Lecture 3

July 29, 2010

Probability

Reminder: start the recording

University of Washington
Linguistics 473: Computational Linguistics Fundamentals
Lecture 3: Probability

Project 1

- Due at 11:45 p.m. on Tuesday
- Unix or system issues?
- Questions?

University of Washington
Linguistics 473: Computational Linguistics Fundamentals
Lecture 3: Probability

Assignment 1

1. Thank you for your essays. Full credit
2. *I saw that gas can explode.*
   - I realized that gases (in general) are able to explode.
   - I saw (with my eyes) that (particular) (metal) gas container explode
   - I realized (that) that (particular) gas is able to explode.

University of Washington
Linguistics 473: Computational Linguistics Fundamentals
Lecture 3: Probability

3. All possible 5-letter words (26^5)

subtract words with all consonants (20^5)

subtract words with all vowels (6^5)

$26^5 - 20^5 - 6^5 = 8{,}673{,}600$

University of Washington
Linguistics 473: Computational Linguistics Fundamentals
Lecture 3: Probability

4. Assuming that we consider identical characters to be indistinguishable in the output:

(萄 萄 萄 萄 橙 橙 苹 梨)

repeated groups: 4, 2, 1, 1

$$\frac{8!}{4! \times 2! \times 1! \times 1!} = 840$$

University of Washington
Linguistics 473: Computational Linguistics Fundamentals
Lecture 3: Probability

5. How many pairwise comparisons are possible between documents on the same topic?

$$\binom{8}{2} + \binom{6}{2} + \binom{3}{2} = 46$$

How many pairwise comparisons are possible between documents on different topics?

$$8 \times 6 + 8 \times 3 + 6 \times 3 = 90$$

6. Extra Credit

Write an expression that gives the number of unordered sets of *k* items that can be formed from a set of *n* distinct items while allowing repetition in the output set.

example: { a, b, c, d } choose 3 (unordered), but allowing repetition in the output:

{ a a a }, { a a b }, { a a c }, { a a d },
{ a b b }, { a b c }, { a b d }, { a c c },
{ a c d }, { a d d }, { b b b }, { b b c },
{ b b d }, { b c c }, { b c d }, { b d d },
{ c c c }, { c c d }, { c d d }, { d d d }

We ensure that order doesn't matter (i.e. we're not including duplicate tuples) by enumerating them using the convention of displaying the letters in sorted order. So we know the answer should be 20.

Divide into groups and count them:

a-group { a a a } { a a b } { a a c } { a a d } { a b b }
{ a b c } { a b d } { a c c } { a c d } { a d d }

$$\binom{4}{2} + 4 = 10$$

b-group { b b b } { b b c }{ b b d } { b c c } { b c d } { b d d }

$$\binom{3}{2} + 3 = 6$$

c-group { c c c } { c c d }{ c d d }

$$\binom{2}{2} + 2 = 3$$

d-group { d d d }

$$\binom{1}{2} + 1 = 1$$

$$\sum_{i}^{n}\left(\binom{i}{k-1} + i\right) = \boxed{\binom{n+k-1}{k}} = \left(\!\binom{n}{k}\!\right)$$

This is called the multiset coefficient

see http://en.wikipedia.org/wiki/Multiset
and http://en.wikipedia.org/wiki/Stars_and_bars_(probability)

Probability

- So far, we have considered counting and arranging sets of items (entities, elements)
- If we consider the event of selecting an element from a set, we enter the world of probability
- This event could also be called an observation or a trial
- The set becomes known as the sample space and—by definition—it is assigned a probability mass of 1.0

Sample Spaces

- Ω (omega) is often used to represent the sample space
- Sample spaces can be discrete or continuous

  discrete:

  Ω = (apple, banana, banana, orange)

  continuous:

  Ω = { *the mass of an orange* }

- P(Ω) = 1 (all possible events are accounted for)

most applications in computational linguistics involve discrete probabilities

Outcomes

- Often, events are often notated with a italic capital letter corresponding to a single outcome (a lower-case letter)

  Ω = (a a b c)

  A is the event of designating 'a' from Ω

  B is the event of designating 'b' from Ω

  C is the event of designating 'c' from Ω
- A^C denotes the complement of event *A*
  - An event *A* partitions the sample space into *A* and A^C
- An event can also be any subset of Ω
  - All individual outcomes are events, but events can also be combinations of individual outcomes

Rolling 2 dice

- The single occurrence of rolling a red die and a black die must have the one following outcomes (red, black)

Ω = { (1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6),
(2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6),
(3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6),
(4, 1), (4, 2), (4, 3), (4, 4), (4, 5), (4, 6),
(5, 1), (5, 2), (5, 3), (5, 4), (5, 5), (5, 6),
(6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6) }

- There are 6 × 6 = 36 outcomes; they are mutually exclusive and collectively exhaustive
- But there are many other events that we can talk about…

Some 2-dice Events

- A particular outcome
  $A = \{ (3, 6) \}$
- Both dice are the same
  $E = \{ (1, 1), (2, 2), (3, 3), (4, 4), (5, 5) \}$
- The total is 5
  $F = \{ (1, 4), (2, 3), (3, 2), (4, 1) \}$
- The total is prime
  $G = \{ (1, 1), (1, 2), (1, 4), (1, 6), (2, 1), (2, 3), (2, 5),(3, 2), (3, 4), (4, 1), (4, 3), (5, 2), (5, 6), (6,1), (6, 5) \}$

Definition of Probability

- Let *P* be a function that satisfies the following:
  $P(\Omega) = 1$
  all possible outcomes are accounted for
  $\forall A \subseteq \Omega : P(A) \geq 0$
  probabilities are non-negative real numbers
  $\forall \{ A, B \} \subseteq \Omega, A \cap B = \emptyset : P(A \cup B) = P(A) + P(B)$
  for any pair of events that are mutually exclusive, the union of their probabilities is the sum of their probabilities

$\emptyset$ denotes the empty set, { }

It follows that:

- If *P* is a valid probability function for sample space Ω
  $\forall A \subseteq \Omega : P(A) \leq 1$
  Probabilities are real numbers in the range [0, 1]
- This must be true because probabilities cannot be negative (by definition), and they must sum to 1
- For every trial, an event either occurs, or does not occur
  $\forall A \subseteq \Omega : P(A^C) = 1 - P(A)$

In this way, each event $A \subseteq \Omega$ can be thought of as defining its own proper probability space containing two outcomes

Mutually Exclusive Outcomes

- It is impossible for two mutually exclusive events to co-occur on the same trial
- For the 2 dice example, each of the 36 basic outcomes are mutually exclusive with each other, and the entire set is collectively exhaustive
- Therefore, one way of defining *P* is to assume that these outcomes are all equally likely:

$$E = \{ (1, 6) \}$$
$$P(E) = \frac{1}{|\Omega|} = \frac{1}{36} = .0278$$

This is the objective when manufacturing a fair gaming die

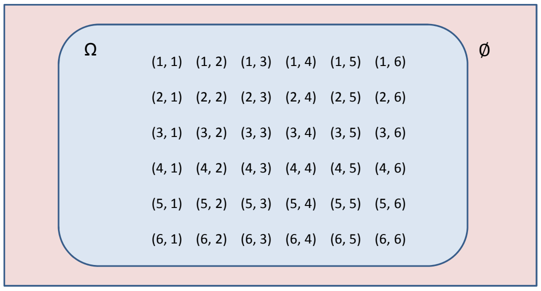
Compositional Events

- The other events we mentioned, they are composed of these basic mutually exclusive outcomes
- If the function *P* describes a valid probability space, then the definition of well-formed *P* allows us to calculate *P* for mutually exclusive compositional events
  $\forall \{ A, B \} \subseteq \Omega, A \cap B = \emptyset : P(A \cup B) = P(A) + P(B)$
- Compositional events are handy for grouping together certain types of events that we might be interested in
- Every trial has an outcome, which may satisfy multiple events; this can be illustrated with Venn Diagrams

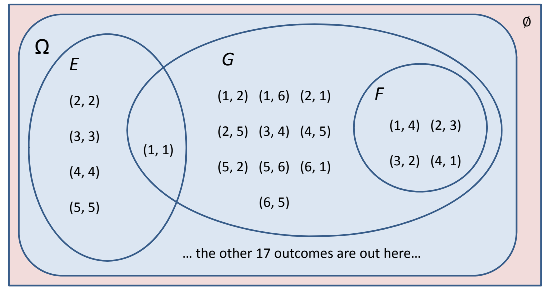
2 Dice Events

- Both dice are the same
  $E = \{ (1, 1), (2, 2), (3, 3), (4, 4), (5, 5) \}$
  $P(E) = .0278 \times 5 = .1389$
- The total is 5
  $F = \{ (1, 4), (2, 3), (3, 2), (4, 1) \}$
  $P(F) = .0278 \times 4 = .1111$
- The total is prime
  $G = \{ (1, 1), (1, 2), (1, 4), (1, 6), (2, 1), (2, 3), (2, 5), (3, 2), (3, 4), (4, 1), (4, 3), (5, 2), (5, 6), (6,1), (6, 5) \}$
  $P(G) = .0278 \times 15 = .4167$

Outcomes in Probability Space



Event Composition



Intersecting Events

- The previous slide shows that compositional events can be mutually exclusive

  E and *F* are mutually exclusive

  $E \cap F = \emptyset$

  E and *G* are not mutually exclusive

  $E \cap G = \{ (1, 1) \}$

  F and *G* are not mutually exclusive

  $F \cap G = \{ (1, 4), (2, 3), (3, 2), (4, 1) \} = F$

More on Adding Probabilities

- We have seen how to calculate probability of P(*A* or *B*) when *A* and *B* are mutually exclusive

  $P(A \cup B) = P(A) + P(B),\ A \cap B = \emptyset$

- If they are not, we can subtract the probability of the intersecting area

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

  Probability of both dice being the same, *or* their total being prime in a single trial

  $P(E \cup G) = P(E) + P(G) - P(E \cap G)$

  $= .1389 + .4167 - .0278$

  $= .5278$

We got this value for $P(E \cap G)$ by $\frac{|E\ \cap\ G|}{|\Omega|}$ but more discussion follows

Joint Probability

- On the previous slide we knew that $P(E \cap G) = .0278$ by noting that only 1 of the 36 mutually exclusive, collectively exhaustive outcomes is in the set intersection $E \cap G$
- More generally though, how can we compute $P(E \cap G)$ from $P(E)$ and $P(G)$?
- $P(E \cap G)$, or $P(E$ and $G)$, or $P(EG)$ is the probability that two events both occur in the same trial
- This is called the joint probability
- For mutually exclusive events, the joint probability is obviously zero:

  $\forall \{ A, B \} \subseteq \Omega, A \cap B = \emptyset : P(A \cap B) = 0$

Joint Probability

Recall our example

$E = \{ (1, 1), (2, 2), (3, 3), (4, 4), (5, 5) \}$

"both dice are the same"

$F = \{ (1, 4), (2, 3), (3, 2), (4, 1) \}$

"the total is 5"

$\forall \{ A, B \} \subseteq \Omega, A \cap B = \emptyset : P(A \cap B) = 0$

$E \cap F = \emptyset,\ \therefore P(E \cap F) = 0$

The probability is zero, meaning it is not possible for both dice to be the same and for the total to be 5 on the same trial.

Joint Probability

- Perhaps it is the case that

  $P(E \cap G) = P(E)\ P(G)$

  Let's try it

  $.0278 \stackrel{?}{=} .1389 \times .4167$

  $.0278 \stackrel{?}{=} .0578$

  No. This means that events *E* and *G* are not independent

Independent Events

- Independence is not the same as mutual exclusivity
  - 2 events are mutually exclusive if they cannot both occur as the outcome of a single trial
  - 2 events are independent if the occurrence of one does not affect the probability of the other occurring in the trial
- Does event A provide any information that would bias the outcome of event B?
- Events *E*, *F*, and *G* in the 2-dice example are not independent of each other

Adding an independent event to our example

Let's start with event *F* and try to think of an event that would be independent of *F*

$F = \{ (1, 4), (2, 3), (3, 2), (4, 1) \}$

"the total is 5"

$P(F) = .1111$

It's not so easy to come up with an event that, in the same trial, will give us no information about *F*. Such an event must meet the following criteria:

•Since F does not partition Ω equally, a event that is independent of F must partition Ω equally, so as not to bias for or against F.

•For the same reason, the event must also partition F equally.

Any ideas?

An event that is independent of *F*

"the red die shows an odd number"

$H = \{ (1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6),$
$(3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6),$
$(5, 1), (5, 2), (5, 3), (5, 4), (5, 5), (5, 6) \}$

$P(H) = .5$

$F = \{ (1, 4), (2, 3), (3, 2), (4, 1) \}$

"the total is 5"

$P(F) = .1111$

$H \cap F = \{ (1, 4), (3, 2) \}$

$P(H \cap F) = .0555 \stackrel{?}{=} P(H)\ P(F) \stackrel{?}{=} .5 \times .111$ ✓

Independent Events

When two events are independent, the probability of both occurring in the same trial is

$P(A \cap B) = P(A)\ P(B)$

- Actually, the reverse of this is the *definition* of independence
- This is how we can test for independence of events
  - we can compare the probability $P(A \cap B)$—obtained from counting—to the product of $P(A)$ and $P(B)$. If they are equal, the events are independent

Conditional Probability

- But what if two events are not independent? How do we compute $P(A \cap B)$ from $P(A)$ and $P(B)$?
- We must know how the events are related
- $P(A|B)$ is notation for the probability of event *A* when assuming that event *B* has co-occurred in the same trial
- This is called conditional probability
- "the probability of A, given B"
- Defines a constrained probability space which contains only those outcomes which satisfy event *B*

Conditional Probability

- Because the reduced sample space is limited to events which satisfy *B*, we exclude from *A* any outcomes that do not satisfy *B*: $P(A \cap B)$
- This lets us express the conditional probability in terms of the reduced sample space

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

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

- If $P(B)$ is 0, then $P(A|B)$ is undefined

Marginal Probability

- Conditional probability introduces the idea that you might have information about *part* of a trial
- In the following equation, we are assuming that we can estimate or provide $P(B)$ for an incomplete trial

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

- P(B) here is called the marginal probability

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

joint probability = conditional probability × marginal probability

Conditional probability and independence

- Note that conditional probability degrades gracefully in the case of independent events
- Assuming *A* and *B* are independent events:

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

If *A* and *B* are independent, then what you may know about one doesn't affect the probability of the other

Summary of Event Probability

- $P(A^C) = 1 - P(A)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- If $P(A \cap B) = P(A)\,P(B)$, then *A* and *B* are called independent events
- Otherwise

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

- Conditional probability

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

Assignment 2: Probability

- Due August 5, 4:30 p.m.
- 2 problems based on this lecture: Event probabilities, independence, mutual exclusivity

Next Tuesday

- Project 1 due at 11:45 p.m.
- Random Variables, the Chain Rule, and Probability Distributions