On the Average Case Analysis of Some Satisfiability Model Problems

by

KHALED BUGRARA and CYNTHIA BROWN

Computer Science Department

Indiana University

Bloomington, Indiana 47405

TECHNICAL REPORT NO. 143

ON THE AVERAGE CASE ANALYSIS OF SOME SATISFIABILITY

MODEL PROBLEMS

1981-82

by

KHALED BUGRARA and CYNTHIA BROWN

December, 1982

Research reported herein was supported in part by the National Science Foundation under grant number MCS 79-06110.

Abstract

The average time required by backtracking is analyzed over four models of random conjunctive form formulas . Each model gives a result of $\exp(O(v^{\frac{s-\alpha}{s-1}}))$, where v is the number of variables from which literals may be formed , s is the number of literals per clause , and v^α , $1 < \alpha < s$, is the number of clauses . This indicates that the analysis methods used are relatively insensitive to small changes in the random model.

1 Introduction

Searching among the values of a set of variables $X = \{x_1, x_2, \ldots, x_v\}$ for all solutions to the problem instance:

$$P(x_1, x_2, \ldots, x_v) = true \tag{1}$$

can be done in time $\exp(O(v))$ when exhaustive search is used. Suppose instead we introduce the intermediate predicate $P(Y)$ which is defined for all $Y \subseteq X$, such that $P(T)$ implies $P(S)$ whenever $S \subseteq T$. Let P_k be the intermediate predicate defined over the set $\{x_1, x_2, \ldots, x_k\}$ of variables, where each variable has a finite number of possible values. The intermediate predicate P_k may be false over the sequence of values $\langle V_1, V_2, \ldots, V_k\rangle$ assigned to variables $x_1, x_2, \ldots, x_k$ only if P_{k+1} is false over $\langle V_1, \ldots, V_{k+1}\rangle$ for all possible values V_{k+1} of x_{k+1} . Hence there is no extended sequence of values assigned to the rest of the variables that would satisfy (1) . Intermediate predicates that are false for most values of the variables reduce the search space significantly, thus making backtracking attractive for hard problems [3,4]. (For a more detailed discussion of backtracking and intermediate predicates see [1].)

In this paper we investigate the asymptotic behavior of the average running time of a simple version of backtracking over several sets of NP complete problems [3]. Purdom and Brown have shown that for a class of random problems with v variables, s literals per clause and v^α clauses, $1 < \alpha < s$, simple backtracking takes time $\exp(O(v^{\frac{s-\alpha}{s-1}}))$ on the average . Although the time is exponential, the improvements over exhaustive search is substantial. Our analytic studies will show the asymptotic equivalence of a number of exact formulas obtained by considering different schemes for forming the random problems. Our results contribute to the study of the quality of the simple backtracking algorithm and the previous average result by comparing the average performance of the algorithm over various classes of random problems . In section 2 we outline a description of the models used to derive the exact formulas . A detailed explanation of the asymptotic results is given in section 3, and section 4 summarizes the significance of the results .

2 Model problems

Let $X = \{x_1, x_2, \ldots, x_v\}$, so that v is the number of variables in the set. Each predicate P in the model is a formula in conjunctive normal form (CNF). A formula is in CNF if it is a conjunction $C_1 \wedge C_2 \wedge \ldots \wedge C_t$ in which each clause $C_i, 1 \le i \le t$, is the disjunction of s literals, $L_{i_1} \vee \ldots \vee L_{i_s}$, and where each literal $L_{i_j}, 1 \le j \le s$, corresponds to a variable $x_i \in X$ or its negation. We may think of CNF formulas as sets. The members of the sets are the clauses of the formulas. The formula is true iff all its clauses are true.

For a predicate P in this class, the intermediate predicate P_k is the conjunction of those clauses of P that use only variables $x_1, \ldots, x_k$. For example, if $P = \{\bar{x}_1 \vee x_1, x_1 \vee x_2\}$, then $P_1 = \{\bar{x}_1 \vee x_1\}$ and P_2 is the same as the original P.

To analyze the average performance of simple backtracking we need a representative problem set. A problem set may be obtained by considering sets of random CNF predicates. In our model a random predicate is one that is the conjunction of t clauses randomly selected from a set of random clauses. A random clause is the disjunction of randomly selected s literals . To obtain the various sets of random problems, we consider selection with and without replacement of the literals and clauses.

Consider the set of clauses that can be formed from the literals of the set $X = \{x_1, x_2, \ldots, x_v\}$ of variables. Each variable can take on the truth value *true* or *false*. Let N denote the total number of distinct predicates that are the conjunction of t such clauses. Let $f(k)$ be the total number of clauses that are false when the first k variables have been set (this number is the same for all assignments of values to the variables.) Table 1 gives the value of N and f under the different selection schemes considered.

selection scheme	N	$f(k)$	clauses with replacement?	literals with replacements?
1	$(2v)^{st}$	k^s	*Yes*	*Yes*
2	$(2^s\binom{v}{s})^t$	$\binom{k}{s}$	*Yes*	*No*
3	$\binom{(2v)^s}{t}$	k^s	*No*	*Yes*
4	$\binom{2^s\binom{v}{s}}{t}$	$\binom{k}{s}$	*No*	*No*

Table 1. The values of N and f under four selection schemes

The exact formulas given below were devloped for the average number of binary nodes in a backtrack tree produced by the simple backtracking algorithm given in [1]; the derivation will not be repeated here. The formulas (eq.(2) -eq.(5)) assume the uniform probability distribution on the predicates, and correspond to each of the four selection schemes of table 1 in order.

$$A_s(v,t) = 1 + \sum_{1<j\le v} 2^j\left(1 - \left(\frac{j-1}{2v}\right)^s\right)^t \tag{2}$$

$$B_s(v,t) = 1 + \sum_{1<j\le v} 2^j\left(1 - \frac{\binom{j-1}{s}}{2^s\binom{v}{s}}\right)^t \tag{3}$$

$$C_s(v,t) = 1 + \sum_{1<j\le v} 2^j\binom{(2v)^s - (j-1)^s}{t} \Big/ \binom{(2v)^s}{t} \tag{4}$$

$$D_s(v,t) = 1 + \sum_{1<j\le v} 2^j\binom{2^s\binom{v}{s} - \binom{j-1}{s}}{t} \Big/ \binom{2^s\binom{v}{s}}{t} \tag{5}$$

3 Asymptotic analysis

The total running time of the simple backtracking algorithm on any of the models would be the number of nodes times a polynomial in v. The above formulas were obtained by combinatorial methods. Unfortunately, we cannot find closed expressions for them. The summation expressions are difficult to evaluate even for reasonably small values of v. Thus we find it necessary to turn to asymptotic methods in order to understand the behavior of the summation formulas.

The asymptotic analysis of A_s is given in [1]. The second selection scheme gives

$$B_s(v,t) = 1 + \sum_{0<j<v} 2^{j+1}\left(1 - \frac{\binom{j}{s}}{2^s\binom{v}{s}}\right)^t \tag{6}$$

Let $t = v^\alpha$ and s be fixed. Rewrite eq.(6) as

$$B_s(v,v^\alpha) = 1 + \sum_{0<j<v} \exp((j+1)\ln 2 + v^\alpha \ln(1 - \frac{\binom{j}{s}}{2^s\binom{v}{s}})). \tag{7}$$

The first step in analyzing the asymptotic behavior is to locate the term making the main contribution. Inspection of the terms of eq.(7) suggests that the summand has a sharp peak. One way to get the value of j that maximizes eq.(7) is to take the derivative with respect to j and set it to zero. Solving the resulting equation asymptotically as $v \to \infty$ shows that the term making the largest contribution occurs when

$$j = 2(\frac{2\ln 2}{s})^{\frac{1}{s-1}} v^{\frac{s-\alpha}{s-1}} + O(1), \quad 1 < \alpha < s. \tag{8}$$

(Appendix A1 has the details of the derivation of formula (8).)

We will find it more convenient to have the main contribution at the value 0, so let k be the value of j that is given in eq.(8)

$$k = 2(\frac{2\ln 2}{s})^{\frac{1}{s-1}} v^{\frac{s-\alpha}{s-1}} + O(1) + x, \quad -k \le x \le v - k - 1.$$

Now we may study the behavior of (7) as a function of x. On substitution in eq.(7) we obtain

$$\exp((k + x + 1)\ln 2 + v^\alpha \ln(-\frac{\binom{k+x}{s}}{2^s\binom{v}{s}})). \tag{9}$$

Next we break down the binomials using the fact that $\binom{n}{r} = \frac{n!}{r!(n-r)!}$, and using the Stirling approximation to get rid of the factorials:

$$\begin{aligned}\frac{\binom{k+x}{s}}{2^s\binom{v}{s}} &= \frac{(k+x)!\,(v-s)!}{2^s v!\,(k+x-s)!}\\ &= \exp((\frac{k+x}{2v})^s + (s-k-x-\frac{1}{2})\ln(1-\frac{s}{k+x}) + (v-s+\frac{1}{2})\ln(1-\frac{s}{v}) + O(v^{2\frac{\alpha-s}{s-1}}))\end{aligned} \tag{10}$$

Expanding the last two logarithms in power series and collecting the individual terms, eq.(10) may be rewritten as:

$$(\frac{k+x}{2v})^s \exp(\sum_{n\ge 0} C_n + \frac{s(s-1)}{2} v^{-1} + O(v^{2\frac{\alpha-s}{s-1}})), \tag{11}$$

where

$$C_n = (-1)^{n+1}\frac{s(s-1)}{2}2^{-(n+1)}\left(\frac{2\ln 2}{s}\right)^{-\frac{n+1}{s-1}}v^{(n+1)\frac{\alpha-s}{s-1}}, \quad 1 < \alpha < s.$$

Expanding the exponential in eq.(11) gives

$$\left(\frac{k+x}{2v}\right)^s\left(1+\sum_{n\geq 0} C_n x^n + \frac{s(s-1)}{2}v^{-1} + O\left(v^{2\frac{\alpha-s}{s-1}}\right)\right). \tag{12}$$

Now expanding the natural log in eq.(9) , using eq.(12) and multiplying out we get

$$\begin{aligned}
1 + &\sum_{-k<x\leq v-k-1} \exp\left((k+x+1)\ln 2 - v^\alpha \sum_n \sum_{p\geq 1} \frac{1}{p}\binom{ps}{n}\left(\frac{k}{2v}\right)^{ps-n}\left(\frac{x}{2v}\right)^n\right. \\
&+ \sum_{n\geq 0} \frac{s(s-1)}{2}2^{-(n+1)}\binom{s-1}{n}\left(\frac{2\ln 2}{s}\right)^{1-\frac{n}{s-1}}v^{n\frac{\alpha-s}{s-1}}x^n \\
&- \sum_{n\geq 0} 2^{-n}\binom{s}{n}\left(\frac{2\ln 2}{s}\right)^{\frac{s-n}{s-1}}v^{\frac{\alpha-s}{s-1}(n-1)-1}x^n \\
&+ \left.O\left(v^{\frac{\alpha-s}{s-1}}\right)\right).
\end{aligned} \tag{13}$$

Let

$$\begin{aligned}
h(x) &= a_2x^2 + a_3x^3 + \cdots, \\
g(x) &= \sum_{n\geq 0} \frac{s(s-1)}{2}2^{-(n+1)}\binom{s-1}{n}\left(\frac{2\ln 2}{s}\right)^{1-\frac{n}{s-1}}v^{n\frac{\alpha-s}{s-1}}x^n \\
r(x) &= -\sum_{n\geq 0} \frac{s(s-1)}{2}2^{-n}\binom{s}{n}\left(\frac{2\ln 2}{s}\right)^{\frac{s-n}{s-1}}v^{\frac{\alpha-s}{s-1}(n-1)-1}x^n.
\end{aligned} \tag{14}$$

where $a_i = -\sum_{p\geq 1}\frac{1}{p}\binom{ps}{i}\left(\frac{k}{2v}\right)^{ps-i}(2v)^{-i}$. We can remove the factor $2^{s+1}\exp(k\ln 2 + v^\alpha\ln(1-(\frac{k}{2v})^s))$ from eq.(13) . This leaves us with the following formula to evaluate over the range $x = \theta = -k, \ldots, x = \beta = v - k - 1$:

$$\sum \exp(v^\alpha h(x))\left(1 + g(x) + r(x) + O\left(v^{\frac{\alpha-s}{s-1}}\right)\right). \tag{15}$$

The Euler summation formula [5] tells us that (see Appendix A2)

$$\sum \exp\left(v^\alpha h(x)\right)q(x) = \int_{-\infty}^{+\infty} \exp\left(v^\alpha h(x)\right)q(x)\,dx + \text{ exponentially small terms }, \tag{16}$$

for which DE Bruijn[2] gives

$$\int_{-\infty}^{+\infty} \exp\left(v^\alpha h(x)\right)q(x)\,dx = \sum_{i\geq 0} d_i v^{(-\frac{1}{2}-i)\alpha}, \qquad v\to\infty, \tag{17}$$

where

$$d_i = (-a_2)^{-i-\frac{1}{2}}\sum_{m=0}^{2i} C_{m,2i-m}(-a_2)^{-m}\Gamma(m+i+\frac{1}{2});$$

and the $C_{m,n}$'s are defined in Appendix A2; recall that $\Gamma(\frac{1}{2}) = \pi$, and for $j = 1, 2, \ldots$ we have

$$\Gamma(j + \frac{1}{2}) = \prod_{i \geq 1}^{j} (i - \frac{1}{2}).$$

Notice that the main term in eq.(17) is

$$d_0 = \sqrt{-2\pi / v^{\alpha} h''(0)} \times q(0). \tag{18}$$

Applying eq. (16) and eq. (17) to eq. (15) we end up with

$$4\left[\sqrt{\frac{2\pi}{s(s-1)}}\left(\frac{s}{2\ln 2}\right)^{\frac{s-2}{2(s-1)}} v^{\frac{s-\alpha}{2(s-1)}} - \frac{s(s-1)}{2}\sqrt{\frac{2\pi}{s(s-1)}}\left(\frac{2\ln 2}{s}\right)^{\frac{s+2}{2(s-1)}} v^{\frac{3}{2}\frac{s-\alpha}{s-1}-1} + O(v^{\frac{\alpha-s}{2(s-1)}})\right];$$

hence

$$\begin{aligned} B_s(v, v^{\alpha}) = 1 + 2^{\frac{(s-1)}{2}} \exp\left((s-1)\left(\frac{2\ln 2}{s}\right)^{\frac{s}{s-1}} v^{\frac{s-\alpha}{s-1}} - \frac{1}{2}\left(\frac{2\ln 2}{s}\right)^{\frac{2s}{s-1}} v^{\frac{s}{s-1}(1-\alpha)+\frac{s-\alpha}{s-1}}\right) \\ \times \left[\sqrt{\frac{2\pi}{s(s-1)}}\left(\frac{s}{2\ln 2}\right)^{\frac{s-2}{2(s-1)}} v^{\frac{s-\alpha}{2(s-1)}} - \frac{s(s-1)}{2}\sqrt{\frac{2\pi}{s(s-1)}}\left(\frac{2\ln 2}{s}\right)^{\frac{s+2}{2(s-1)}} v^{\frac{3}{2}\frac{s-\alpha}{s-1}-1} + O(v^{\frac{\alpha-s}{2(s-1)}})\right] \end{aligned}$$

For example

$$B_3(v, v^{1.5}) = \exp(0.62824 v^{3/4})[9.45135 v^{3/8} - 8.9066 v^{1/8} + O(v^{-3/8})]$$

Similarly we obtain an asymptotic power series for formula (4) given by

$$\begin{aligned} C_s(v, v^{\alpha}) = 1 + \exp((s-1)\left(\frac{2\ln 2}{s}\right)^{\frac{s}{s-1}} v^{\frac{s-\alpha}{s-1}} - \frac{1}{2}\left(\frac{2\ln 2}{s}\right)^{\frac{2s}{s-1}} v^{\frac{s}{s-1}(1-\alpha)+\frac{s-\alpha}{s-1}}) \\ \left[4\sqrt{\frac{2\pi}{s(s-1)}}\left(\frac{s}{2\ln 2}\right)^{\frac{s-2}{2(s-1)}} v^{\frac{1}{2}\frac{s-\alpha}{s-1}} + O(1)\right] \end{aligned} \tag{20}$$

And

$$C_3(v, v^{1.5}) = \exp(0.62824 v^{3/4})[4.7256 v^{3/8} + O(v^{-3/8})].$$

The derivation of formula 5 is in the appendix where we have shown its asymptotic equivalence to formula 3. Finally we should note that the derived formulas were verified using computer calculations by computing the values of the exact and the asymptotic formulas and showing that the ratio is converging to one as v increases.

4 Conclusions

Simple backtracking was analyzed over the set of CNF predicates with s literals per clause and v^{α} clauses in the predicate, $1 < \alpha < s$, where v distinct variables and their negations are available for selection. We showed that when all the schemes were considered for forming the random predicates, and when searching for all possible solutions over the problem set, the the average number of nodes in a backtrack tree is

$$\exp((s-1)(\frac{2\ln 2}{s})^{\frac{s}{s-1}} v^{\frac{s-\alpha}{s-1}}) \times \text{polynomial in } v.$$

Thus, we establish the asymptotic equivalence of eq.(2) through eq.(5) . Moreover, formulas (2) and (4) have the same asymptotic polynomial expansion with slightly varying coefficients, and formulas (3) and (5) have exactly the same asymptotic expansion. The otherwise insignificant improvement of the latter pair over the former seems to depend on the literals being distinct within the clauses.

Each of the selection schemes gives a different set of predicates. However, the numbers of the nodes in the backtrack trees they produce differ at most polynomially from one another, so it is the case that the exponential performance of the algorithm is the same under all the schemes studied. We believe that these schemes are "reasonable", in the sense that it is not easy to think of reasonable selection processes that are not polynomially related to them.

In summary, so long as we restrict ourselves to the standard model of CNF formulas, the exponential behavior of the average time for simple backtracking will be unaffected by the particular scheme used for forming the predicates. This suggests that we can make our choice for the analysis of more complex backtracking algorithms based on convenience without affecting the results in any significant way.

5 Appendix A1

Taking the derivative of formula (6) with respect to j and setting it to zero, we get:

$$2^s \ln 2\binom{v}{s} = (j)^{\underline{s}}(\frac{v^\alpha}{s!} \sum_{0\le i\le s-1} \frac{1}{j-i} + \frac{\ln 2}{s!})$$

using the fact that

$$(j)^{\underline{s}} = j(j-1)\ldots(j-s+1) = j^s \sum_i (-1)^i \left[{s \atop s-i}\right]\left(\frac{1}{j}\right)^i,$$

and

$$\frac{d}{dj}\binom{j}{s} = \frac{1}{s!} \sum_{0\le i\le s-1} \frac{(j)^{\underline{s}}}{j-i},$$

where $[\ ]$ denotes Stirling numbers of the first kind [5]. Now, successive approximation methods give :

$$j = 2\left(\frac{2\ln 2}{s}\right)^{\frac{1}{s-1}} v^{\frac{s-\alpha}{s-1}} + \frac{s(s-2)}{2(s-1)} + O\left(v^{\frac{\alpha-s}{s-1}}\right),$$

$$1 < \alpha < s, \quad v \to \infty,$$

using

$$\sum_{0\le i\le s-1} \frac{1}{j-i} \approx \frac{s}{j} + \frac{s^2}{j^2},$$

$$\frac{\ln 2}{s!} \approx 0 \qquad \text{and} \qquad \binom{v}{s} \approx \frac{v^s}{s!} \qquad \text{as } v \to \infty .$$

6 APPENDIX A2

Let

$$Q(x) = U(x)\exp(v^{\alpha} h(x)),$$

$$h(x) \leq -dx^2 \qquad (\theta \leq x \leq \beta) \qquad \text{for some } d > 0 \tag{21}$$

where h is as defined in section 3, and U is the sum of a convergent power series.

One way to obtain an asymptotic power series for $\sum Q(x)$ is to consider $\exp(a_2 v^{\alpha} x^2)$ as the main factor of Q . The remaining factor

$$U(x)\exp(v^{\alpha} x^3(a_3 + a_4 x + \cdots)) \tag{22}$$

can be expanded in a double power series in the two arguments x and $v^{\alpha} x^3$. Therefore we have :

$$\sum Q(x) = \sum H(x, v^{\alpha} x^3)\exp(a_2 v^{\alpha} x^2).$$

Where

$$H(x, v^{\alpha} x^3) = \sum_{\substack{0 \leq m < \infty \\ 0 \leq n < \infty}} C_{m,n}(v^{\alpha} x^3)^m x^n$$

and

$$C_{m,n} = \frac{1}{m!} \sum_{0 \leq j \leq n} c_j b_{n-j}.$$

Notice that

$$c_0 = a_3^m, \quad c_k = \frac{1}{k a_3} \sum_{1 \leq i \leq k} (im - k + i) a_{i+2} c_{k-i+2}, \qquad k \geq 1.$$

The Euler summation formula can be written

$$\sum_{\theta \leq x \leq \beta} f(x) = \int_{\theta}^{\beta} f(x)\, dx - \frac{1}{2} f(x)|_{\theta}^{\beta}$$
$$+ \frac{1}{2} B_2 . f^{(1)}(x)|_{\theta}^{\beta} + \cdots + \frac{1}{m+1} B_{m+1} \frac{f^{(m)}(x)}{m!}|_{\theta}^{\beta} + R_{m+1},$$

where $R_{m+1} = (-1)^{m+1}/m! \int_{\theta}^{\beta} B_m(\{y\}) f^m(y)\, dy$, $f^{(m)}$ indicates the mth derivative, $\{y\} = y - \lfloor y \rfloor$, and the B_i are the Bernouli numbers and polynomials[5]. The error term is $O(v f^{(m)}(z))$, where z is the value of y that maximizes $f^{(m)}$. Let $a = a_2 v^{\alpha}$, p and q be nonnegative integers. Then

$$Q(x) = C_{p,q} x^p (v^{\alpha} x^3)^q e^{-ax^2}.$$

The Euler summation formula gives an asymptotic power series for $\sum Q(x)$ as $v \to \infty$ since

$$Q^{(m)} = a_2^{\frac{m-p-3q}{2}} v^{aq} g(ay), \quad g(x) = e^{-x^2} x^{p+3q}.$$

Now $g(x)$ is a well behaved function independent of a. So the error term is $O(a_2^{\frac{m-p-3q}{2}} v^{aq+1}) = O(v^{1+\frac{m-p+3q}{2}(\frac{\alpha-s}{s-1})})$, $1 < \alpha < s$. It is easy to see that $Q^{(m)}(-\infty) = Q^{(m)}(+\infty) = 0$, we thus have reduced the summation problem to one about integrals:

$$\sum_{\theta \leq x \leq \beta} Q(x) = \int_{-\infty}^{+\infty} Q(x)\, dx + O(1).$$

Notice that $\int_{-k}^{v-k} Q(x)\, dx$ differ from $\int_{-\infty}^{+\infty} Q(x)\, dx$ by an exponentially small amount.

7 APPENDIX A3

We derive the expression C_s which is given by (20), as follows: Let $t = v^{\alpha}$, s fixed, $1 < \alpha < s$ and v increasing. We proceed as before by considering the value of j that maximizes the ratio in (4). The maximum value of the summand is reached when

$$j = 2(\frac{2\ln 2}{s})^{\frac{1}{s-1}} v^{\frac{s-\alpha}{s-1}} (1 + O(v^{\frac{\alpha-s}{s-1}}) + O(v^{\frac{s}{s-1}(1-\alpha)})) \tag{23}$$

If we shift the index in eq. (4) so that it is zero at the largest term, then

$$r_0 = \binom{(2v)^s - (k+x)^s}{v^{\alpha}} \Big/ \binom{(2v)^s}{v^{\alpha}} = \frac{((2v)^s - (k+x)^s)!\, ((2v)^s - v^{\alpha})!}{((2v)^s - (k+x)^s - v^{\alpha})!\, (2v)^s!}.$$

Where we let k denote eq. (23) and eq. (4) will be treated as a function of x. Using the Stirling formula for the factorials, and after laborious calculations we obtain :

$$\begin{aligned} r_0 = ((2v)^s + \frac{1}{2})&\Big[\ln\Big(1 - (\frac{k+x}{2v})^s\Big) + \ln\Big(1 - \frac{v^{\alpha}}{(2v)^s}\Big) - \ln\Big(1 - (\frac{k+x}{2v})^s - \frac{v^{\alpha}}{(2v)^s}\Big)\Big] \\ &+ (k+x)^s\Big[\ln\Big(1 - (\frac{k+x}{2v})^s + \frac{v^{\alpha}}{(2v)^s}\Big) - \ln\Big(1 - (\frac{k+x}{2v})^s\Big)\Big] \\ &+ v^{\alpha}\Big[\ln\Big(1 - (\frac{k+x}{2v})^s - \frac{v^{\alpha}}{(2v)^s}\Big) - \ln\Big(1 - \frac{v^{\alpha}}{(2v)^s}\Big)\Big] \\ &+ O(v^{-s}) \qquad \text{as } v \to \infty \end{aligned} \tag{24}$$

Combining the natural logarithms in (24), expanding the resultant expression in a power series, and using the binomial theorem as many times as needed we get:

$$\ln r_0 = ((2v)^s + \frac{1}{2}) \sum_{i \geq 1} \sum_{q,l} \frac{(-1)^{i+q-1}}{i} \binom{-i}{q} \binom{q}{l} \binom{s(i+1)}{n} y^{i+q-1} u^{s(i+1)} (\frac{x}{2v})^n \tag{25}$$

$$- (2v)^s \sum_{i \geq 1} \sum_{q,n} \frac{(-1)^q}{i} \binom{-i}{q} \binom{s(q+1)}{n} y^i u^{s(q+1)-n} (\frac{x}{2v})^n \tag{26}$$

$$- v^{\alpha} \sum_{i \geq 1} \sum_{q,n} \frac{1}{i} \binom{i}{q} \binom{sq}{n} y^{i-q} u^{sq-n} (\frac{x}{2v})^n \tag{27}$$

$$+ v^{\alpha} \sum_{i \geq 1} \frac{1}{i} y^i \tag{28}$$

$$+ O(v^{-s}).$$

Where $y = v^\alpha/(2v)^s$, $u = j/2v$. Now eq. (25) through eq. (28) are in a convenient form. To see why rewrite eq. (27) as

$$-v^\alpha \sum_{i\geq 1}\sum_{q|i=q}\sum_{n}\frac{1}{i}\binom{i}{q}\binom{sq}{n}y^{i-q}u^{sq-n}(\frac{x}{2v})^n + v^\alpha \sum_{i\geq 1}\sum_{q<i}\sum_{n}\frac{1}{i}\binom{i}{q}\binom{sq}{n}y^{i-q}u^{sq-n}(\frac{x}{2v})^n$$

$$= -v^\alpha \sum_{i\geq 1}\sum_{0<q<i}\sum_{n}\frac{1}{i}\binom{i}{q}\binom{sq}{n}y^{i-q}u^{sq-n}(\frac{x}{2v})^n \tag{29}$$

$$-v^\alpha \sum_{i\geq 1}\binom{si}{n}u^{si-n}(\frac{x}{2v})^n \tag{30}$$

$$-v^\alpha \sum_{i\geq 1}\frac{1}{i}y^i. \tag{31}$$

Notice that (31) cancels with (28), so we consider formula (30), and need only the most important term in (29). Therefore (4) reduces to

$$C_s(v,v^\alpha) = 1 + \sum_{\theta\leq x\leq\beta}\exp\left((k+x+1)\ln 2 - v^\alpha\sum_{i\geq 1}\sum_{n}\frac{1}{i}\binom{si}{n}u^{si-n}(\frac{x}{2v})^n\right)$$
$$\exp\left(-\frac{1}{2}v^\alpha y\sum_{n}\binom{s}{n}u^{s-n}(\frac{x}{2v})^n\right)(1+O(v^{\alpha-s})+O(v^{\frac{s}{s-1}(1-\alpha)})+O(v^{-s}))$$

Expanding the second exponential, where $e^{-a} = (1 - a + O(a^2))$ as $a \to 0$, we have

$$1 + \sum_{\theta\leq x\leq\beta}\exp\left((k+x+1)\ln 2 - v^\alpha\sum_{i\geq 1}\sum_{n}\frac{1}{i}\binom{si}{n}u^{si-n}\frac{x}{2v}^n\right)$$
$$\left[1-\frac{1}{2}yv^\alpha\sum_{n}\binom{s}{n}u^{s-n}(\frac{x}{2v})^n(1+O(v^{\alpha-s})+O(v^{\frac{s}{s-1}(1-\alpha)}))+O(v^{-s})\right]$$

Taking the constant terms outside the sum, and cancelling out the term proporitonal to x, the remaining expression to evaluate is of the following form :

$$\sum e^{v^\alpha h(x)}(1-r(x)+O(f(v))),\ \text{for some function } f\to 0,\ \text{as } v\to\infty. \tag{32}$$

Using the Euler summation formula we find that (11) has the asymptotic expansion shown in formula (20) :

8 Appendix A4 Derivation of formula (5)

Rewrite the expression in (5) using factorials , and changing the index in the formula

$$D(v,t) = 1 + \sum_{0<j<v} 2^j\frac{(2^s\binom{v}{s}-\binom{j}{s})!\,(2^s\binom{v}{s}-t)!}{(2^s\binom{v}{s}-\binom{j}{s})!\,(2^s\binom{v}{s})!}.$$

Let

$$r_j = \frac{(2^s\binom{v}{s} - \binom{j}{s})!\,(2^s\binom{v}{s} - t)!}{(2^s\binom{v}{s} - \binom{j}{s})!\,(2^s\binom{v}{s})!}.$$

Expanding r_j using the Stirling formula for the factorials , we get

$$\begin{aligned}
\ln r_j = {} & (2^s\binom{v}{s} + \frac{1}{2})\left(\ln\left(1 - \frac{\binom{j}{s}}{2^s\binom{v}{s}}\right) + \ln\left(1 - \frac{t}{2^s\binom{v}{s}}\right) - \ln\left(1 - \frac{\binom{j}{s}}{2^s\binom{v}{s}} - \frac{t}{2^s\binom{v}{s}}\right)\right) \\
& + \binom{j}{s}\left(\ln\left(1 - \frac{\binom{j}{s}}{2^s\binom{v}{s}} - \frac{t}{2^s\binom{v}{s}}\right) - \ln\left(1 - \frac{\binom{j}{s}}{2^s\binom{v}{s}}\right)\right) \\
& + t\left(\ln\left(1 - \frac{\binom{j}{s}}{2^s\binom{v}{s}} - \frac{t}{2^s\binom{v}{s}}\right) - \ln\left(1 - \frac{t}{2^s\binom{v}{s}}\right)\right) \\
& + O(v^{-s}) \qquad \text{as } v \to \infty .
\end{aligned}$$

Combining the natural logarithms as appropriate and expanding, we obtain

$$\begin{aligned}
D_s(v,t) = 1 + \sum_{0<j<v} \exp(j \ln 2 + {} & (2^s\binom{v}{s} + \frac{1}{2}) \sum_{i\geq 1} \sum_{q,l} \frac{1}{i}(-1)^{i+q-1}\binom{-i}{q}\binom{q}{l}\left(\frac{t}{2^s\binom{v}{s}}\right)^{i+q-l}\left(\frac{\binom{j}{s}}{2^s\binom{v}{s}}\right)^{i+1} \\
& - \binom{j}{s} \sum_{i\geq 1} \sum_{q} \frac{1}{i}(-1)^{q}\binom{-i}{q}\left(\frac{t}{2^s\binom{v}{s}}\right)^{i}\left(\frac{\binom{j}{s}}{2^s\binom{v}{s}}\right)^{q+1} \\
& - t \sum_{i\geq 1} \sum_{q} \frac{1}{i}\binom{i}{q}\left(\frac{t}{2^s\binom{v}{s}}\right)^{i-q}\left(\frac{\binom{j}{s}}{2^s\binom{v}{s}}\right)^{q} \\
& + t \sum_{i\geq 1} \frac{1}{i}\left(\frac{t}{2^s\binom{v}{s}}\right)^{i} + O(v^{-s})).
\end{aligned}$$

In a similar technique to that of appendix A2, we get:

$$D_s(v, v^\alpha) = 1 + \sum_{1<j\leq v} 2^j\left(1 - \frac{\binom{j-1}{s}}{2^s\binom{v}{s}}\right)^{v^\alpha} + O(1), \text{ as } v \to \infty .$$

Therefore D_s is asymptotically equivelent to B_s.

References

[1] Cynthia A. Brown and Paul W. Purdom, *An average time analysis of backtracking,* SIAM J. Comput. 10(1981).

[2] N. G. DeBruijn, *Asymptotic methods in analysis,* third edition, North-Holland Publishing Co. 1970.

[3] Michael R. Garey and David S. Johnson, *Computers and Intractability,* W. H. Freeman and Co., San Franciso (1979).

[4] Donald E. Knuth, *Estimating the efficiency of backtracking programs,* Math. Comput. 29 (1975), pp. 121–136.

[5] Donald E. Knuth, *The Art of Computer Programming, Vol. 1, Fundamental Algorithm,* Addison-Wesley, Reading, second edition (1973).