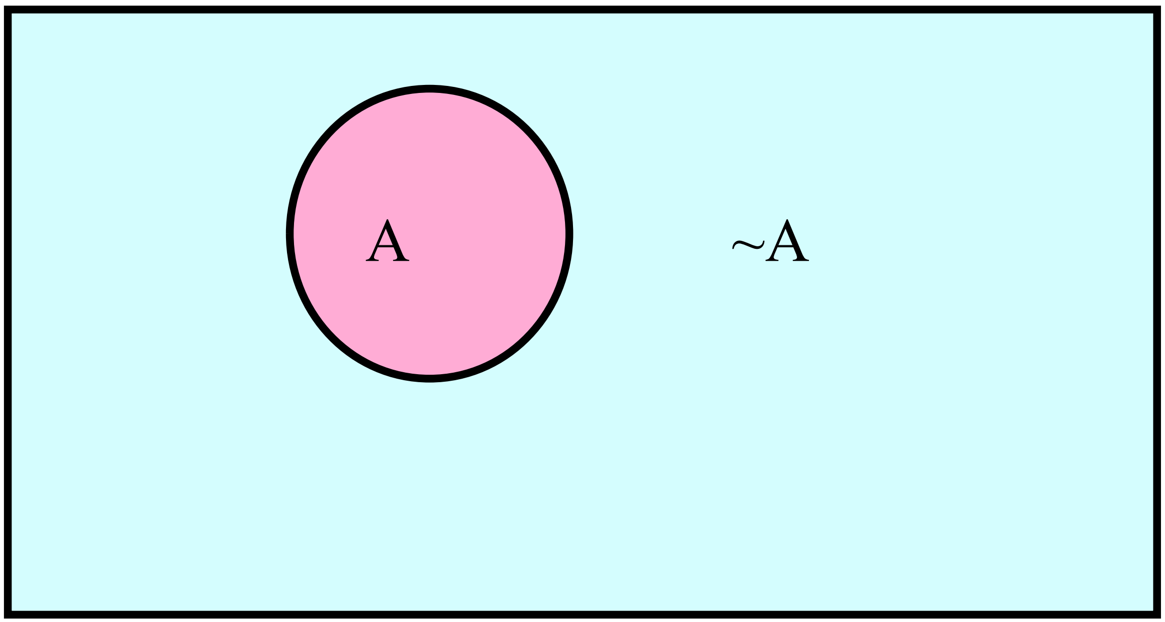
when gambling based on uncertainty formalism A you can be exploited by an opponent

iff

your uncertainty formalism A violates these axioms

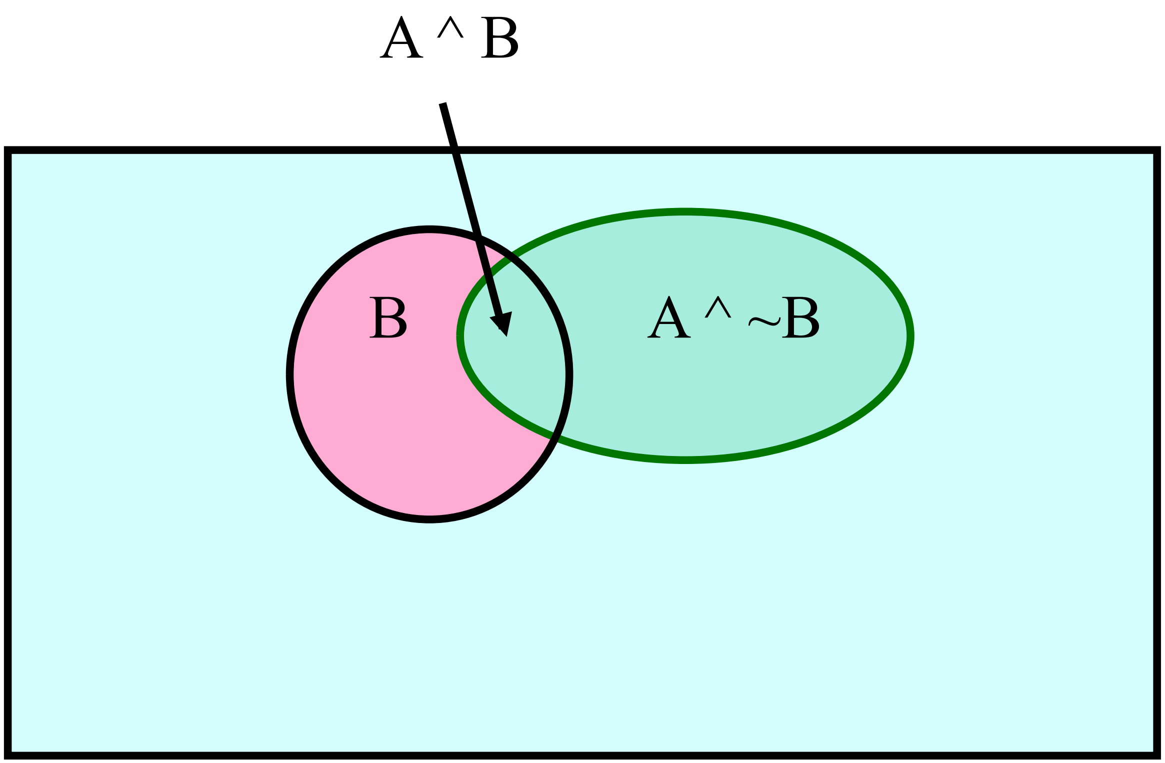
Elementary Probability in Pictures

$$P(\sim A) + P(A) = 1$$



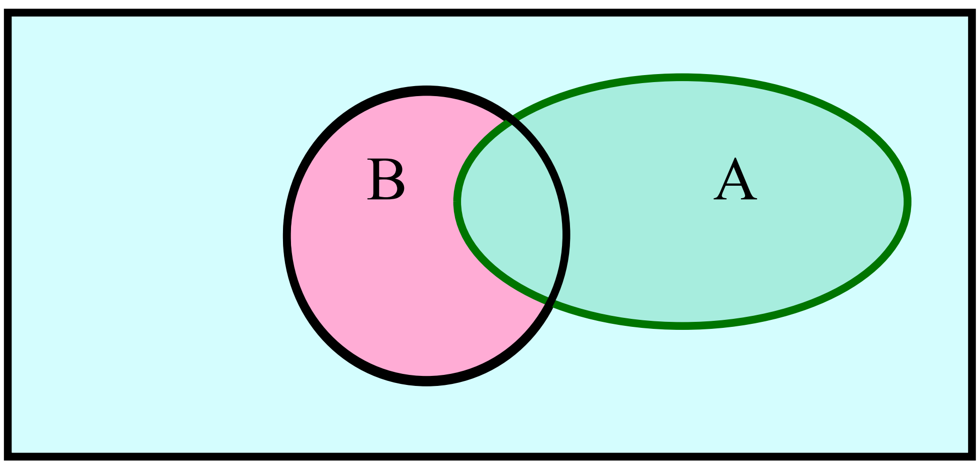
Elementary Probability in Pictures

$$P(A) = P(A \wedge B) + P(A \wedge \sim B)$$



Definition of Conditional Probability

$$P(A|B) = \frac{P(A \wedge B)}{P(B)}$$



Definition of Conditional Probability

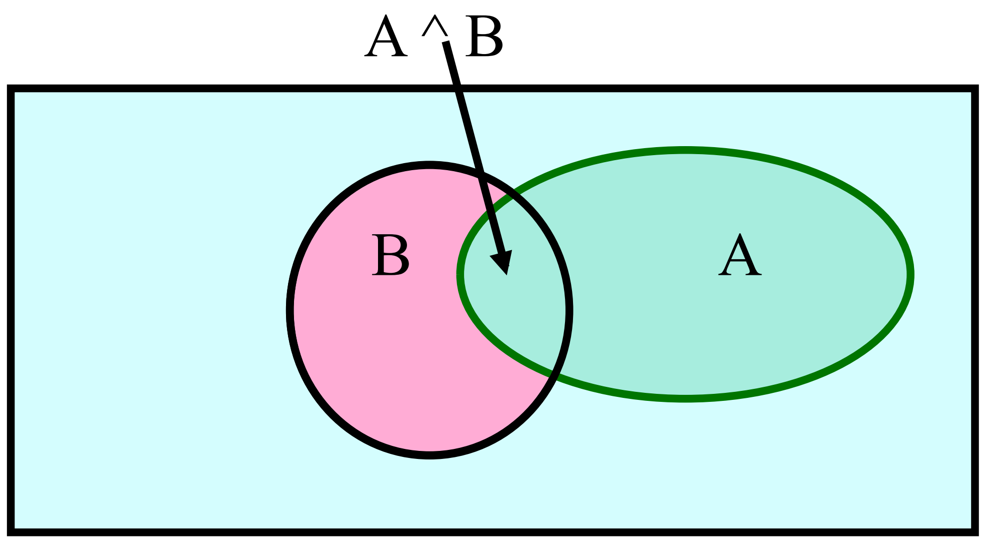
$$P(A|B) = \frac{P(A \wedge B)}{P(B)}$$

Corollary: The Chain Rule

$$P(A \wedge B) = P(A|B)\, P(B)$$

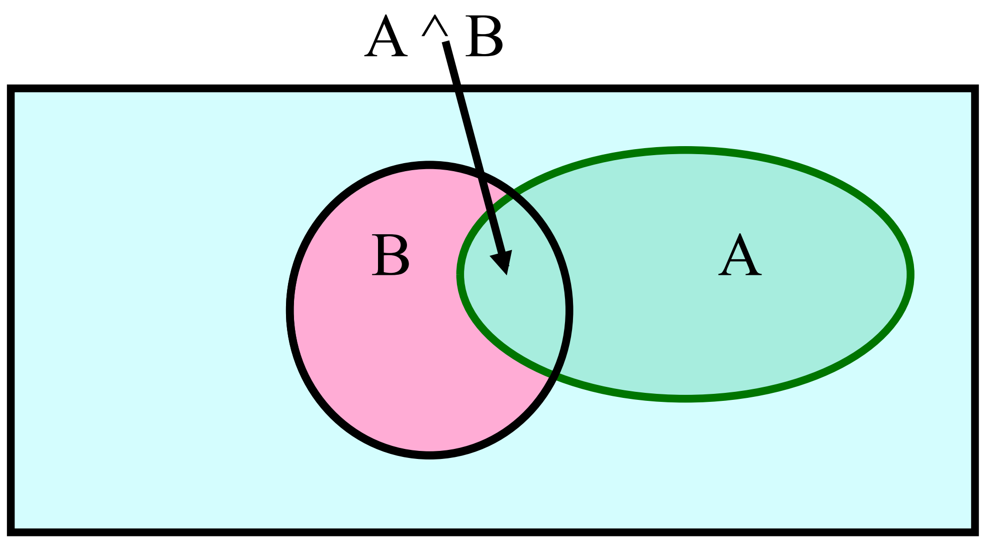
Bayes Rule

- let's write 2 expressions for $P(A \wedge B)$



Bayes Rule

- let's write 2 expressions for P(A ^ B)

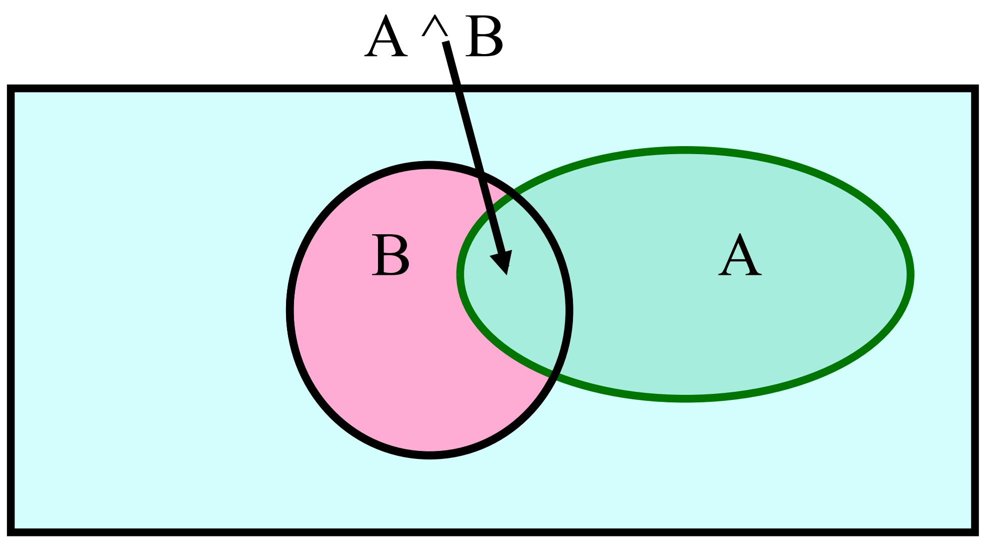


$$P(A|B) = \frac{P(A \wedge B)}{P(B)}$$

$$P(A \wedge B) = P(A|B)\,P(B)$$

Bayes Rule

- let's write 2 expressions for $P(A \wedge B)$



$$P(A|B) = \frac{P(A \wedge B)}{P(B)}$$

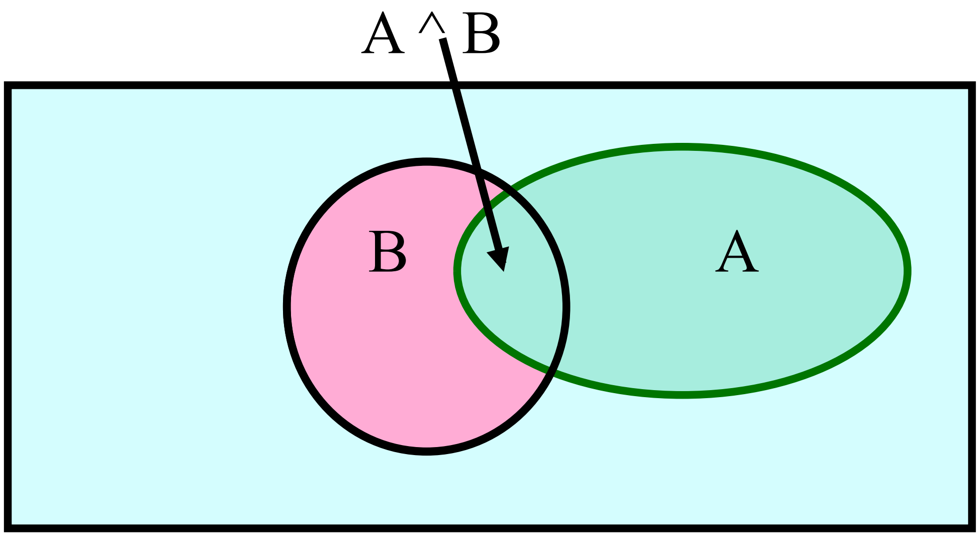
$$P(A \wedge B) = P(A|B)\,P(B)$$

$$P(B|A) = \frac{P(B \wedge A)}{P(B)}$$

$$P(B \wedge A) = P(B|A)\,P(A)$$

Bayes Rule

- let's write 2 expressions for P(A ^ B)



$$P(A|B) = \frac{P(A \wedge B)}{P(B)}$$

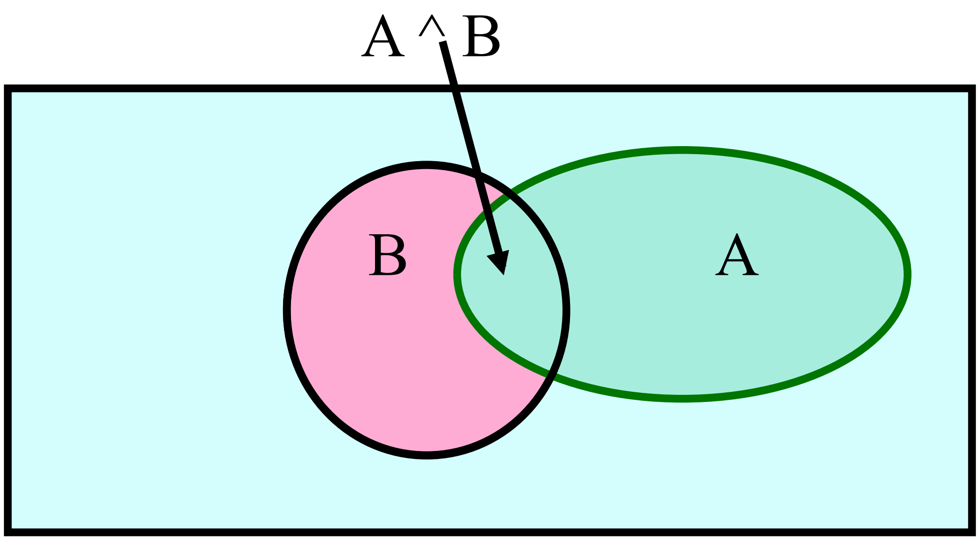
$$P(A \wedge B) = P(A|B)\, P(B)$$

$$P(B|A) = \frac{P(B \wedge A)}{P(B)}$$

$$P(B \wedge A) = P(B|A)\, P(A)$$

Bayes Rule

- let's write 2 expressions for P(A ^ B)



$$P(A \wedge B) = P(A|B)\,P(B) \quad \| \quad P(B \wedge A) = P(B|A)\,P(A)$$

$$P(B|A)\,P(A) = P(A|B)\,P(B)$$

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$

Bayes Rule

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$

we call P(A) the prior

and P(A|B) the posterior

Other Forms of Bayes Rule

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B|A)\,P(A) + P(B|\sim A)\,P(\sim A)}$$

$$P(A|B \wedge X) = \frac{P(B|A \wedge X)\,P(A \wedge X)}{P(B \wedge X)}$$

Applying Bayes Rule

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\sim A)P(\sim A)}$$

Find the probability of having a flu given that you have just coughed

A = you have the flu, B = you just coughed

Assume:
P(A) = 0.05
P(B|A) = 0.80
P(B| ~A) = 0.2

Applying Bayes Rule

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\sim A)P(\sim A)}$$

Find the probability of having a flu given that you have just coughed

A = you have the flu, B = you just coughed

Assume:
P(A) = 0.05
P(B|A) = 0.80
P(B| ~A) = 0.2

what is P(flu | cough) = P(A|B)?

Applying Bayes Rule

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\sim A)P(\sim A)}$$

Find the probability of having a flu given that you have just coughed

A = you have the flu, B = you just coughed

Assume:
P(A) = 0.05
P(B|A) = 0.80
P(B| ~A) = 0.2

P(~A) = 0.95
P (A |B) = 0.17

what is P(flu | cough) = P(A|B)?

The Joint Distribution

Example: Boolean variables A, B, C

Recipe for making a joint distribution of M variables:

The Joint Distribution

Example: Boolean variables A, B, C

Recipe for making a joint distribution of M variables:

1. Make a truth table listing all combinations of values of your variables (if there are M Boolean variables then the table will have 2^M rows).

A	**B**	**C**
0	0	0
0	0	1
0	1	0
0	1	1
1	0	0
1	0	1
1	1	0
1	1	1

The Joint Distribution

Example: Boolean variables A, B, C

Recipe for making a joint distribution of M variables:

1. Make a truth table listing all combinations of values of your variables (if there are M Boolean variables then the table will have 2^M rows).
2. For each combination of values, say how probable it is.

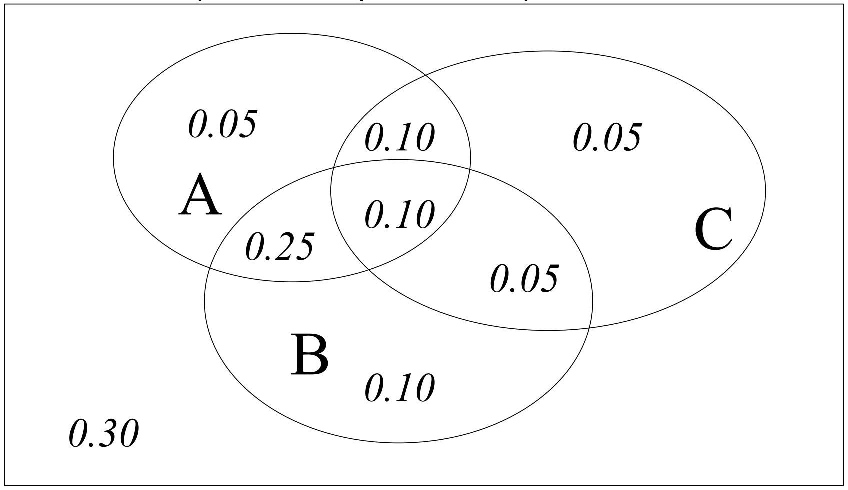
A	B	C	Prob
0	0	0	0.30
0	0	1	0.05
0	1	0	0.10
0	1	1	0.05
1	0	0	0.05
1	0	1	0.10
1	1	0	0.25
1	1	1	0.10

The Joint Distribution

Recipe for making a joint distribution of M variables:

1. Make a truth table listing all combinations of values of your variables (if there are M Boolean variables then the table will have 2^M rows).
2. For each combination of values, say how probable it is.
3. If you subscribe to the axioms of probability, those numbers must sum to 1.

A	B	C	Prob
0	0	0	0.30
0	0	1	0.05
0	1	0	0.10
0	1	1	0.05
1	0	0	0.05
1	0	1	0.10
1	1	0	0.25
1	1	1	0.10



Using the Joint Distribution

gender	hours_worked	wealth	
Female	v0:40.5-	poor	0.253122
		rich	0.0245895
	v1:40.5+	poor	0.0421768
		rich	0.0116293
Male	v0:40.5-	poor	0.331313
		rich	0.0971295
	v1:40.5+	poor	0.134106
		rich	0.105933

One you have the JD you can ask for the probability of **any** logical expression involving these variables

$$P(E) = \sum_{\text{rows matching } E} P(\text{row})$$

Using the Joint

gender	hours_worked	wealth	
Female	v0:40.5-	poor	0.253122
		rich	0.0245895
	v1:40.5+	poor	0.0421768
		rich	0.0116293
Male	v0:40.5-	poor	0.331313
		rich	0.0971295
	v1:40.5+	poor	0.134106
		rich	0.105933

P(Poor Male) = 0.4654

$$P(E) = \sum_{\text{rows matching } E} P(\text{ row })$$

Using the Joint

gender	hours_worked	wealth	
Female	v0:40.5-	poor	0.253122
		rich	0.0245895
	v1:40.5+	poor	0.0421768
		rich	0.0116293
Male	v0:40.5-	poor	0.331313
		rich	0.0971295
	v1:40.5+	poor	0.134106
		rich	0.105933

P(Poor) = 0.7604

$$P(E) = \sum_{\text{rows matching } E} P(\text{row})$$

Inference with the Joint

gender	hours_worked	wealth	
Female	v0:40.5-	poor	0.253122
		rich	0.0245895
	v1:40.5+	poor	0.0421768
		rich	0.0116293
Male	v0:40.5-	poor	0.331313
		rich	0.0971295
	v1:40.5+	poor	0.134106
		rich	0.105933

$$P(E_1|E_2) = \frac{P(E_1 \wedge E_2)}{P(E_2)} = \frac{\sum_{\text{rows matching } E_1 \text{ and } E_2} P(\text{ row })}{\sum_{\text{rows matching } E_2} P(\text{ row })}$$

$$P(\text{Male} \mid \text{Poor}) = 0.4654 / 0.7604 = 0.612$$

Learning and the Joint Distribution

gender	hours_worked	wealth	
Female	v0:40.5-	poor	0.253122
		rich	0.0245895
	v1:40.5+	poor	0.0421768
		rich	0.0116293
Male	v0:40.5-	poor	0.331313
		rich	0.0971295
	v1:40.5+	poor	0.134106
		rich	0.105933

Suppose we want to learn the function f: <G, H> ✎ W

Equivalently, P(W | G, H)

Solution: learn joint distribution from data, calculate P(W | G, H)

e.g., P(W=rich | G = female, H = 40.5-) =