Vagueness, Uncertainty, And Degrees of Belief

Shai Embon

Agenda

- Motivation
- Introduction
- Objective probability
- Subjective Probability
- Vagueness

Motivation

- The classical logic cannot handle statements that are not absolute.

- We would like to handle statements that are not total, statements that are with the form of “**Ps are always**, **purely**, **exactly and unarguably Qs**”.

Motivation

- We have seen cases where “Ps might usually be Qs”.

Example: ***Birds usually fly but not always.***

- In other cases “Ps might be fair, but not excellent examples of Qs”.

Example: we may prefer to say that some one is some what tall (and not just tall or short).

- In cases where we use physical sensors we might also have some unavoidable imprecision, as with thermometer.

Motivation

- As we seen earlier it might be hard to estimate something precisely or categorically.

- In addition to the imperfection of statements, the way we generate conclusions may also be imprecise.

Motivation

- **<u>Example:</u>** if we learn a fact or rule form other person we may need to discount for that persons untrustworthiness.
  Similarly, we may understand some system to a modest level of depth, and not be able to apply rules in 100%.

- In cases like this (where the information is equivocal and imperfect), the conclusions that we come to may not be categorical, meaning, we may not be confident in an answer, or can only come within some error range of the true answer.

Introduction

- In this lecture we will look at some of the more common ways to expand our core representation.

Introduction

- Let us start by looking at a typical sentence of the form $\forall x P(x)$,as in “everyone in my class is tall.
- We can distinguish 3 different types of modification to this logical structure, in order to make it more flexible:

1. We can relax the strength of the quantifier. Instead of “**for all x…**” we might want to say “**for 95% of x…**”, for example, “**95% of the people in class has brown eyes**”. The probability in such sentence is ***objective***, because it doesn’t subject to interpretation or degrees of confidence.

Introduction

2. We could relax the applicability of the predicate. Example: Instead of strict assertion like "**Everyone in my class is (absolutely) tall**", we could have statements like ***"Everyone in my class is somewhat tall"***.

   We call this ***vague predicates***.

Introduction

3. We could relax our degree of belief in the sentence as whole. Instead of saying, "**Everyone in the room is married**", we might say, ***"I believe that everyone in the room are married, but I am not very sure".***

   Here we are dealing with uncertain knowledge.

   When we can quantify our lack of certainty, we are using the notion of ***subjective probability***.

Objective probability

- Objective probability are about frequency.

  even though we talk in terms of probability or chance of a single event happening, we actually refer to frequency.
- The **“chance of x”** is really the percentage of times x is expected to happen out of a sequence of many events, when the basic process is repeated many times, each event is independent, and the conditions each time are exactly the same.

Objective probability

- **<u>Examples:</u>**

  The chance the next card I am dealt with will be the ace of hearts, or whether tomorrow will be raining.
- The notion of objective probability, or chance of something, is best applied to processes like coin flipping and card drawing.
- This kind of probability is called **objective** because it doesn't depend on who is assessing the probability.

Objective probability – postulates

- Probability is a number between 0 and 1, representing the frequency of an event in a large enough space of random samples.
- An event with probability 1 is considered to always to happen, and one with probability 0 is considered to never to happen.

Objective probability – postulates

More formally:

- **U** – set of all possible occurrences.
- **a** – any subset of U.
- **Pr –** function from events to the interval [0,1] satisfying the following:
  1. Pr(U) = 1.
  2. If $a_1,\ldots,a_n$ are disjoint events, then

$$\Pr(a_1 \cup \ldots \cup a_n) = \Pr(a_1) + \ldots + \Pr(a_n)$$

Objective probability – postulates

- It follows from these two postulates that:

$$\mathbf{Pr}(\bar{a}) = 1 - \mathbf{Pr}(a).$$

$$\mathbf{Pr}(\{\}) = 0.$$

$$\Pr(a \cup b) = \Pr(a) + \Pr(b) - \Pr(a \cap b)$$

- If $\mathbf{b_1, b_2 \ldots b_n}$ are disjoint events and exhaust all the possibilities then

$$\mathbf{Pr}(a) = \mathbf{Pr}(a \cap b_1) + \ldots + \mathbf{Pr}(a \cap b_n)$$

Objective probability – conditions and independence.

- A key idea in probability is conditioning, one event may depend on its interaction with another.

- We write conditional probability with the vertical bar ("|") between the events.

  Pr(a|b) means the probability of a, given that b has occurred.

Objective probability – conditions and independence.

- More formally:

$$\Pr(a|b) = \frac{\Pr(a \cap b)}{\Pr(b)}$$

- It does follow from the definition of conditioning that

$$\mathbf{Pr}(a \cap b) = \mathbf{Pr}(a|b) \times \mathbf{Pr}(b)$$

- More generally we have the following chain rule:

$\mathbf{Pr}(a_1 \cap \ldots \cap a_n) = \mathbf{Pr}(a_1 | a_2 \cap \ldots \cap a_n) \times \mathbf{Pr}(a_2 | a_3 \cap \ldots \cap a_n) \times \mathbf{Pr}(a_{n-1} | a_n) \times \mathbf{Pr}(a_n)$.

Objective probability – conditions and independence.

- We also get:

$$\mathbf{Pr}(\bar{a}|b) = 1 - \mathbf{Pr}(a|b)$$

- And the following

  If $\mathbf{b_1}\ \mathbf{b_2}\ldots\mathbf{b_n}$ are disjoint events and exhaust all the possibilities, then:

$$\mathbf{Pr}(a|c) = \mathbf{Pr}(a \cap b_1|c) + \ldots + \mathbf{Pr}(a \cap b_n|c)$$

Objective probability – conditions and independence.

- Bayes rule, uses the definition of conditional probability to relate the probability of **a** given **b** to the probability of **b** given **a**:

$$\Pr(a \mid b) = \frac{\Pr(a) \times \Pr(b \mid a)}{\Pr(b)}$$

Objective probability – conditions and independence.

- An event a is ***conditionally independent*** of event **b**, if **b** does not affect the probability of **a**, that is, if

$$\mathbf{Pr}(a|b) = \mathbf{Pr}(a)$$

Subjective Probability

- Persons subjective degree of confidence or certainty in a sentence, is separable from the content of the sentence itself.
- Regardless of how vague or categorical a sentence may be, the degree of belief in it can vary.
- Degrees of belief of this sort are often derived from observations about groups of things in the world, and the statistics of events occurring.

Subjective Probability

- Moving from statistics to graded beliefs about individuals, seems similar to the move we make from general facts about the world to defaults.

  Example: we may conclude that Tweety flies based on a belief that birds generally fly.

- This sort of conclusion tend to be all or nothing, and in subjective belief we are interested in expressing levels of confidence.

- Because degrees of belief often derive from statistical considerations, they are usually referred to as ***subjective probability***.

Subjective Probability

- In the world of subjective probability, we define two types of probability relative to drawing a conclusion:

1. The prior probability of a sentence **α** involves the prior state of information or background knowledge. (**Pr**(α|β))
2. A posterior probability is derived when new evidence is taken into account. ($Pr(\alpha|\beta \wedge \gamma)$ where γ is the new evidence)

From statistics to belief

- The traditional approach for doing so, is to find a reference class for which we have statistical information, and use the statistics about the class to compute an appropriate degree of belief.

- **Reference class** (definition) – general class into which the individual in question would fit and information about which would comfortably seem to apply.

- **Direct inference** (definition) – the move from pure statistics to belief.

From statistics to belief

Example: lets try to assign a degree of belief to the proposition “Eric is tall” where Eric is an American male, if all we knew is that:

A. 20% of American males are tall.

We would probably assign the value of 0.2 to our belief about Eric’s height.

Now lets assume that Eric is from California and that:

B. 32% of Californian males are tall.

We would probably assign a higher degree of belief, 0.32 perhaps.

From statistics to belief

Suppose we also know that:

C. 1% of hi-tech people are tall

If we don’t Eric’s occupation, should we leave our degree of belief unchanged? or should we have to estimate the probability of Eric being hi-tech person.

- Simple direct inference as we saw in the previous example, are full of problems because of multiple reference class.

A Basic Bayesian approach

- We want to have a more principled way for calculating subjective probabilities, and see how those are effected by new evidence.

- Assume we have number of propositional variables (or atomic sentences) of interest, $p_{1, \ldots,} p_n$.
  <u>**Example**</u>: $\boldsymbol{p_1}$ might be the proposition that Eric is tall, and $\boldsymbol{p_2}$ might be the proposition that Linda is rich and so on.

- In different states of the world , different combination of this sentence will be true.

A Basic Bayesian approach

- Interpretation I specifies which atomic sentences are true and which are false.
- J is defined to be joint probability distribution, which is the specification of the degree of belief for each of the 2^n truth assignments for the propositional variables.
- For each interpretation I, $J(I)$ is a number between 0 and 1 such that $\sum J(I) = 1$, where the sum is over all 2^n possibilities.

A Basic Bayesian approach

- Using the joint probability like this, we can calculate the degree of belief in any sentence involving any subset of the variables.

The degree of belief in **α** is the sum of $\mathcal{J}$ over all the interpretations where **α** is true.

We believe **α** to the extent we believe in the world states that satisfy **α** , formally: $\Pr(a) = \sum_{I \models a} J(I)$

And as before:

$$\Pr(a \mid b) = \frac{\Pr(a \wedge b)}{\Pr(b)}$$

A Basic Bayesian approach

- **Example:**

  The degree of belief that Eric is tall given that he is male and from California is:

$$\frac{\sum_{I \models t,m,c} J(I)}{\sum_{I \models m,c} J(I)}$$

t – tall.

m – male.

c – from California.

A Basic Bayesian approach

- This approach is not good enough.
- For n atomic sentences, we would need to specify the values of $2^n - 1$ numbers, this is unworkable for any practical application.

Belief Networks

- We want to cut down on what we know, so we will make some simplifying assumptions.
- We will introduce new notation:

  For sentences $p_1, \ldots, p_n$, we specify an interpretation using $\langle P_1, \ldots, P_n \rangle$, where each uppercase P_i is either p_i or $\neg p_i$.
  From this definition we see that :

$$J(\langle P_1, \ldots, P_n \rangle) = \Pr(P_1 \wedge P_2 \wedge \ldots \wedge P_n)$$

Belief Networks

- In case all the atomic sentences are conditionally independent we get:

$$J(\langle P_1,\ldots,P_n\rangle) = \Pr(P_1)\cdot\Pr(P_2)\cdots\Pr(P_n)$$

- Now we only need to know n numbers to fully specify the disjoint probability, but this assumption is too extreme.

Belief Networks

- Better idea: we would represent the variables p_i in a directed acyclic graph, which we call a ***belief network*** (or Bayesian network).
- There should be an arc from p_i to p_j if we think of the truth of p_i as directly affecting the truth of p_j .
- We say that p_i is a parent of p_j in the belief network.

Belief Networks

- We will number the variables in such a way that the parents of any variable p_j appear earlier in the ordering than p_j .

- From this we get that:

$$J(\langle P_1,\ldots,P_n\rangle) = \Pr(P_1)\cdot\Pr(P_2 \mid P_1)\cdot\Pr(P_3 \mid P_1 \wedge P_2)\cdots\Pr(P_n \mid P_1 \wedge \cdots \wedge P_{n-1})$$

- We also make the following assumption:

  Each propositional variable in the belief network is conditionally independent from the nonparent variables given the parent variable.

Belief Networks

- More formally we assume that:

$$\Pr(P_{j+1} \mid P_1 \wedge \ldots \wedge P_j) = \Pr(P_{j+1} \mid parents(P_{j+1}))$$

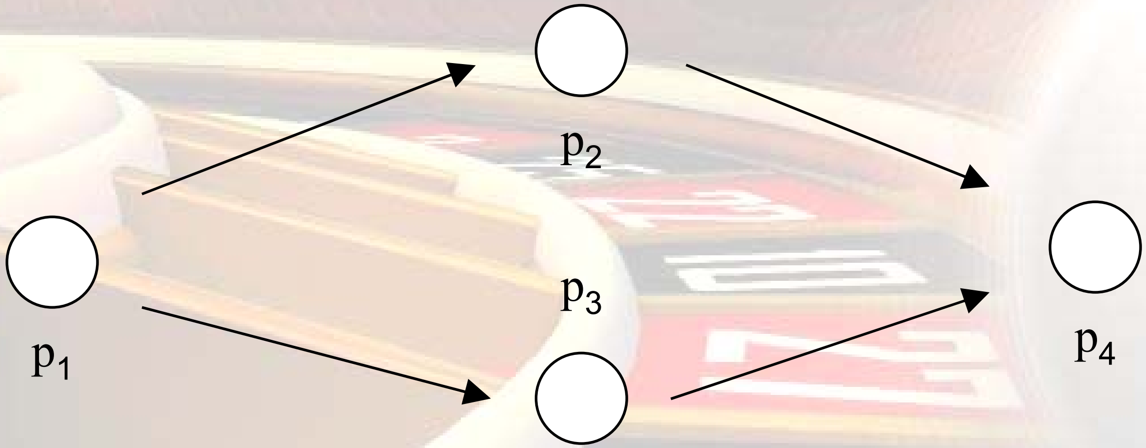
- Where parents(P_{j+1}) is the conjunction of those $P_1,\ldots,P_j$ literals that are parents of P_{j+1} in the graph.

- With those independence assumption it follows that:

$$J(\langle P_1,\ldots,P_n \rangle) = \Pr(P_1 \mid parents(P_1)) \cdots \Pr(P_n \mid parents(P_n))$$

Belief Networks-example 1

- Let us look at the following net



This graph represents the assumption that:

$$J(\langle P_1, P_2, P_3, P_4\rangle) = \Pr(P_1)\cdot\Pr(P_2 \mid P_1)\cdot\Pr(P_3 \mid P_1)\cdot\Pr(P_4 \mid P_2 \wedge P_3)$$

The full joint probability distribution is specified by:

(1+2+2+4)=9, rather then 15.

Belief Networks-example 2

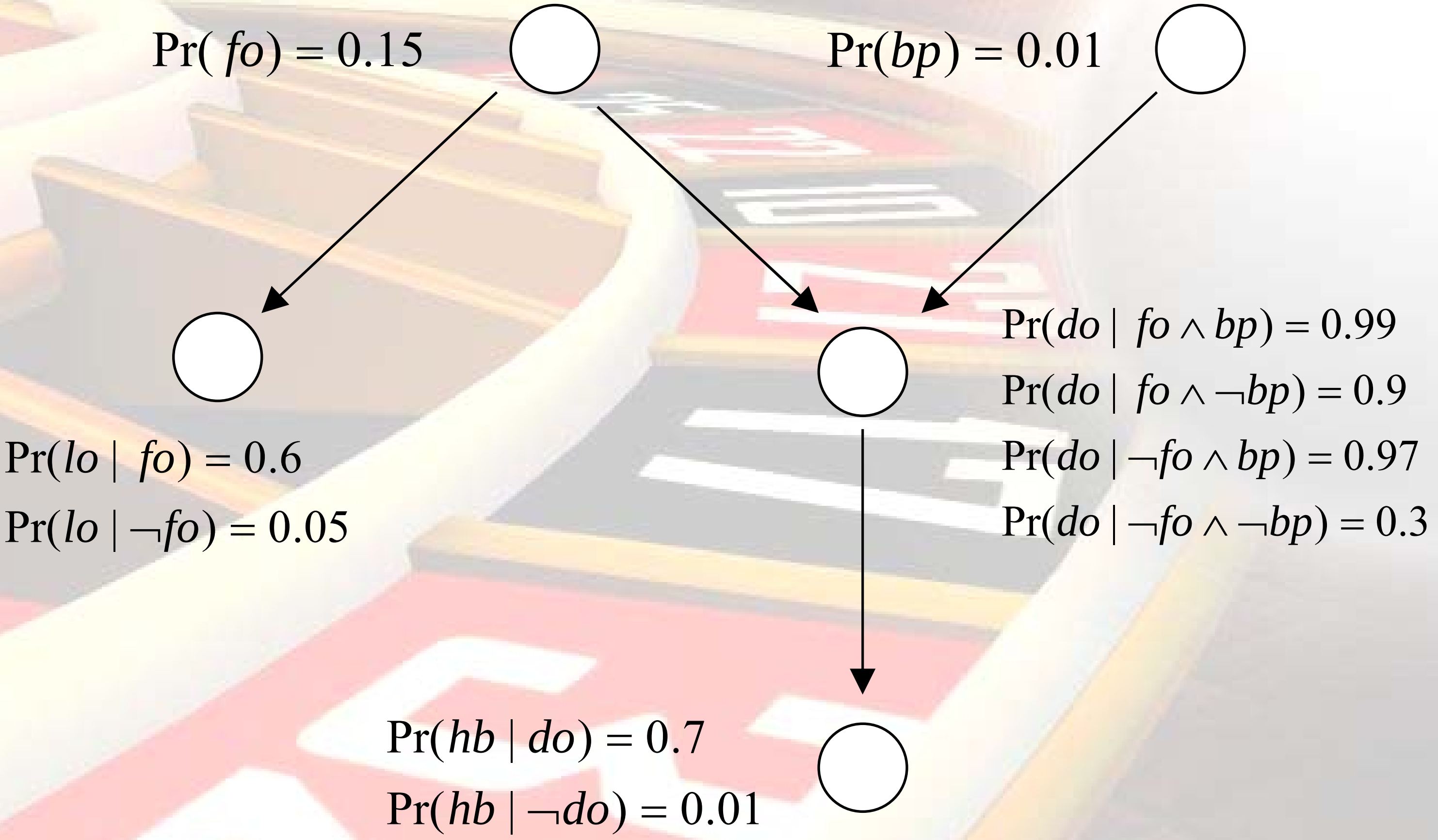
- We want to do some reasoning whether or not my family is out of the house.

- Let as look at the facts:

1. The dog is out (do) when the family is out (fo).
2. The dog is also out when it has bowel problem (bp).
3. A reasonable proportion of the time when the dog is out, you can hear him barking (hb).
4. Usually (but not always) the light on (lo) outside the house when the family is out.

Belief Networks-example 2

The corresponding belief network:



Belief Networks-example 2

- The graph represents the following assumption about the joint probability distribution:

$$J(\langle FO, LO, BP, DO, HB \rangle) =$$

$$\Pr(FO) \cdot \Pr(LO \mid FO) \cdot \Pr(BP) \cdot \Pr(DO \mid FO \wedge BP) \cdot \Pr(HB \mid DO)$$

- We want to calculate the probability that the family is out, given that the light is on but we don't hear barking: $\Pr(fo \mid lo \wedge \neg hb)$

Belief Networks-example 2

- Using the definition of conditional probability, this translates to the following:

$$\Pr(fo \mid lo \wedge \neg hb) = \frac{\Pr(fo \wedge lo \wedge \neg hb)}{\Pr(lo \wedge \neg hb)} = \frac{\sum J(\langle fo, lo, BP, DO, \neg hb \rangle)}{\sum J(\langle FO, lo, BP, DO, \neg hb \rangle)}$$

As we can see the sum in the numerator has four terms, and the sum in the denominator has eight terms.

Belief Networks-example 2

- We will compute the eight needed elements from the numbers in the figure:

1. $J(\langle fo, lo, bp, do, \neg hb \rangle) = 0.15 \times 0.6 \times 0.01 \times 0.99 \times 0.3 = 0.002673$
2. $J(\langle fo, lo, bp, \neg do, \neg hb \rangle) = 0.15 \times 0.6 \times 0.01 \times 0.01 \times 0.99 = 0.0000891$
3. $J(\langle fo, lo, \neg bp, do, \neg hb \rangle) = 0.15 \times 0.6 \times 0.99 \times 0.9 \times 0.3 = 0.024057$
4. $J(\langle fo, lo, \neg bp, \neg do, \neg hb \rangle) = 0.15 \times 0.6 \times 0.99 \times 0.1 \times 0.99 = 0.088209$
5. $J(\langle \neg fo, lo, bp, do, \neg hb \rangle) = 0.85 \times 0.05 \times 0.01 \times 0.03 \times 0.99 = 0.000123675$
6. $J(\langle \neg fo, lo, bp, \neg do, \neg hb \rangle) = 0.85 \times 0.05 \times 0.01 \times 0.03 \times 0.99 = 0.000126225$
7. $J(\langle \neg fo, lo, \neg bp, do, \neg hb \rangle) = 0.15 \times 0.6 \times 0.99 \times 0.3 \times 0.3 = 0.0378675$
8. $J(\langle \neg fo, lo, \neg bp, \neg do, \neg hb \rangle) = 0.15 \times 0.6 \times 0.99 \times 0.7 \times 0.99 = 0.029157975$

Vagueness

- Now we will look at statements like: ***"Is a man tall if his height is 1.8 meters?"***
- Obviously, the tallness of a man depends to who we are comparing him to.
- Predicates like tall that are thought of as holding to a degree are called ***vague predicates***.

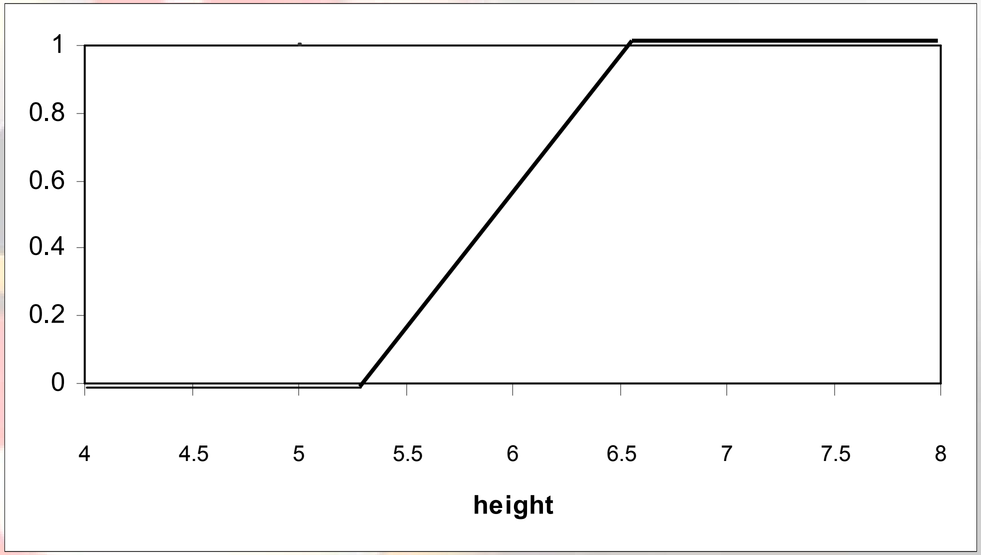
Vagueness

- These corresponds to adjectives that can be modified by the adverb “**very**", unlike, “**married**” or “**dead**”.
- We assume that for each vague predicate, there is a corresponding precise ***base function*** in terms of which the predicate is understood.

Vagueness

- Example: for “tall” the base function is “height”, for “bald” it might be “percentage hair cover”.
- ***degree curve*** – the function between the predicates and their base function.

  Example function:



Vagueness

- This definition of tall would yield the following values for various individuals and their height:

Individual	Height	Degree of tallness
Larry	4’6	0
Roger	5’6	0.25
Henry	5’9	0.5
Michael	6’2	0.9
wilt	7’1	1

Vagueness

- We need to consider Boolean combinations of vague properties and to what degree they are satisfied:
  - With negation its exactly as we know: $\Pr(\neg p) = 1 - \Pr(p)$
  - For conjunctions and disjunctions it is different from what we know:

    Example: lets say we are looking for a basketball player.

    We might be looking for someone who is tall, physically coordinated, strong and so on.

    Lets say we have a person who rates high in all those qualities.

    We would expect this person to be considered good candidate.

    If we would actually dealt with conjunction as we know we would take the product of all the degrees of his quality. Lets say we have 20 measurements and each are satisfied in level of 0.95, then we would get that the conjoined criteria to be only 0.36.

Vagueness

- There is a difference between the probability of satisfying the conjoined criterion, and the degree of which the conjoined criterion is satisfied.
- The degree of which an individual is P and Q is the **minimum** of the degrees to which the individual is P and is Q.
- Similarly , the degree to which a disjoined criterion is satisfied, is the maximum degree to which each individual criterion is satisfied.

Vagueness

- We will show the use of vague predicates in set of production rules.

- Rule is constructed as follows:
  - If x then y.
  - Example: "***If*** the apartment is big ***and*** the neighborhood is safe ***then*** the rent will be expensive.

- The term of a rule will concern quantities that can be measured or evaluated, and the consequent will concern some control action.

- Unlike standard production system where a rule either does or doesn't apply, here the terms of a rule will apply to some degree, and the control action will be affected in proportionate degree.

Vagueness

- The advantage of rules using vague predicates is that they enable inferences even when the conditions are partially satisfied.
- In this kind of system the terms apply to values from the same base functions.
- The rules are usually developed in groups and are not significant independent from one another.
- Their main goal is to work in concert to jointly affect the output variable.

Vagueness

- Example:

  we are trying to decide on a tip at restaurant based on the quality of the food and the service.

  Assume that the service and food can be described by a simple number on a linear scale.

  The amount of the tip will be the percentage of the cost of the meal.

Vagueness

- We will assume the following 3 rules:
  1. If the service is poor or the food is bad then the tip is stingy.
  2. If the service is good then the tip is normal.
  3. If the service is excellent or the food is delicious then the tip is generous.
- We assume that for each of the eight vague predicates we are given a degree curve relating the predicate to one of the three base quantities: service, food quality, or tip.

Vagueness

- The problem we wish to follow is the following: Given a specific numeric rating for the service and another specific rating for the food, calculate a specific amount for the tip, subject to these rules.
- We will see a popular method which is used to solve this problem.

Vagueness-5 step method

1. Transform the input – determine the degree to which each of the vague predicated used in the terms hold for each of the inputs.
2. Evaluate the terms – determine the degree to which each rule is applicable using the appropriate combinations for the logical operators.
3. Evaluate the consequence – determine the degree to which predicates “stingy”, “normal” and “generous” should be satisfied.

Vagueness-5 step method

4. Aggregate the consequence – obtain a single degree curve for the tip that combines the “stingy”, “normal” and “generous” ones in light of the applicability of the rules.
5. Use the aggregated degree curve to generate a weighted average value for the tip. One way to do this in our example is to take the aggregated curve from step 4 and find the center of the area under the curve.

The End