CS4850 Lecture #7 – Feb 4, 2009 **Christie Brandt (ccb35) and Chris Parsons (csp33)**

Phase transition for CNF

CNF (SAT): conjunctive normal form satisfiability

n variables

k literals per clause

c clauses

we will also use a variable 0<r<1 which satisfies the following:

c=rn.

Upper Bound for CNF-SAT:

With a random assignment to the variables, what is the probability that a given assignment is satisfied (true)?

Pr(random assignment to variable is false) = ½.

Pr(random assignment to clause is false) = $\left(\frac{1}{2}\right)^k$

Therefore, Pr(random assignment to clause is true) = $1-\left(\frac{1}{2}\right)^k$

Pr(random assignment to c clauses is satisfied) = $\left(1-\left(\frac{1}{2}\right)^k\right)^c = \left(1-\left(\frac{1}{2}\right)^k\right)^{rn}$

What is the expected number of assignments which satisfy all of the clauses?

We know there are 2^n possible assignments. Therefore,

E(number of assignments that satisfy the CNF) = $2^n\left(1-\left(\frac{1}{2}\right)^k\right)^{rn}$

Now, let $r=2^k\ln(2)$.

Then:

E(# assignments that satisfy the CNF for r =2^kln(2)) = $2^n\left(1-\left(\frac{1}{2}\right)^k\right)^{rn}$

$$=2^n\left(1-\left(\frac{1}{2}\right)^k\right)^{2^k\ln(2)(n)}$$

$$\approx 2^n e^{-n\ln(2)}$$

$$\approx 2^n e^{-n}$$

$$=1$$

Therefore, Pr(∃ an assignment that satisfies the CNF formula for r= 2^kln(2)) ≈1

However, if r> 2^kln(2), then almost surely, there is no satisfying assignment!

Using the Second Moment Method:

We would like to be able to use the second moment method to provide a lower bound for CNF-SAT. However, the second moment method is only helpful in one of the following two cases:

Second moment method is helpful if:

- The variables are statistically independent, or:
- The amount of statistical independence is very small.

On the CNF-SAT problem, it is hard or impossible to use the second moment method because clearly the assignments to the variables are not statistically independent.

Lower Bound for CNF-SAT:

Chao and Franco (1990) gives us an algorithm for finding a satisfying assignment to a CNF formula which, with very high probability, will work if $r<\frac{2^k}{k}$.

The algorithm works as follows:

Begin with all literals unassigned to boolean values.

While there are still clauses in the formula:

Choose a clause.

Examine each literal with a preassigned (non-null) value in the clause.

If the literal's value is false, cross it out (remove the literal from the clause)

If the literal's value is true, then the clause is satisfied. Cross out this clause (that is, remove the clause from the formula)

If the clause has only one unassigned (null) literal, then set it to true.

Else

Pick an unassigned literal at random and set it to true.

<u>Threshold Bounds for CNF-SAT:</u>

Using the upper bound provided by solving for the probabilities and the lower bound found by using the Chao and Franco algorithm, we know:

$$\frac{2^k}{k} \leq \text{threshold value for r} \leq 2^k \ln(2)$$

For example, when k=3, we know ($0.2667 \leq r \leq 5.5432$).

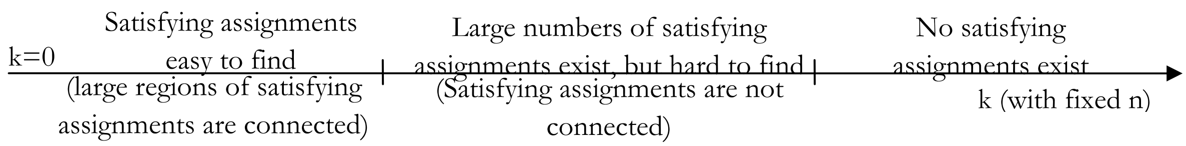
<u>Multiple Phases in CNF-SAT:</u>

CNF formulas are easy to satisfy if assignments are connected.

Two assignments are **<u>connected</u>** if they differ by one variable.

However, if a CNF formula is not connected, there may be a large number of satisfying assignments, but is very, very hard to find one.

Phases of Satisfiability of CNF formulas with a fixed number of variables as the number of clauses (k) increases:



This situation where three phases appears to exist also occurs in other problems. For instance, for the clique problem in G(n, ½), a clique of size (2-∈)log(n) almost surely exists, but we can't find it!

Homework #1 Review (High Dimensions):

<u>How to generate points on surface of sphere</u>

$X = (x_1, x_2, \ldots, x_d)$

Probability distribution for x_i = $\frac{1}{\sqrt{2\pi}} e^{\frac{-x_i}{2}}$

(This is the normal distribution of standard deviation 1 for each X_i)

This by itself would generate X at roughly distance $\sqrt{d}$ from the origin.

Prob(X) = $\left(\frac{1}{2\pi}\right)^{\frac{d}{2}} e^{\frac{-\left(x_1^2+x_2^2+\cdots+x_n^2\right)}{2}}$

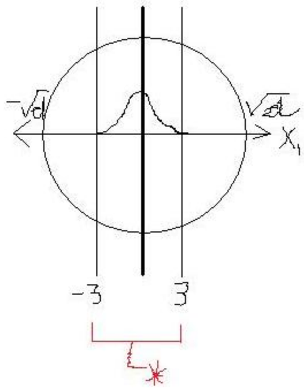
What would happen if we used some distribution other than Gaussian? If it drops off slower than the increase of the volume of the sphere, we get points out at arbitrarily large distances from the origin.

What if we want to project points on a hypersphere onto the x_1-axis? This is equivalent to dropping all coordinates besides x_1 to 0. That is:

$(x_1, x_2, \ldots, x_n) \rightarrow (x_1, 0, \ldots, 0)$

The distances of the projects from the origin are just the same as the distribution for x_1: Gaussian (roughly no points more than 3 standard deviations away from the mean of 0:

FIGURE 1



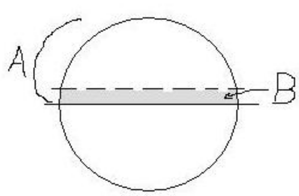
* Suggests that all the surface area is at a band around the equator.

So that's where the surface area is, but where is the volume of the sphere?

We conclude this lecture with two main facts that we will prove next lecture:

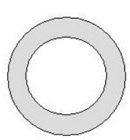
1. Most of the volume is near the equator:

FIGURE 2



2. Most of the volume is at the edge of the sphere.

FIGURE 3

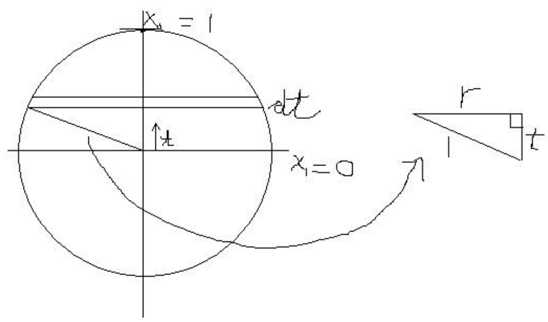


This suggests that most of the volume is at the intersection of the two.

Plan of attack: We worry only about the upper hemisphere to prove (1). If we show that B is a decreasing fraction of A in FIGURE 2.

How do we go about finding the volume of the hemisphere? We set up a variable t that goes along the x_1 axis from $x_1 = 0$ to $x_1 = 1$. We take slices and integrate these dt. Each slice is a d-1 dimensional sphere with radius r = sqrt($1-t^2$). The volume of each with respect to t are: V(d-1)*$(1-t^2)^{(d-1)/2}$

FIGURE 4



To be continued…