Phil/LPS 31 Introduction to Inductive Logic
Lecture 12

David Mwakima
dmwakima@uci.edu
Department of Logic and Philosophy of Science
University of California, Irvine

May 15th 2023

Topics

- Joint Probability
- Marginal Probability
- Probabilistic Independence
- Conditional Probability
- Bayes' Theorem

Joint Probability and Marginal Probability

- In Homework 6 you will prove the General Disjunction Rule, which will bring the total number of rules to 8, so far.

Joint Probability and Marginal Probability

- In Homework 6 you will prove the General Disjunction Rule, which will bring the total number of rules to 8, so far.

Rule 8 $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Joint Probability and Marginal Probability

- In Homework 6 you will prove the General Disjunction Rule, which will bring the total number of rules to 8, so far.

Rule 8 $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

- In Rule 8 we have $P(A \cap B)$ for two sets A and B, which are not necessarily disjoint. But how do you calculate $P(A \cap B)$ if all we know is $P(A)$ and $P(B)$? Do we have a rule for that?!

Joint Probability and Marginal Probability

- In Homework 6 you will prove the General Disjunction Rule, which will bring the total number of rules to 8, so far.

Rule 8 $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

- In Rule 8 we have $P(A \cap B)$ for two sets A and B, which are not necessarily disjoint. But how do you calculate $P(A \cap B)$ if all we know is $P(A)$ and $P(B)$? Do we have a rule for that?!
- We call $P(A \cap B)$, the joint probability of A and B.

Joint Probability and Marginal Probability

- In Homework 6 you will prove the General Disjunction Rule, which will bring the total number of rules to 8, so far.

Rule 8 $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

- In Rule 8 we have $P(A \cap B)$ for two sets A and B, which are not necessarily disjoint. But how do you calculate $P(A \cap B)$ if all we know is $P(A)$ and $P(B)$? Do we have a rule for that?!
- We call $P(A \cap B)$, the joint probability of A and B.
- In the context of joint probability, and a concept we shall introduce later of conditional probability, $P(A)$ and $P(B)$ are called the marginal probability of A and B, respectively. The reason for the adjective "marginal" will become clear in a minute.

Joint Probability and Marginal Probability: First Illustration

	Spades	Hearts	Diamonds	Clubs	
Ace	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{13}$
2	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{13}$
3	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{13}$
4	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{13}$
5	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{13}$
6	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{13}$
7	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{13}$
8	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{13}$
9	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{13}$
10	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{13}$
Jack	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{13}$
Queen	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{13}$
King	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{13}$
	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	1

Joint Probability and Marginal Probability: First Illustration

- What is $P(\text{King})$, $P(\text{Spades})$, $P(\text{Ace})$, $P(\text{Diamonds})$, $P(7)$, $P(\text{Hearts})$? All these are examples of marginal probabilities

- What is $P(\text{King of Spades})$, $P(\text{Ace of Diamonds})$, $P(\text{7 of Hearts})$? All these are are examples of joint probabilities

- What is $P(\text{King}) \times P(\text{Spades})$, $P(\text{Ace}) \times P(\text{Diamonds})$, $P(7) \times P(\text{Hearts})$? Are the joint probabilities equal to the product of the respective marginal probabilities?
- We say that the suit and rank of a card are probabilistically independent.

Joint Probability and Marginal Probability: Second Illustration

	Hair Color				
Eye Color	Black	Brunette	Red	Blond	
Brown	0.11	0.20	0.04	0.01	0.37
Blue	0.03	0.14	0.03	0.16	0.36
Hazel	0.03	0.09	0.02	0.02	0.16
Green	0.01	0.05	0.02	0.03	0.11
	0.18	0.48	0.12	0.21	1

Joint Probability and Marginal Probability: Second Illustration

- What is P(Brown eye color), P(Black Hair), P(Blue eye color), P(Red Hair color), P(Green eye color), P(Blond Hair color)? All these are examples of marginal probabilities.

- What is P(Brown eyes and Black Hair), P(Blue eyes and Red Hair), P(Green eyes and Blond Hair)? All these are are examples of joint probabilities.

- What is P(Brown Eye color) $\times$ P(Black Hair), P(Blue eye color) $\times$ P(Red Hair color), P(Green Eyes) $\times$ P(Blond Hair)? Are the joint probabilities equal to the product of the respective marginal probabilities?
- We say that eye-color and hair color are probabilistically dependent.

Probabilitistic Independence

- We say two events E and F are probabilistically independent, or simply just independent, if:

$$P(E \cap F) = P(E) \times P(F)$$

Probabilitistic Independence

- We say two events E and F are probabilistically independent, or simply just independent, if:
$$P(E \cap F) = P(E) \times P(F)$$
- From this we can get one way of computing $P(A \cap B)$ only in the special case where A and B are independent.

Probabilitistic Independence

- We say two events E and F are probabilistically independent, or simply just independent, if:
$$P(E \cap F) = P(E) \times P(F)$$
- From this we can get one way of computing $P(A \cap B)$ only in the special case where A and B are independent.

Rule 9 $P(E \cap F) = P(E) \times P(F)$ if E and F are (probabilistically) independent.

Probabilitistic Independence

- We say two events E and F are probabilistically independent, or simply just independent, if:
$$P(E \cap F) = P(E) \times P(F)$$
- From this we can get one way of computing $P(A \cap B)$ only in the special case where A and B are independent.

Rule 9 $P(E \cap F) = P(E) \times P(F)$ if E and F are (probabilistically) independent.

- How about when A and B are dependent, like in the eye color and hair color example?

Probabilitistic Independence

- We say two events E and F are probabilistically independent, or simply just independent, if:

$$P(E \cap F) = P(E) \times P(F)$$

- From this we can get one way of computing $P(A \cap B)$ only in the special case where A and B are independent.

Rule 9 $P(E \cap F) = P(E) \times P(F)$ if E and F are (probabilistically) independent.

- How about when A and B are dependent, like in the eye color and hair color example?
- For that we need the concept of conditional probability.

Probabilitistic Independence

- We say two events E and F are probabilistically independent, or simply just independent, if:
$$P(E \cap F) = P(E) \times P(F)$$
- From this we can get one way of computing $P(A \cap B)$ only in the special case where A and B are independent.

Rule 9 $P(E \cap F) = P(E) \times P(F)$ if E and F are (probabilistically) independent.

- How about when A and B are dependent, like in the eye color and hair color example?
- For that we need the concept of conditional probability.
- But first let us talk about the theorem of total probability.

Joint Probability and Marginal Probability: Some Facts

We begin with some facts. Consider the two illustrations once more.

	Spades	Hearts	Diamonds	Clubs	
Ace	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{13}$
2	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{13}$
3	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{13}$
4	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{13}$
5	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{13}$
6	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{13}$
7	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{13}$
8	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{13}$
9	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{13}$
10	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{13}$
Jack	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{13}$
Queen	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{13}$
King	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{52}$	$\frac{1}{13}$
	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	1

Joint Probability and Marginal Probability: Some Facts

	Hair Color				
Eye Color	Black	Brunette	Red	Blond	
Brown	0.11	0.20	0.04	0.01	0.37
Blue	0.03	0.14	0.03	0.16	0.36
Hazel	0.03	0.09	0.02	0.02	0.16
Green	0.01	0.05	0.02	0.03	0.11
	0.18	0.48	0.12	0.21	1

Joint Probability and Marginal Probability: Some Facts

- The columns and rows of the table of joint probabilities are two different partitions of the set we are considering in a given context. For example, in the first illustration, the deck of cards is partitioned by suit column-wise and by rank row-wise. The set of people included in the eye and hair color data is partitioned by hair color column-wise and by eye-color row-wise.

Joint Probability and Marginal Probability: Some Facts

- The columns and rows of the table of joint probabilities are two different partitions of the set we are considering in a given context. For example, in the first illustration, the deck of cards is partitioned by suit column-wise and by rank row-wise. The set of people included in the eye and hair color data is partitioned by hair color column-wise and by eye-color row-wise.
- The column totals are the marginal probabilities of each set in the partition by column (suit or hair-color). This is an example of the theorem of total probability.

Joint Probability and Marginal Probability: Some Facts

- The columns and rows of the table of joint probabilities are two different partitions of the set we are considering in a given context. For example, in the first illustration, the deck of cards is partitioned by suit column-wise and by rank row-wise. The set of people included in the eye and hair color data is partitioned by hair color column-wise and by eye-color row-wise.
- The column totals are the marginal probabilities of each set in the partition by column (suit or hair-color). This is an example of the theorem of total probability.
- The row totals are the marginal probabilities of each set in the partition by row (rank or eye-color). This is also an example of the theorem of total probability.

Joint Probability and Marginal Probability: Some Facts

- The columns and rows of the table of joint probabilities are two different partitions of the set we are considering in a given context. For example, in the first illustration, the deck of cards is partitioned by suit column-wise and by rank row-wise. The set of people included in the eye and hair color data is partitioned by hair color column-wise and by eye-color row-wise.
- The column totals are the marginal probabilities of each set in the partition by column (suit or hair-color). This is an example of the theorem of total probability.
- The row totals are the marginal probabilities of each set in the partition by row (rank or eye-color). This is also an example of the theorem of total probability.
- The grand total of the marginal probabilities (by row or by column) is 1 (as it should be since the union of the sets in a partition is the whole set).

Theorem of Total Probability

- The Theorem of Total Probability is a generalization of Rule 7, which we called the Partition of an Event Rule.

Theorem of Total Probability

- The Theorem of Total Probability is a generalization of Rule 7, which we called the Partition of an Event Rule.
- Recall Rule 7 says that $P(E) = P(E \cap F) + P(E \cap F^c)$. Here we have partitioned E into disjoint sets $(E \cap F)$ and $(E \cap F^c)$, whose union is E.

Theorem of Total Probability

- The Theorem of Total Probability is a generalization of Rule 7, which we called the Partition of an Event Rule.
- Recall Rule 7 says that $P(E) = P(E \cap F) + P(E \cap F^c)$. Here we have partitioned E into disjoint sets $(E \cap F)$ and $(E \cap F^c)$, whose union is E.
- The theorem of total probability provides an alternative way of computing $P(E)$ if we can somehow partition E.

Theorem of Total Probability

- The Theorem of Total Probability is a generalization of Rule 7, which we called the Partition of an Event Rule.
- Recall Rule 7 says that $P(E) = P(E \cap F) + P(E \cap F^c)$. Here we have partitioned E into disjoint sets $(E \cap F)$ and $(E \cap F^c)$, whose union is E.
- The theorem of total probability provides an alternative way of computing $P(E)$ if we can somehow partition E.
- Suppose $\{E_1, E_2, E_3, \ldots, E_n\}$ is partition of E. The theorem of total probability says that:

$$\begin{aligned} P(E) &= P(E \cap E_1) + P(E \cap E_2) + P(E \cap E_3) + \cdots + P(E \cap E_n) \\ &= \sum_{i}^{n} P(E \cap E_i) \end{aligned}$$

Conditional Probability

- With the theorem of total probability under our belt, we can now finally talk about conditional probability.

Conditional Probability

- With the theorem of total probability under our belt, we can now finally talk about conditional probability.
- The conditional probability of A given B is written as $P(A \mid B)$. Key words that indicate you should use conditional probability in a problem include words like “given”, “if”, “dependent”, “without replacement” etc. You will see lots of examples in your Homework 6.

Conditional Probability

- With the theorem of total probability under our belt, we can now finally talk about conditional probability.
- The conditional probability of A given B is written as $P(A \mid B)$. Key words that indicate you should use conditional probability in a problem include words like "given", "if", "dependent", "without replacement" etc. You will see lots of examples in your Homework 6.
- Suppose you have a bag containing 13 yellow and 17 white marbles. You decide to sample one marble at a time, with replacement. What is $P(\text{Yellow marble})$, $P(\text{White marble} \mid \text{Yellow marble})$, $P(\text{Yellow marble} \mid \text{White marble})$?

Conditional Probability

- Now suppose you decide to sample one marble at a time, without replacement? This kind of sampling introduces dependency between events. We need to condition on the new information we are given (or that we obtain) once we see the color of the marble we sampled earlier in the sequence. What is $P(\text{White marble} \mid \text{Yellow marble})$, $P(\text{Yellow marble} \mid \text{White marble})$?

Conditional Probability

- Now suppose you decide to sample one marble at a time, without replacement? This kind of sampling introduces dependency between events. We need to condition on the new information we are given (or that we obtain) once we see the color of the marble we sampled earlier in the sequence. What is $P(\text{White marble} \mid \text{Yellow marble})$, $P(\text{Yellow marble} \mid \text{White marble})$?
- Notice that in general $P(E \mid F) \neq P(F \mid E)$. Recall that you showed in Quiz 1 that $\rightarrow$ is not symmetric. Be careful!

Conditional Probability

- Now suppose you decide to sample one marble at a time, without replacement? This kind of sampling introduces dependency between events. We need to condition on the new information we are given (or that we obtain) once we see the color of the marble we sampled earlier in the sequence. What is $P(\text{White marble} \mid \text{Yellow marble})$, $P(\text{Yellow marble} \mid \text{White marble})$?
- Notice that in general $P(E \mid F) \neq P(F \mid E)$. Recall that you showed in Quiz 1 that $\rightarrow$ is not symmetric. Be careful!
- Exercise. Consider the deck of cards illustration from earlier in the slides. Suppose you're drawing cards from a standard deck of 52 cards with replacement. What is:
  $P(\text{Ace of Hearts} \mid \text{King of Spades})$,
  $P(\text{Jack of Diamonds} \mid \text{Queen of Hearts})$,
  $P(\text{A red card} \mid \text{A black card})$?

Joint Probability from Conditional Probability

- Exercise. Consider the same deck of cards illustration from earlier in the slides. Suppose you're drawing cards from a standard deck of 52 cards without replacement. What is:
  $P(\text{Ace of Hearts} \mid \text{King of Spades})$,
  $P(\text{Jack of Diamonds} \mid \text{Queen of Hearts})$,
  $P(\text{A red card} \mid \text{A black card})$?

Joint Probability from Conditional Probability

- Exercise. Consider the same deck of cards illustration from earlier in the slides. Suppose you're drawing cards from a standard deck of 52 cards without replacement. What is: $P(\text{Ace of Hearts} \mid \text{King of Spades})$, $P(\text{Jack of Diamonds} \mid \text{Queen of Hearts})$, $P(\text{A red card} \mid \text{A black card})$?
- See Homework 6 for more exercises.

Joint Probability from Conditional Probability

- Exercise. Consider the same deck of cards illustration from earlier in the slides. Suppose you're drawing cards from a standard deck of 52 cards without replacement. What is:
  $P(\text{Ace of Hearts} \mid \text{King of Spades})$,
  $P(\text{Jack of Diamonds} \mid \text{Queen of Hearts})$,
  $P(\text{A red card} \mid \text{A black card})$?
- See Homework 6 for more exercises.
- These exercises illustrate a general principle concerning independent events:

  Event E is independent of F if and only $P(E) = P(E \mid F)$.

Joint Probability from Conditional Probability

- Exercise. Consider the same deck of cards illustration from earlier in the slides. Suppose you're drawing cards from a standard deck of 52 cards without replacement. What is:
  $P(\text{Ace of Hearts} \mid \text{King of Spades})$,
  $P(\text{Jack of Diamonds} \mid \text{Queen of Hearts})$,
  $P(\text{A red card} \mid \text{A black card})$?

- See Homework 6 for more exercises.

- These exercises illustrate a general principle concerning independent events:

  Event E is independent of F if and only $P(E) = P(E \mid F)$.

- So the independence between two events captures, in a sense, what we mean when we say that the occurrence of one event E doesn't give any information we can conditionalize upon in order to learn about the chances of another event F occurring.

Joint Probability from Conditional Probability

- Suppose we sample twice with replacement from the bag with 13 yellow marbles and 17 white marbles. What is:
  P(2 white marbles in a row),
  P(A white marble followed by a yellow marble),
  P(A yellow marble followed by a white marble)?

Joint Probability from Conditional Probability

- Suppose we sample twice with replacement from the bag with 13 yellow marbles and 17 white marbles. What is:
  P(2 white marbles in a row),
  P(A white marble followed by a yellow marble),
  P(A yellow marble followed by a white marble)?
- Sampling with replacement means that the sequence of samples are independent, i.e., the contents of the bag don't change, so we don't learn anything new that we can conditionalize upon.

Joint Probability from Conditional Probability

- Suppose we sample twice with replacement from the bag with 13 yellow marbles and 17 white marbles. What is:
  P(2 white marbles in a row),
  P(A white marble followed by a yellow marble),
  P(A yellow marble followed by a white marble)?
- Sampling with replacement means that the sequence of samples are independent, i.e., the contents of the bag don't change, so we don't learn anything new that we can conditionalize upon.
- Now suppose that we sample twice without replacement from the bag with 13 yellow marbles and 17 white marbles. What is:
  P(2 white marbles in a row),
  P(A white marble followed by a yellow marble),
  P(A yellow marble followed by a white marble)?

Joint Probability from Conditional Probability

- Sampling without replacement means that the sequence of samples are dependent, i.e., the contents of the bag change. So after each sample we learn something new that we can conditionalize upon.

Joint Probability from Conditional Probability

- Sampling without replacement means that the sequence of samples are dependent, i.e., the contents of the bag change. So after each sample we learn something new that we can conditionalize upon.

- What these examples, have illustrated is that when two events (E and F) are dependent, their joint probability is not the product of the marginal probabilities. In fact, you can use this fact as a litmus test for dependence.

Joint Probability from Conditional Probability

- Sampling without replacement means that the sequence of samples are dependent, i.e., the contents of the bag change. So after each sample we learn something new that we can conditionalize upon.
- What these examples, have illustrated is that when two events (E and F) are dependent, their joint probability is not the product of the marginal probabilities. In fact, you can use this fact as a litmus test for dependence.
- In general, when E and F are not independent, $P(E \cap F) = P(F) \times P(E|F)$. So we can add this as our final rule.

Joint Probability from Conditional Probability

- Sampling without replacement means that the sequence of samples are dependent, i.e., the contents of the bag change. So after each sample we learn something new that we can conditionalize upon.

- What these examples, have illustrated is that when two events (E and F) are dependent, their joint probability is not the product of the marginal probabilities. In fact, you can use this fact as a litmus test for dependence.

- In general, when E and F are not independent, $P(E \cap F) = P(F) \times P(E|F)$. So we can add this as our final rule.

Rule 10 $P(E \cap F) = P(F) \times P(E|F)$

Joint Probability from Conditional Probability

- Sampling without replacement means that the sequence of samples are dependent, i.e., the contents of the bag change. So after each sample we learn something new that we can conditionalize upon.
- What these examples, have illustrated is that when two events (E and F) are dependent, their joint probability is not the product of the marginal probabilities. In fact, you can use this fact as a litmus test for dependence.
- In general, when E and F are not independent, $P(E \cap F) = P(F) \times P(E|F)$. So we can add this as our final rule.

Rule 10 $P(E \cap F) = P(F) \times P(E|F)$

- From Rule 10 we define the conditional probability of E given F as:

$$\boxed{P(E \mid F) = \frac{P(E \cap F)}{P(F)}}$$

Bayes' Theorem

- A very powerful (immediate) consequence of the definition of conditional probability is Bayes' Theorem, named after Rev. Thomas Bayes. In Homework 6 you will derive Bayes' Theorem. Here I just state this remarkable theorem, which we shall get a lot of mileage out of.

$$\boxed{P(E \mid F) = \frac{P(F|E)P(E)}{P(F|E)P(E) + P(F|E^c)P(E^c)}}$$

Bayes' Theorem

- A very powerful (immediate) consequence of the definition of conditional probability is Bayes' Theorem, named after Rev. Thomas Bayes. In Homework 6 you will derive Bayes' Theorem. Here I just state this remarkable theorem, which we shall get a lot of mileage out of.

$$\boxed{P(E \mid F) = \frac{P(F|E)P(E)}{P(F|E)P(E) + P(F|E^c)P(E^c)}}$$

- A slightly more general form of Bayes Theorem says that if $\{E_1, E_2, E_3, E_4, \ldots, E_n\}$ is a partition of E, then:

$$\boxed{P(E \mid F) = \frac{P(F|E)P(E)}{\sum_i^n P(F|E_i)P(E_i)}}$$

Bayes' Theorem

- A very powerful (immediate) consequence of the definition of conditional probability is Bayes' Theorem, named after Rev. Thomas Bayes. In Homework 6 you will derive Bayes' Theorem. Here I just state this remarkable theorem, which we shall get a lot of mileage out of.

$$\boxed{P(E \mid F) = \frac{P(F|E)P(E)}{P(F|E)P(E) + P(F|E^c)P(E^c)}}$$

- A slightly more general form of Bayes Theorem says that if $\{E_1, E_2, E_3, E_4, \ldots, E_n\}$ is a partition of E, then:

$$\boxed{P(E \mid F) = \frac{P(F|E)P(E)}{\sum_i^n P(F|E_i)P(E_i)}}$$

- Memorize this!