CS 362: Intelligent Systems

Unit 5

STOCHASTIC METHODS

Outline

What is Uncertainty?

- **Uncertainty is essentially lack of information to formulate a decision.**
- **Uncertainty may result in making poor or bad decisions.**
- **As living creatures, we are accustomed to dealing with uncertainty – that's how we survive.**
- **Dealing with uncertainty requires reasoning under uncertainty along with possessing a lot of common sense.**

Theories to Deal with Uncertainty

- Bayesian Probability
- Hartley Theory (1928)
- Shannon Theory (1948)
- Dempster-Shafer Theory (1976)
- Markov Models
- Zadeh's Fuzzy Theory(1965)

Dealing with Uncertainty

- **Deductive reasoning – deals with exact facts and exact conclusions**
- **Inductive reasoning – not as strong as deductive – premises support the conclusion but do not guarantee it.**
- **There are a number of methods to pick the best solution in light of uncertainty.**
- **When dealing with uncertainty, we may have to settle for just a good solution.**

Introduction

- **One important application domain for the use of the stochastic methodology is diagnostic reasoning where cause/effect relationships are not always captured in a purely deterministic fashion, as is often possible in the knowledge-based approaches to problem solving that we saw in Chapters 2, 3, 4, and will see again in Chapter 8.**

- **A diagnostic situation usually presents evidence, such as fever or headache, without further causative justification. In fact, the evidence can often be indicative of several different causes, e.g., fever can be caused by either flu or an infection. In these situations probabilistic information can often indicate and prioritize possible explanations for the evidence.**

- **Another interesting application for the stochastic methodology is gambling, where supposedly random events such as the roll of dice, the dealing of shuffled cards, or the spin of a roulette wheel produce possible player payoff.**

Introduction

In fact, in the 18th century, the attempt to provide a mathematical foundation for gambling was an important motivation for Pascal (and later Laplace) to develop a probabilistic calculus.

- **We next describe several problem areas, among many, where the stochastic methodology is often used in the computational implementation of intelligence; these areas will be major topics in later chapters.**

Some Applications for Stochastic Methods

We describe several problem areas, where the stochastic methodology is often used in the computational implementation of intelligence.

- **Diagnostic Reasoning.**

  In medical diagnosis, for example, there is not always an obvious cause/effect relationship between the set of symptoms presented by the patient and the causes of these symptoms. In fact, the same sets of symptoms often suggest multiple possible causes.

- **Natural language understanding.**

  If a computer is to understand and use a human language, that computer must be able to characterize how humans themselves use that language. Words, expressions, and metaphors are learned, but also change and evolve as they are used over time.

Some Applications for Stochastic Methods

- **Planning and scheduling.**

  When an agent forms a plan, for example, a vacation trip by automobile, it is often the case that no deterministic sequence of operations is guaranteed to succeed. What happens if the car breaks down, if the car ferry is cancelled on a specific day, if a hotel is fully booked, even though a reservation was made?

- **Learning.**

  The three previous areas mentioned for stochastic technology can also be seen as domains for automated learning. An important component of many stochastic systems is that they have the ability to sample situations and learn over time.

The stochastic methodology has its foundation in the properties of counting.

The Elements of Counting

- **The foundation for the stochastic methodology is the ability to count the elements of an application domain.**
- **The basis for collecting and counting elements is, set theory, in which we must be able to determine whether an element is or is not a member of a set of elements.**
- **Once this is determined, there are methodologies for counting elements of sets, of the *complement* of a set, and the *union* and *intersection* of multiple sets.**

The Elements of Counting

- If we have a set A, the number of the elements in set A is denoted by |A|, called the *cardinality* of A.
- Of course, A may be
  - Empty (the number of elements is zero),
  - Finite,
  - Countably infinite,
  - or uncountably infinite.
- Each set is defined in terms of a domain of interest or *universe* , U, of elements that might be in that set

The Elements of Counting Example

Suppose the universe, U, is the set $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$

Let A be the set $\{1, 3, 5, 7, 9\}$

Let B be the set $\{0, 2, 4, 6, 8\}$

Let C be the set $\{4, 5, 6\}$

Then $|A|$ is 5, $|B|$ is 5, $|C|$ is 3, and $|U|$ is 10.

Also, $A \subseteq U$, $B \subseteq U$ and $A \cup B = U$, $|B| = |A|$, $A = \overline{B}$

Further, $|A \cup B| = |A| + |B| = 10$, since $A \cap B = \{\ \}$, but

$|A \cup C| = |A| + |C| - |A \cap C| = 7$, since $A \cap C = \{5\}$

The Elements of Counting
The Addition and Multiplication Rules

The *addition rule* for combining two sets.

- **For any two sets A and C, the number of elements in the union of these sets is:**

$$|A \cup C| = |A| + |C| - |A \cap C|$$

- **A similar addition rule holds for three sets A,B, and C,**

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|$$

This Addition rule may be generalized to any *finite* number of sets

The Elements of Counting
The Addition and Multiplication Rules

- The *Cartesian Product* of two sets A and B, is the set of all ordered pairs (a, b) where a is an element of set A and b is an element of set B; or more formally:

$$A \times B = \{(a, b) \mid (a \in A) \wedge (b \in B)\}$$

- and by the multiplication principle of counting for two sets

$$|A \times B| = |A| \times |B|$$

The Cartesian product can, of course, be defined across any number of sets.

The Elements of Counting
Permutations and Combinations

- **A *permutation* of a set of elements is an arranged sequence of the elements of that set.**
- **The *permutations* of a set of n elements taken r at a time**

$$n \times (n - 1) \times (n - 2) \times (n - 3) \times \ldots \times (n - (r - 1))$$

- **The *combinations* of a set of n elements taken r at a time**

$$_{n}C_{r} = \frac{_{n}P_{r}}{r!} = \frac{n!}{(n-r)!\,r!} \qquad _{n}P_{r} = \frac{n!}{(n-r)!}$$

The Elements of Counting
Permutations and Combinations

- A *permutation* of a set of elements is an arranged sequence of the elements of that set.
- The *permutations* of a set of **n** elements taken **r** at a time ($0 \le r \le n$)

$$n \times (n - 1) \times (n - 2) \times (n - 3) \times \ldots \times (n - (r - 1))$$

we can r

$$\frac{n \times (n - 1) \times (n - 2) \times (n - 3) \times \ldots \times (n - (r - 1)) \times (n - r) \times (n - r - 1) \times \ldots \times 2 \times 1}{(n - r) \times (n - r - 1) \times \ldots \times 2 \times 1}$$

- e

time, which is symbolized as ${}_nP_r$, is:

$${}_nP_r = \frac{n!}{(n-r)!}$$

The Elements of Counting
Permutations and Combinations

- The combination of a set of n elements is any subset of these elements that can be formed.
- The *combinations* of a set of **n** elements taken **r** at a time ($0 \le r \le n$)

$$ {}_nC_r = \frac{{}_nP_r}{r!} = \frac{n!}{(n-r)!\, r!} $$

The Elements of Counting
Permutations and Combinations

Permutations: there are basically two types of permutation:

(remember the order does matter now):

- Repetition is Allowed: such as the lock example. It could be "333".
- No Repetition: for example the first three people in a running race. You can't be first *and* second.

Combinations: there are also two types of combinations (remember the order does not matter now):

- Repetition is Allowed: such as coins in your pocket (5,5,5,10,10)
- No Repetition: such as lottery numbers (2,14,15,27,30,33)

The Elements of Counting
Permutations with Repetition Example

Example: in the lock below, there are 10 numbers to choose from (0,1,...9) and we choose 3 of them:

$10 \times 10 \times \ldots$ (3 times) $= 10^3 = 1{,}000$ permutations

So, the formula is simply:

$$n^r$$

where n is the number of things to choose from, and we choose r of them (Repetition allowed, order matters)

The Elements of Counting
Permutations without Repetition Example

For example, what order could 16 pool balls be in?

After choosing, say, number "14" we can't choose it again.

So, our first choice would have 16 possibilities, and our next choice would then have 15 possibilities, then 14, 13, etc. And the total permutations would be:

$$16 \times 15 \times 14 \times 13 \times \ldots = 20{,}922{,}789{,}888{,}000$$

But maybe we don't want to choose them all, just 3 of them, so that would be only:

$$16 \times 15 \times 14 = 3{,}360$$

In other words, there are 3,360 different ways that 3 pool balls could be selected out of 16 balls.

The Elements of Counting
Permutations without Repetition Example

So, if we wanted to select all of the billiard balls the permutations would be:

$$16! = 20{,}922{,}789{,}888{,}000$$

But if we wanted to select just 3, then we have to stop the multiplying after 14. How do we do that? There is a neat trick ... we divide by 13! ...

$$\frac{16 \times 15 \times 14 \times 13 \times 12 \ldots}{13 \times 12 \ldots} = 16 \times 15 \times 14 = 3{,}360$$

Do you see? 16! / 13! = 16 × 15 × 14

- The formula is written:

$$\frac{n!}{(n-r)!}$$

where n is the number of things to choose from, and we choose r of them (No repetition, order matters)

The Elements of Counting
Permutations without Repetition Example

Examples:

Our "order of 3 out of 16 pool balls example" would be:

$$\frac{16!}{(16-3)!} = \frac{16!}{13!} = \frac{20{,}922{,}789{,}888{,}000}{6{,}227{,}020{,}800} = 3{,}360$$ **(which is just the same as: 16 × 15 × 14 = 3,360)**

- **How many ways can first and second place be awarded to 10 people?**

$$\frac{10!}{(10-2)!} = \frac{10!}{8!} = \frac{3{,}628{,}800}{40{,}320} = 90$$ **(which is just the same as: 10 × 9 = 90)**

- **Notation**
- **Instead of writing the whole formula, people use different notations such as these:**

$$P(n,r) = {}^{n}P_{r} = {}_{n}P_{r} = \frac{n!}{(n-r)!}$$

Example: *P*(10,2) = 90

The Elements of Counting
Combinations without Repetition Example

Going back to our pool ball example, let's say we just want to know which 3 pool balls were chosen, not the order.

We already know that 3 out of 16 gave us 3,360 permutations.

But many of those will be the same to us now, because we don't care what order!

- **For example, let us say balls 1, 2 and 3 were chosen. These are the possibilities:**

Order does matter	Order doesn't matter
1 2 3 1 3 2 2 1 3 2 3 1 3 1 2 3 2 1	1 2 3

So, the permutations will have 6 times as many possibilities.

- **In fact there is an easy way to work out how many ways "1 2 3" could be placed in order, and we have already talked about it. The answer is:**

$$3! = 3 \times 2 \times 1 = 6$$

The Elements of Counting
Combinations without Repetition Example

- **So we adjust our permutations formula to reduce it by how many ways the objects could be in order (because we aren't interested in their order any more):**

$$\frac{n!}{(n-r)!} \times \frac{1}{r!} = \frac{n!}{r!(n-r)!}$$

- **That formula is so important it is often just written in big parentheses like this:**

$$\frac{n!}{r!(n-r)!} = \binom{n}{r}$$

where n is the number of things to choose from, and we choose r of them (No repetition, order doesn't matter)

- **It is often called "n choose r" (such as "16 choose 3") and is also known as the "Binomial Coefficient“**

Notation

- **As well as th [illegible] le also use these notations:**

$$C(n,r) = {}^{n}C_{r} = {}_{n}C_{r} = \binom{n}{r} = \frac{n!}{r!(n-r)!}$$

The Elements of Counting
Combinations without Repetition Example

So, our pool ball example (now without order) is:

$$\frac{16!}{3!(16-3)!} = \frac{16!}{3!\times 13!} = \frac{20,922,789,888,000}{6\times 6,227,020,800} = 560$$

Or we could do it this way:

$$\frac{16\times 15\times 14}{3\times 2\times 1} = \frac{3360}{6} = 560$$

So remember, do the permutation, then reduce by a further "r!"

In other words choosing 3 balls out of 16, or choosing 13 balls out of 16 have the same number of combinations.

$$\frac{16!}{3!(16-3)!} = \frac{16!}{13!(16-13)!} = \frac{16!}{3!\times 13!} = 560$$

The Elements of Counting Combinations with Repetition Example

- **Actually, these are the hardest to explain.**

We would write it like this:

$$\binom{n+r-1}{r} = \frac{(n+r-1)!}{r!(n-1)!}$$

where n is the number of things to choose from, and we choose r of them (Repetition allowed, order doesn't matter)

Elements of Probability Theory

The following definitions, the foundation for a theory for probability, were first formalized by the French mathematician Laplace (1816) in the early nineteenth century.

D E F I N I T I O N S:

- **The Sample Space,**
- **Probabilities, and**
- **Independence**

Elements of Probability Theory
The Sample Space, Probabilities, and Independence

DEFINITION

- **ELEMENTARY EVENT**

  An *elementary* or *atomic event* is a happening or occurrence that cannot be made up of other events.

- **EVENT, E**

  An *event* is a set of elementary events.

- **SAMPLE SPACE, S**

  The set of all possible outcomes of an event E is the *sample space* S or universe for that event.

- **PROBABILITY, p**

  The *probability* of an event E in a sample space S is the ratio of the number of elements in E to the total number of possible outcomes of the sample space S of E. Thus,

  $p(E) = |E| / |S|$.

Example

For example, what is the probability that a 7 or an 11 are the result of the roll of two fair dice?

- **We first determine the sample space for this situation. Using the multiplication principle of counting, each die has 6 outcomes, so the total set of outcomes of the two dice is 36.**
- **The number of combinations of the two dice that can give a 7 is 1,6; 2,5; 3,4; 4,3; 5,2; and 6,1 - 6 altogether. The probability of rolling a 7 is thus 6/36 = 1/6.**
- **The number of combinations of the two dice that can give an 11 is 5,6; 6,5 - or 2, and the probability of rolling an 11 is 2/36 = 1/18.**

Using the additive property of distinct outcomes, there is 1/6 + 1/18 or 2/9 probability of rolling either a 7 or 11 with two fair dice.

Example

In this 7/11 example, the two events are getting a 7 and getting an 11. The elementary events are the distinct results of rolling the two dice. Thus the event of a 7 is made up of the six atomic events (1,6), (2,5), (3,4), (4,3), (5,2), and (6,1).

The full sample space is the union of all thirty-six possible atomic events, the set of all pairs that result from rolling the dice.

As we see soon, because the events of getting a 7 and getting an 11 have no atomic events in common, they are independent, and the probability of their sum (union) is just the sum of their individual probabilities.

Elements of Probability Theory
The Sample Space, Probabilities, and Independence

- The probability of any event **E** from the sample space **S** is:

$$0 \le p(E) \le 1, \text{ where } E \subseteq S$$

- The *sum of the probabilities* of all possible outcomes is 1
- The probability of the *compliment* of an event is

$$p(\overline{E}) = (|S| - |E|) \ / \ |S| = (|S| \ / \ |S|) - (|E| \ / \ |S|) = 1 - p(E).$$

- The probability of the contradictory or *false* outcome of an event

$$p(\{\ \}) = 1 - p(\overline{\{\ \}}) = 1 - p(S) = 1 - 1 = 0, \text{ or alternatively,}$$
$$= |\{\ \}| \ / \ |S| = 0 \ / \ |S| = 0$$

Elements of Probability Theory
The Sample Space, Probabilities, and Independence

- A final important relationship, the probability of the union of two sets of events.

- Namely that for any two sets A and B:

$$|A \cup B| = |A| + |B| - |A \cap B|.$$

- From this relationship we can determine the probability of the union of any two sets taken from the sample space S:

$$\begin{aligned} p(A \cup B) &= |A \cup B| \; / \; |S| = (|A| + |B| - |A \cap B|) \; / \; |S| \\ &= |A| \; / \; |S| + |B| \; / \; |S| - |A \cap B| \; / \; |S| = p(A) + p(B) - p(A \cap B) \end{aligned}$$

Elements of Probability Theory
The Sample Space, Probabilities, and Independence

DEFINITION

INDEPENDENT EVENTS

Two events A and B are *independent* if and only if the probability of their both occurring is equal to the product of their occurring individually. This independence relation is expressed:

$$p(A \cap B) = p(A) * p(B)$$

We sometimes use the equivalent notation p(s,d) for $p(s \cap d)$. We clarify the notion of independence further in the context of conditional probabilities in Section 5.2.4.

Example

- Consider the situation where bit strings of length four are randomly generated.
- We want to know whether the event of the bit string containing an even number of 1s is independent of the event where the bit string ends with a 0.

- Using the multiplication principle,
- each bit having 2 values, there are a total of $2^4 = 16$ bit strings of length 4.

Example

- There are 8 bit strings of length four that end with a 0:
  {1110, 1100, 1010, 1000,0010, 0100, 0110, 0000}.

- There are also 8 bit strings that have an even number of 1s:
  {1111, 1100, 1010, 1001, 0110, 0101, 0011, 0000}.

- The number of bit strings that have both an even number of 1s and end with a 0 is 4: {1100, 1010, 0110, 0000}. Now these two events are independent since

$$p(\{\text{even number of 1s}\} \cap \{\text{end with 0}\}) = p(\{\text{even number of 1s}\}) \times p(\{\text{end with 0}\})$$

$$4/16 = 8/16 \times 8/16 = 1/4$$

The three Kolmogorov Axioms

The three Kolmogorov Axioms:

- We note that other axiom systems supporting the foundations of probability theory are possible, for instance, as an extension to the propositional calculus .
- As an example of the set-based approach we have taken, the Russian mathematician Kolmogorov (1950) proposed a variant of the following axioms, equivalent to our definitions.
- From these three axioms Kolmogorov systematically constructed all of probability theory.

1. The probability of event E in sample space S is between 0 and 1, i.e., $0 \le p(E) \le 1$.
2. When the union of all E = S, $p(S) = 1$, and $p(\overline{S}) = 0$.
3. The probability of the union of two sets of events A and B is:

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

Probabilistic Inference Example

For this example we assume we have three true or false parameters (we will define this type parameter as a boolean random variable).

- First, there is whether or not the traffic - and you - are slowing down. This situation will be labeled S, with assignment of t or f.
- Second, there is the probability of whether or not there is an accident, A, with assignments t or f.
- Finally, the probability of whether or not there is road construction at the time, C; again either t or f.

We can express these relationships for the interstate highway traffic, thanks to our car-based data download system, in Table 5.1.

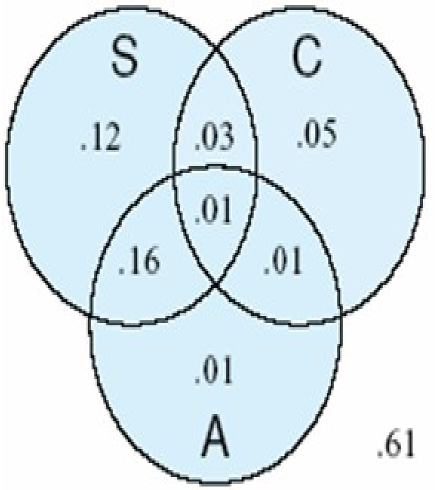
A Venn diagram representation of the probability distributions of Table 5.1; S is traffic slowdown, A is accident, C is construction.

Probabilistic Inference Example

- Table 5.1 are interpreted, of course, just like the truth tables, except that the right hand column gives the probability of the situation on the left hand side happening. Thus, the third row of the table gives the probability of the traffic slowing down and there being an accident but with no construction as 0.16:

S	C	A	p
t	t	t	0.01
t	t	f	0.03
t	f	t	0.16
t	f	f	0.12
f	t	t	0.01
f	t	f	0.05
f	f	t	0.01
f	f	f	0.61

$$S \cap \overline{C} \cap A = 0.16$$



Probabilistic Inference Example

We can also work out the probability of any simple or complex set of events.

- For example, we can calculate the probability that there is traffic slowdown S. The value for slow traffic is 0.32, the sum of the first four lines of Table 5.1; that is, all the situations where S = t.

  This is sometimes called the unconditional or marginal probability of slow traffic, S.

  This process is called marginalization because all the probabilities other than slow traffic are summed out. That is, the distribution of a variable can be obtained by summing out all the other variables from the joint distribution containing that variable.

Probabilistic Inference Example

In a like manner, we can calculate the probability of construction C with no slow-down $\overline{S}$ - a phenomenon not uncommon in the State of New Mexico! This situation is captured by $p(C \cap \overline{S}) = t$, as the sum of the 5th and the 6th lines of Table 5.1, or 0.06. If we consider the negation of the situation $C \cap \overline{S}$, we would get (using deMorgan's laws), $p(\overline{C} \cup S)$. Calculating the probability of the union of two sets, as presented in Section 5.2.1, we obtain:

$$0.16 + 0.12 + 0.01 + 0.61 + 0.01 + 0.03 + 0.16 + 0.12 - (0.16 + 0.12) = .94$$

And again the total probability of $C \cap \overline{S}$ and its complement (negation) is 1.0.

Random Variables

DEFINITION

RANDOM VARIABLE

A *random variable* is a function whose domain is a sample space and range a set of outcomes, most often real numbers. Rather than using a problem-specific event space, a random variable allows us to talk about probabilities as numerical values that are related to an event space.

BOOLEAN, DISCRETE, and CONTINUOUS RANDOM VARIABLES

A *boolean random variable* is a function from an event space to {true, false} or to the subset of real numbers {0.0, 1.0}. A boolean random variable is sometimes called a *Bernoulli trial*.

A *discrete random variable*, which includes boolean random variables as a subset, is a function from the sample space to (a countable subset of) real numbers in [0.0, 1.0].

A *continuous random variable* has as its range the set of real numbers.

Random Variables Example

- An example using a discrete random variable on the domain of Season, where the atomic events of

  Season are{spring, summer, fall, winter}, assigns 0.75,

  say, to the domain element Season = spring.

  In this situation we say

  p(Season = spring) = 0.75.

- An example of a boolean random variable in the same domain would be the mapping p(Season = spring) = true.

Random Variables - EXPECTATION OF AN EVENT

DEFINITION

EXPECTATION OF AN EVENT

If the reward for the occurrence of an event E, with probability p(E), is r, and the cost of the event not occurring, 1 – p(E), is c, then the *expectation* for an event occurring, ex(E), is:

$$ex(E) = r \times p(E) + c \times (1 - p(E))$$

Example : **Suppose that a fair roulette wheel has integers 0 through 36 equally spaced on the slots of the wheel.**

- **In the game each player places $1 on any number she chooses:**
- **if the wheel stops on the number chosen, she wins $35; otherwise she loses the dollar.**

The reward of a win is $35; the cost of a loss, $1.

Since the probability of winning is 1/37, of losing, 36/37, the expected value for this event, ex(E), is:

$$ex(E) = 35\ (1/37) + (-1)\ (36/37) \approx -0.027$$

Thus the player loses, on average, about $0.03 per play!

Conditional Probability

- *Prior* **probability and (*Conditional or Posterior*) probability**

DEFINITION

PRIOR PROBABILITY

The *prior probability*, generally an *unconditioned probability*, of an event is the probability assigned based on all knowledge supporting its occurrence or absence, that is, the probability of the event prior to any new evidence. The prior probability of an event is symbolized: p(event).

POSTERIOR PROBABILITY

The *posterior* (after the fact) *probability*, generally a *conditional probability*, of an event is the probability of an event given some new evidence. The posterior probability of an event given some evidence is symbolized: p(event | evidence).

Conditional Probability

Based on the previous definitions,

- ***The posterior probability*** **of a person having disease d, from a set of diseases D, with symptom or evidence, s, from a set of symptoms S, is:**

$$p(d|s) = p(d \cap s) / p(s).$$

There is an equivalent relationship for $p(s|d)$; again, see Figure 5.2:

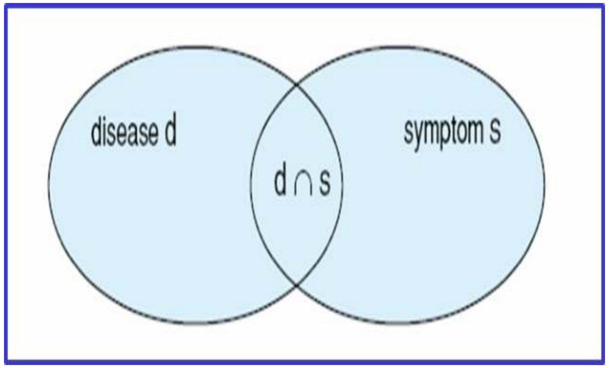
$$p(s|d) = p(s \cap d) / p(d).$$

We next solve the $p(s|d)$ equation to determine the value for $p(s \cap d)$:

$$p(s \cap d) = p(s|d)\, p(d).$$

Substituting this result in the previous equation for $p(d|s)$ produces Bayes' rule for one disease and one symptom:

$$p(d|s) = \frac{p(s|d)p(d)}{p(s)}$$



A Venn diagram illustrating the calculations of P(d | s) as a function of p(s | d).

Conditional Probability
The chain rule

- **Chain rule, an important technique used across many domains of stochastic reasoning, especially in natural language processing. We have just developed the equations for any two sets, A1 and A2:**

$$p(A_1 \cap A_2) = p(A_1 \mid A_2)\, p(A_2) = p(A_2 \mid A_1)\, p(A_1).$$

- **and now, the generalization to multiple sets Ai, called the chain rule:**

$$p(A_1 \cap A_2 \cap \ldots \cap A_n) = p(A_1)\, p(A_2 \mid A_1)\, p(A_3 \mid A_1 \cap A_2) \ldots p(A_n \mid \bigcap_{i=1}^{n-1} A_i)$$

- **We make an inductive argument to prove the chain rule, consider the *nth* case:**

$$p(A_1 \cap A_2 \cap \ldots \cap A_{n-1} \cap A_n) = p((A_1 \cap A_2 \cap \ldots \cap A_{n-1}) \cap A_n),$$

Conditional Probability
The chain rule for two sets

- We apply the intersection of two sets of rules to get:

$$p((A_1 \cap A_2 \cap \ldots \cap A_{n-1}) \cap A_n) = p(A_1 \cap A_2 \cap \ldots \cap A_{n-1})\, p(A_n \mid A_1 \cap A_2 \cap \ldots \cap A_{n-1})$$

- And then reduce again, considering that:

$$p(A_1 \cap A_2 \cap \ldots A_{n-1}) = p((A_1 \cap A_2 \cap \ldots \cap A_{n-2}) \cap A_{n-1})$$

- Until $p(A_1 \cap A_2)$ is reached, the base case, which we have already demonstrated.

CONDITIONALLY INDEPENDENT EVENTS

- **We redefine independent events (see Section 5.2.1) in the context of conditional probabilities, and then we define conditionally independent events, or the notion of how events can be independent of each other, given some third event.**

DEFINITION

INDEPENDENT EVENTS

Two events A and B are *independent* of each other if and only if $p(A \cap B) = p(A)\ p(B)$. When $p(B) \neq 0$ this is the same as saying that $p(A) = p(A|B)$. That is, knowing that B is true does not affect the probability of A being true.

CONDITIONALLY INDEPENDENT EVENTS

Two events A and B are said to be *conditionally independent* of each other, given event C if and only if $p((A \cap B) \mid C) = p(A \mid C)\ p(B \mid C)$.

Conditional Probability
Bayes’ theorem 3

- The Reverend Thomas Bayes was a mathematician and a minister. His famous theorem was published in 1763, four years after his death.
- Bayes’ theorem relates cause and effect in such a way that by understanding the effect we can learn the probability of its causes. As a result Bayes’ theorem is important both for determining the causes of diseases, such as cancer, as well as useful for determining the effects of some particular medication on that disease.

Conditional Probability
Bayes’ theorem

- We revisit one of the results of Section 5.2.4, Bayes’ equation for one disease and one symptom. To help keep our diagnostic relationships in context, we rename the variables used previously to indicate individual hypotheses, h_i, from a set of hypotheses, H, and a set of evidence, E. Furthermore, we will now consider the set of individual hypotheses hi as disjoint, and having the union of all hi to be equal to H.

$$p(h_i|E) = (p(E|h_i) \times p(h_i)) / p(E)$$

This equation may be read, “The probability of an hypothesis hi given a set of evidence E is . . .”

Conditional Probability
Bayes' theorem

- **The general form of Bayes' theorem where we assume the set of hypotheses H partition the evidence set E:**

$$p(H_i|E) = \frac{p(E|H_i) \times p(H_i)}{\sum_{k=1}^{n} p(E|H_k) \times p(H_k)}$$

$p(h_i|E)$ is the probability that h_i is true given evidence E.
$p(h_i)$ is the probability that h_i is true overall.
$p(E|h_i)$ is the probability of observing evidence E when h_i is true.
n is the number of possible hypotheses.

- ***Naïve Bayes*, or the *Bayes classifier*, that uses the partition assumption, even when it is not justified:**

$$p(E|h_j) \approx \prod_{i=1}^{n} p(e_i|h_j)$$

Conditional Probability Bayes' theorem Example

simple numerical example demonstrating Bayes' theorem

- **Suppose that you go out to purchase an automobile. The probability that you will go to dealer 1, d1, is 0.2. The probability of going to dealer 2, d2, is 0.4. There are only three dealers you are considering and the probability that you go to the third, d3, is also 0.4. At d1 the probability of purchasing a particular automobile, a1, is 0.2; at dealer d2 the probability of purchasing a1 is 0.4. Finally, at dealer d3, the probability of purchasing a1 is 0.3. Suppose you purchase automobile a1. What is the probability that you purchased it at dealer d2?**
  - **First, we want to know, given that you purchased automobile a1, that you bought it from dealer d2, i.e., to determine p(d2|a1). We present Bayes' theorem in variable form for determining p(d2|a1) and then with variables bound to the situation in the example.**

$$
\begin{aligned}
p(d_2|a_1) &= (p(a_1|d_2)\ p(d_2)) \,/\, ((d_1 \times a_1) + (d_2 \times a_1) + (d_3 \times a_1)) \\
&= (0.4)\ (0.4) \,/\, ((0.2)\ (0.2) + (0.4)\ (0.4) + (0.4)\ (0.3)) \\
&= 0.16 \,/\, 0.32 \\
&= 0.5
\end{aligned}
$$

Applications of the Stochastic Methodology
PROBABILISTIC FINITE STATE MACHINE

We present two examples that use probability measures to reason about the interpretation of ambiguous information.

First, we define an important modeling tool based on the finite state machine, the *probabilistic finite state machine*.

DEFINITION

PROBABILISTIC FINITE STATE MACHINE

A *probabilistic finite state machine* is a finite state machine where the next state function is a probability distribution over the full set of states of the machine.

PROBABILISTIC FINITE STATE ACCEPTOR

A *probabilistic finite state machine* is an *acceptor*, when one or more states are indicated as the *start* states and one or more as the *accept* states.

Applications of the Stochastic Methodology
PROBABILISTIC FINITE STATE MACHINE

It can be seen that these two definitions are simple extensions to the finite state. The addition for non-determinism is that the next state function is no longer a function in the strict sense. That is, there is no unique range state for each input value and each state of the domain. Rather, for any state, the next state function is a probability distribution over all possible next states.

Applications of the Stochastic Methodology
PROBABILISTIC FINITE STATE MACHINE Example

Example : How Is "Tomato" Pronounced?

Figure 5.3 presents a probabilistic finite state acceptor that represents different pronunciations of the word "tomato". A particular acceptable pronunciation for the word tomato is characterized by a path from the start state to the accept state. The decimal values that label the arcs of the graph represent the probability that the speaker will make that particular transition in the state machine. For example, 60% of all speakers in this data set go directly from the t phone to the m without producing any vowel phone between.

You say

[*to ow m ey tow*]

and I say

[*t ow m aa t ow*]...

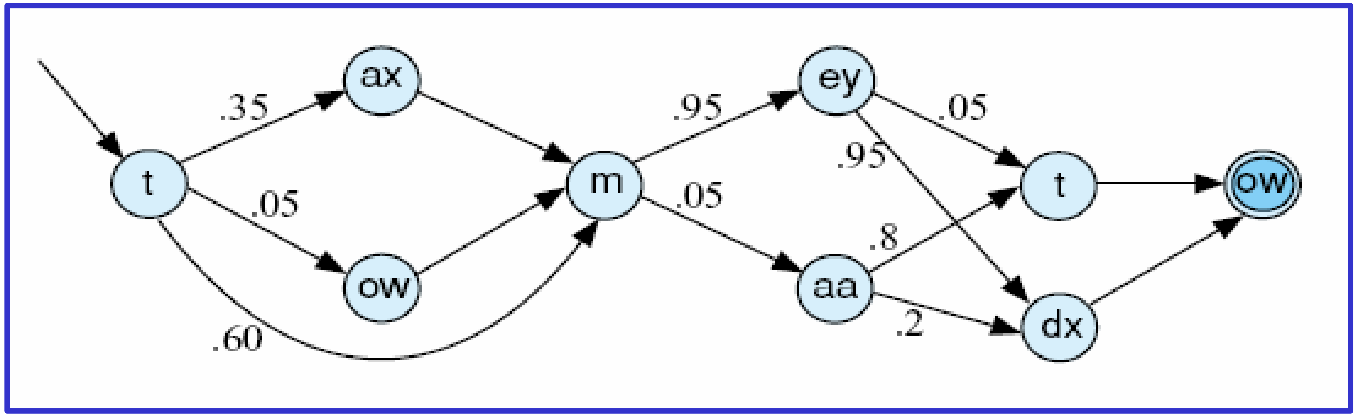


Figure 5.3 A probabilistic finite state acceptor for the pronunciation of tomato , adapted from Jurafsky and Martin (2009).

Applications of the Stochastic Methodology PROBABILISTIC FINITE STATE MACHINE Example

- **Besides characterizing the various ways people in the pronunciation database speak the word "tomato", this model can be used to help interpret ambiguous collections of phonemes. This is done by seeing how well the phonemes match possible paths through the state machines of related words. Further, given a partially formed word, the machine can try to determine the probabilistic path that best completes that word.**

Extending the Road/Traffic Example

- Figure 5.4 presents a Bayesian account of what we have just seen. road construction is correlated with orange barrels and bad traffic. Similarly, accident correlates with flashing lights and bad traffic. We examine Figure 5.4 and build a joint probability distribution for the road construction and bad traffic relationship.

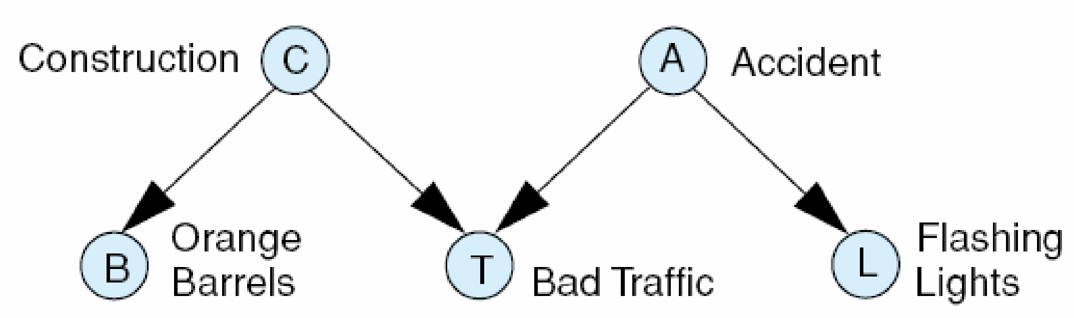


Fig 5.4 The Bayesian representation of the traffic problem with potential explanations.

Extending the Road/Traffic Example

- **We simplify both of these variables to be either true (t) or false (f) and represent the probability distribution in Table 5.4. Note that if construction is f there is not likely to be bad traffic and if it is t then bad traffic is likely. Note also that the probability of road construction on the interstate, C = true, is .5 and the probability of having bad traffic, T = true, is .4.**

C is true = .5

C	T	p
t	t	.3
t	f	.2
f	t	.1
f	f	.4

T is true = .4

Table 5.4 The joint probability distribution for the traffic and construction variables of Fig 5.3.

Extending the Road/Traffic Example

- **We next consider the change in the probability of road construction given the fact that we have bad traffic, or p(C|T) or p(C = t | T = t).**

p(C|T) = p(C = t , T = t) / (p(C = t , T = t) + p(C = f , T = t)) = .3 / (.3 + .1) = .75

- **So now, with the probability of road construction being .5, given that there actually is bad traffic, the probability for road construction goes up to .75. This probability will increase even further with the presence of orange barrels, explaining away the hypothesis of accident.**

Homework

1- if you have 16 pool balls and you want to choose 3 of them, then the combination $_nC_r$ is :

a. 16! / 3! X 13!
b. 16!
c. 13!
d. 16! / 13!

2- if A is a set with 3 elements and B is a set with 4 element and the intersection between them is 2 elements then the union between A and B is:

a. 7 elements
b. 5 elements
c. 12 elements
d. 3 elements

3- what is the result of permutations $_5P_5$:

a. 0
b.120
c. 10
d. 5

Homework

4- Let *X* represent the outcome of a roll of a six-sided die. More specifically, X will be the number of pips showing on the top face of the die after the toss. The possible values for *X* are 1, 2, 3, 4, 5, 6, all equally likely (each having the probability of 1/6). The expectation of *X* is:

a. 1/6 b. 3.5

c. 6 d. None of the above

5- If the sample space is {win, lose} and *p* (win) = 2/5. The random variable for net (reward) winnings is *X* (win) = 5, *X* (lose) = 2, then the expected value for this event is:

a. 4/5 b.2

c. 1 d. 8/5

6- If given event A {1, 3, 4, 5} and event B {2, 3, 4, 6}, then P(A|B) is :

a. 2/16 b.4/16

c. 1/16 d. 0.5

7- if given event A is 3 elements and event B is 7 elements, and P(B|A) is 0.2 then P(A|B) is :

a. (0.2 X 0.7) / 0.3 b.(0.2 X 0.3) / 0.7

c. 0.2 / 0.7 d. 0.3 / 0.7

$$E[X] = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3.5.$$

Homework

8- If given A is a set {1, 2} and B is a set {4, 5, 6}, then |A X B| by using the multiplication principle of counting is:

a. 6 . b. 9.

c. 3. d. 8.

9- If a lock have 10 numbers to choose from (0,1,..9) and you choose 3 of them, then the permutations with Repetition is:

a. 10 b. 100

c. 1000 d. 9

10- if you have 16 pool balls and you want to choose 3 of them and $_nP_r$ is 16! then the combination $_nC_r$ is :

a. 16! / 3! b. 16!

c. 13! d. 16 X 15 X 13 / 3!

11- if A is a set with 4 elements and B is a set with 4 element and the intersection between them is 2 elements then the union between A and B is:

a. 7 elements b. 6 elements

c. 12 elements d. 3 elements

12- what is the result of permutations $_6P_2$:

a. 360 b.30

c. 6 d. 4

13- if given event A is 7 elements and event B is 3 elements, and P(B|A) is 0.2 then P(A|B) is :

a. (0.2 X 0.7) / 0.3 b.(0.2 X 0.3) / 0.7

c. 0.2 / 0.7 d. 0.3 / 0.7

Homework

14- if you have 16 pool balls and you want to choose 3 of them, then the permutations without Repetition is :

a. 16! / 13! b. 16!

c. 13! d. 3!

15- If you have 16 pool balls and you want to choose 3 of them and nPr is 16! then the combination nCr is :

a. 16! / 3! b. 16!

c. 13! d. 16X15X13 / 3!

16- what is the result of permutations 6 P 2 is :

a. 360 b. 30

c. 6 d. 4

Homework

- **Based on the probabilistic finite state acceptor bellow, what is the percentage for the number of people that pronounce towmeytow?**

  a. 0.6
  b. 0.2
  c. 0.8
  d. 0.05

Suppose the fire department mandates that all fire fighters must weigh between 150 and 250 pounds. The weight of a fire fighter would be an example of a:

a. Boolean random variable
b. Discrete random variable
c. Continuous random variable
d. None of the above