Validity of Normality Assumption in CSP Research

Alvin C. M. Kwan
Department of Computer Science, University of Essex
Wivenhoe Park, Colchester, Essex CO4 3SQ, United Kingdom

email: alvin@essex.ac.uk

December 1995

Abstract

There are many new methods for solving constraint satisfaction problems proposed in recent years. Due to their complexity, a theoretical analysis on their average-case behaviours seems to be very difficult. Researchers tend to adopt an empirical approach to evaluate constraint satisfaction techniques. When empirical results are ready, statistical techniques are often employed for analysis. The question is which statistics to use. Some recent research uses parametric tests such as *t*-test and ANOVA. However those tests assume that the characteristic of the normal curve can be applied. In this paper, we provide evidence that the normality assumption is often not valid in the results produced by a range of constraint satisfaction algorithm-heuristic combinations on random binary constraint satisfaction problems and 3-colouring problems, particularly when a problem is within the "mushy region", which are popular benchmark problems for evaluating CSP methods. The failure of normality assumption highlights the need for some statistics which do not rely on the normality assumption to analyse empirical results from CSP research. We believe that non-parametric techniques could be the right tools for the purpose.

1 Introduction

A constraint satisfaction problem (CSP) can be informally defined as a set of variables, each of them associated with a set of feasible domain values which the variable can be instantiated to, and a set of constraints, each of them imposes some restrictions on the values of its constrained variables can take. The aim is to find a solution to the problem such that every variable is assigned a value without violating any constraints [Tsang 93].

Many new CSP methods, notably new algorithms or heuristics, have been proposed in recent years. Since analysing the average-case behaviour of those techniques is very difficult, most evaluations are done in an empirical way. The most frequently adopted performance indicators for evaluating CSP methods appear to be the number of compatibility checks and the cpu time that they need to solve a problem. In a typical CSP method evaluation exercise, sets of problems with each set usually contains a few dozen to a few thousand problem instances which are characterised by some parameterised value are generated and each CSP method under investigation is applied to solve them.

After collecting the required data from an experiment, the immediate question is how to analyse them. The most frequently used statistics are simple measures such as mean or median. However mean is vulnerable to outliers whereas median totally ignores them. Parametric tests such as ANOVA and *t*-test are used in some research [Freuder & Wallace 91] [Frost & Dechter 95]. However such parametric tests often rely on the fact that the characteristics of the normal curve can be applied. The validity of this assumption in CSP research is in doubt. The problem of non-normality of data in a *NP*-complete problem is observed and reported in [Mitchell 94]. In that research, the Davis-Putnum algorithm was applied to test the satisfiability of randomly generated 3-SAT problems. The number of steps that the algorithm took in each run was recorded. Sample means for different sample sizes were then calculated. Mitchell found that the sample means varied significantly as the sample size was increased. This suggested that the central limit theorem did not hold unless a very large sample size was used. As a result, the normality assumption was undermined.

According to our previous computational experience, the number of compatibility checks or the cpu time needed for a CSP algorithm-heuristic combination to solve CSPs may have a great variation even for CSPs which share very similar characteristics.[1] We conjecture that the normality assumption could be violated in some occasions. If this is true, the implication is twofold. Firstly, previous reported research results based on parametric statistics could be prone to error. Secondly, the CSP research community has to consider some statistical tools which do not rely on normality assumption for analysing experimental results.

The aim of this paper is to provide evidence that the normality of data should not be taken for granted when evaluating CSP methods. To make this paper more self-contained, our procedures to verify normality of data will be included in the next section.[2] The design and results of experiments on random binary CSPs and 3-colouring problems, two popular problems for benchmarking CSP methods, will be described in §3 and §4 respectively. We shall conclude the lesson that we learn in the final section.

2 Procedures for verifying normality of data

The deviation of data from normality can be measured in terms of *skewness* and *kurtosis*. Skewness is a measure of the central tendency of the peak of the data whereas kurtosis is a measure of the relative peakedness of the curve of the data. We denote skewness and

---

[1] For simplicity, we only consider the number of compatibility checks as the performance indicator for comparing algorithm-heuristic combinations in this paper. However, we believe that the observations presented in this paper should be equally valid if cpu time is used as a performance indicator instead.

[2] There are two reasons for us to include the details of this "standard" statistics rather than only making a reference in this paper. First, it appears that the procedures of testing normality of data are not widely available in books on statistics. Second, we found different descriptions on the computation of those statistics in different sources. In order to avoid possible confusion and make this paper more self-contained, we decided to spell out the details of the procedures that we used in the paper. However it is important to emphasise that most of the materials presented in this section are adopted from [Snedecor 56].

kurtosis by g_1 and g_2, statistics based on the sum of the third and the fourth powers of deviations from mean respectively.

Suppose X_i is the value of i-th sample of X and $\overline{X}$ is the sample mean. S_j denotes the sum of the j-th powers of deviations from the sample mean, *i.e.*,

$$S_j = \sum_{i=1}^{N} \left(X_i - \overline{X}\right)^j$$

where N is the sample size. g_1 and g_2 are defined as follows:

$$g_1 = k_3 / (k_2 \sqrt{k_2}) \qquad \text{and} \qquad g_2 = k_4 / k_2^{\,2}$$

where

$$k_2 = S_2/(N-1);$$
$$k_3 = N S_3/(N-1)(N-2);$$
$$k_4 = \left[N(N+1)S_4 - 3(N-1)S_2^{\,2}\right]/(N-1)(N-2)(N-3);$$

If g_1 and g_2 are equal to zero, there is no departure from normality as far as these measures are concerned. g_1 is positive when data are skewed positive and negative when data are negatively skewed.[3] g_2 is positive when the curve is more peaked than normal, *i.e.*, there is an excess of items near the mean, and negative when the distribution curve is flat-topped. Z scores of skewness and kurtosis, denoted by Z_{g1} and Z_{g2} respectively are defined as follows:

$$Z_{g1} = g_1 / \sqrt{6N(N-1)/(N-2)(N+1)(N+3)}$$
$$Z_{g2} = g_2 / \sqrt{24N(N-1)^2/(N-3)(N-2)(N+3)(N+5)}$$

Thus if Z_{g1} or Z_{g2} is greater than 1.96 or less than -1.96, the distribution is non-normal with $p \leq 0.5$.

3 Experiments on random binary CSPs

3.1 Experimental design

In order to investigating whether the normality assumption holds in general, we decided to conduct our experiments on randomly generated binary CSPs which have been widely used as benchmark problems for evaluating CSP techniques.

[3] Note that a positively skewed distribution implies that the median is smaller than the mean and a negatively skewed distribution has its mean smaller than the median.

The binary CSPs are characterised by a 4-tuple <n, m, $p1$, $p2$> such that n is the number of variables, m is the uniform domain size for each variable, constraint density, denoted by $p1$, is the proportion of pairs of variable that a constraint exists between them and constraint tightness, denoted by $p2$, is the proportion of incompatible pairs of values for each pair of variables if there exists a constraint between them. No universal constraints (that allows any pair of values) or null constraints (that do not allow any pair of values) were allowed. At each data point, 50 samples were generated. All the generated problems have connected constraint graphs. Since we are interested in how often that the normality assumption is violated, CSPs with different characteristics were used for testing.

Given a set of CSPs, different algorithm-heuristic combinations might produce results with very different distributions. To counter this, several algorithm-heuristic combinations were used on each problem to find the first solution or exhaust the search space otherwise. These included backmarking and conflict-directed backjumping (BM-CBJ) [Prosser 93] with minimal width heuristic (MWO) [Freuder 82], forward checking with backmarking and conflict-directed backjumping (FC-BM-CBJ) [Prosser 94] with a heuristic called BZ which is derived from Brélaz heuristic [Brélaz 79][4], minimal forward checking with backmarking and conflict-directed backjumping (MFC-BM-CBJ) [Kwan & Tsang 95] with either BZ or MWO. The backmarking element in all these algorithms employed a 2-dimensional data structure for storing minimal backup level as described in [Kondrak & van Beek 95]. These algorithms were chosen because they represent a spectrum of CSP techniques which perform different degree of constraint propagation during search. The purpose of employing backmarking and conflict-directed backjumping in all the chosen algorithms was to reduce the possibility of thrashing during search. We hoped that this allowed us to alleviate the possibility of encountering exceptionally hard problems [Hogg & Williams 94] [Gent & Walsh 94] which usually result in producing non-normal distributions, particularly when a small sample size is adopted.

3.2 Results

In the first series of experiments, we ran the selected algorithm-heuristic combinations on CSPs with the following characteristics.

$n = 10$; $m = 5$; $p1 = 0.2, 0.3, ..., 1$;
$n = 10$; $m = 20$; $p1 = 0.2, 0.3, ..., 1$;
$n = 20$; $m = 10$; $p1 = 0.1, 0.2, ..., 1$;
$n = 30$; $m = 5$; $p1 = 0.1, 0.2, ..., 1$;
$n = 40$; $m = 5$; $p1 = 0.1, 0.2, ..., 1$

$n = 10$; $m = 10$; $p1 = 0.2, 0.3, ..., 1$
$n = 20$; $m = 5$; $p1 = 0.1, 0.2, ..., 1$
$n = 20$; $m = 20$; $p1 = 0.1, 0.2, ..., 1$
$n = 30$; $m = 10$; $p1 = 0.1, 0.2, ..., 1$

For all the above settings, $p2$ was varied from 0.04 to 0.96 with an increment of 0.04. For each of the data points, we computed Z_{g1} and Z_{g2} of the distribution produced by each of the algorithm-heuristic combinations on the 50 samples at that data point. The number of

[4] BZ is a variant of the fail-first heuristic [Haralick & Elliott 80]. The difference between them is that BZ breaks ties by selecting the future variable connected to the most other future variables whereas the fail-first heuristic selects one arbitrarily.

distributions which were found non-normal for the chosen algorithm-heuristic combinations for different n and m is given in Table 1. Note that we regard a distribution as non-normal if either its skewness or its kurtosis is beyond normality with a confident level of 95%.

Each entry of Table 1 has two numbers. The first is the number of non-normal distributions found whereas the second is the number of total data points examined. For example, the top leftmost entry in the table means that when FC-BM-CBJ with BZ is used to tackle the CSPs, which have 10-variable and a uniform domain size of 5, from 216 data points, 107 of them have a non-normal distribution.

From Table 1, it is clear that FC-BM-CBJ with BZ quite often produces results with a non-normal distribution. The other three algorithm-heuristic combinations produce significantly more non-normal results for problems with $m = 5$ than with $m = 10$ or 20. It appears that the normality assumption is often violated when m is small, say $m = 5$, in all the selected combinations.

	FC-BM-CBJ + BZ[5]	MFC-BM-CBJ + BZ	MFC-BM-CBJ + MWO	BM-CBJ + MWO
$n = 10; m = 5$	107 (216)	117 (216)	122 (216)	108 (216)
$n = 10; m = 10$	100 (216)	11 (216)	16 (216)	14 (216)
$n = 10; m = 20$	95 (216)	10 (216)	14 (216)	12 (216)
$n = 20; m = 5$	104 (240)	121 (240)	122 (240)	122 (240)
$n = 20; m = 10$	91 (240)	27 (240)	23 (240)	26 (240)
$n = 20; m = 20$	80 (240)	23 (240)	29 (240)	27 (240)
$n = 30; m = 5$	102 (240)	112 (240)	113 (240)	111 (240)
$n = 30; m = 10$	79 (240)	24 (240)	26 (240)	28 (240)
$n = 40; m = 5$	90 (240)	108 (240)	125 (240)	119 (240)

Table 1. Number of data points which have a non-normal distribution ($p < 0.05$) for random binary CSPs (sample size at each data point = 50)

When we examined the raw results, two trends were observed. First, almost all non-normal distributions that we found are heavily positively skewed, with many of them also associated with a large kurtosis. Second, all the adopted combinations produced non-normal distributions across the "mushy region"[6] [Smith 94] when $p1$ was set to 0.1.[7] In particular, most of the distributions associated with the data points across the "mushy region" produced by FC-BM-CBJ with BZ for small $p1$ (0.1 - 0.3) were found to be non-normal.

The second observation is particularly interesting. CSPs near to the 50% satisfiable point, also known as the crossover point [Crawford & Auton 93], have been used in a

[5] The plus sign "+" stands for "combined with". For example "FC-BM-CBJ+BZ" means FC-BM-CBJ combined with BZ.

[6] The "mushy region" denotes the range of values of the order parameter over which the phase transition from solubility to insolubility takes place [Smith 94]. In our case, $p2$ was the order parameter since n, m and $p1$ were fixed in the experiments.

[7] Note that this excludes the 10-variable CSPs as the smallest $p1$ which allows a connected constraint graph is 0.2.

number of recent research in evaluating CSP methods. In particular, many of those benchmark problems have a rather small $p1$. However, our finding appears to suggest that when m or $p1$ is small, the normality assumption is very often violated near the crossover point. We conjectured that when both m and $p1$ have small values, the chance of having the normality assumption valid would be very slim for problems close to the crossover point. In order to verify this conjecture, the same algorithm-heuristic combinations were run on larger CSPs with $n = 60$, $m = 5$ and $p1 = 0.1$, 0.2 and 0.3. Again $p2$ was varied from 0.04 to 0.96 with a step of 0.04 and 50 samples were generated at each data point. For each chosen $p1$, we recorded the largest range of $p2$ in which the distribution associated with each of our sampled $p2$ within the range is non-normal. The results are summarised in Table 2. For ease of reference, the crude boundaries of the mushy regions are also given. Note that the mushy regions are often very narrow for the 60-variable CSPs. Phase transition [Cheeseman *et. al.* 91] happened "instantaneously" in several occasions because a coarse $p2$ increment was used. To overcome this problem, we define the $p2$'s range associated with a "crude" mushy region starting from the largest $p2$ which has all its 50 CSPs soluble and ending at the smallest $p2$ which has all its 50 CSPs insoluble in our samples.

		FC-BM-CBJ + BZ	MFC-BM-CBJ + BZ	MFC-BM-CBJ + MWO	BM-CBJ + MWO
$p1 = 0.1$	"crude" mushy region	0.32 - 0.40	0.32 - 0.40	0.32 - 0.40	0.32 - 0.40
	non-normal $p2$'s range	0.20 - 0.52	0.20 - 0.68	0.20 - 0.84	0.16 - 0.72
$p1 = 0.2$	"crude" mushy region	0.20 - 0.24	0.20 - 0.24	0.20 -0.24	0.20 -0.24
	non-normal $p2$'s range	0.16 - 0.32	0.04 - 0.32	0.12 - 0.36	0.08 - 0.76
$p1 = 0.3$	"crude" mushy region	0.12 - 0.20	0.12 - 0.20	0.12 - 0.20	0.12 - 0.20
	non-normal $p2$'s range	0.12 - 0.24	0.08 - 0.20	0.08 - 0.36	0.08 - 0.52

Table 2. Range of $p2$ with a non-normal distribution ($p < 0.05$) for 60-variable random binary CSPs (sample size at each data point = 50)

It is clear from Table 2 that all the data points within the "crude" mushy regions of the 60-variable CSPs have a non-normal distribution when $p1 = 0.1$, 0.2 and 0.3. It can also be observed from the table that the lower the $p1$, the greater the $p2$'s range associated with a non-normal distribution. These results suggest that we have to be careful in interpreting the inference drawn from parametric statistics based on the results of running a CSP method on problems close to the crossover point which have small m and $p1$.

4 Experiments on 3-colouring problems

4.1 Experimental design

It has been shown in §3.2 that we should not assume normality of distribution when analysing results based on experimenting random binary CSPs which have small m and $p1$ within or close to mushy regions. In order to see whether the same observation can be

drawn from other types of CSPs, experiments on 3-colouring problems were performed. The reason for choosing 3-colouring problems was that they have a small domain size. In addition, the approximate location of phase transition for 3-colouring problems has been reported to occur when the average node degree of a graph, denoted by d, is about 4.6 in the literature [Hogg & Williams 94]. For graphs with reasonably large number of nodes, denoted by n, it is expected that phase transition happens at a rather small $p1$ because the relationship between d and $p1$ are dictated by the following equation:

$$d = (n\text{-}1) * p1$$

In our experiments, connected graphs of different sizes (n = 20, 30, ..., 110 and 200) were randomly generated. For each value of n, 100 graphs were generated for each value of d starting from 2.0 to 7.0 with an increment of 0.1. 56,100 graphs were used in total. The four algorithm-heuristic combinations mentioned in §3.1 were used to colour each graph except that MFC-BM-CBJ with BZ was not used for colouring the 200-node graphs as this combination was found much less efficient than the other combinations when colouring the other graphs ($n \leq 100$).

4.2 Results

The empirical results were analysed as what we did for the random binary CSPs. Table 3 displays the number of data points which are associated with a non-normal distribution ($p < 0.05$). The interpretation of Table 3 is the same as Table 1.

	FC-BM-CBJ + BZ	MFC-BM-CBJ + BZ	MFC-BM-CBJ + MWO	BM-CBJ + MWO
n = 20	44 (51)	43 (51)	44 (51)	43 (51)
n = 30	47 (51)	45 (51)	43 (51)	43 (51)
n = 40	44 (51)	46 (51)	39 (51)	39 (51)
n = 50	43 (51)	48 (51)	39 (51)	38 (51)
n = 60	43 (51)	48 (51)	40 (51)	40 (51)
n = 70	43 (51)	48 (51)	40 (51)	39 (51)
n = 80	45 (51)	48 (51)	40 (51)	40 (51)
n = 90	44 (51)	48 (51)	40 (51)	38 (51)
n = 100	44 (51)	49 (51)	40 (51)	39 (51)
n = 110	46(51)	48 (51)	44 (51)	44 (51)
n = 200	45(51)	N/A	38 (51)	38 (51)

Table 3. Number of data points which have a non-normal distribution ($p < 0.05$) for 3-colouring problems (sample size at each data point = 100)

From Table 3, it is clear that for all the three algorithm-heuristic combinations that we used, most of the results (on average, over 80%) produce a non-normal distribution. When examining the raw results, we found that almost all the non-normal distributions have a large skewness as well as a large kurtosis. In addition, all the data points across the "mushy region", no matter which algorithm-heuristic combination was used, are associated

with a non-normal distribution (see Table 4). These findings are consistent with the results we have found earlier when experimenting random binary CSPs. Note that Table 4 is interpreted in a similar way as in Table 2. The main differences are that mushy regions are used instead of "crude" mushy regions and the non-normal *d*'s range is identified instead of the *p*2's range.

		FC-BM-CBJ + BZ	MFC-BM-CBJ + BZ	MFC-BM-CBJ + MWO	BM-CBJ + MWO
n = 20	mushy region	2.8 - 5.8	2.8 - 5.8	2.8 - 5.8	2.8 - 5.8
	non-normal *d*'s range	2.7 - 7.0	2.7 - 6.8	2.8 - 7.0	2.9 - 7.0
n = 30	mushy region	3.1 - 5.0	3.1 - 5.0	3.1 - 5.0	3.1 - 5.0
	non-normal *d*'s range	2.4 - 7.0	2.6 - 7.0	3.0 - 7.0	3.0 - 7.0
n = 40	mushy region	3.2 - 5.1	3.2 - 5.1	3.2 - 5.1	3.2 - 5.1
	non-normal *d*'s range	2.7 - 7.0	2.5 - 7.0	3.2 - 7.0	3.2 - 7.0
n = 50	mushy region	3.4 - 5.1	3.4 - 5.1	3.4 - 5.1	3.4 - 5.1
	non-normal *d*'s range	2.9 - 7.0	2.3 - 7.0	3.2 - 7.0	3.3 - 7.0
n = 60	mushy region	3.8 - 4.9	3.8 - 4.9	3.8 - 4.9	3.8 - 4.9
	non-normal *d*'s range	3.1 - 7.0	2.3 - 7.0	2.8 - 7.0	3.1 - 7.0
n = 70	mushy region	3.2 - 5.1	3.2 - 5.1	3.2 - 5.1	3.2 - 5.1
	non-normal *d*'s range	2.8 - 7.0	2.3 - 7.0	3.2 - 7.0	3.3 - 7.0
n = 80	mushy region	3.1 - 4.9	3.1 - 4.9	3.1 - 4.9	3.1 - 4.9
	non-normal *d*'s range	2.7 - 7.0	2.3 - 7.0	3.1 - 7.0	3.1 - 7.0
n = 90	mushy region	4.0 - 4.9	4.0 - 4.9	4.0 - 4.9	4.0 - 4.9
	non-normal *d*'s range	3.1 - 7.0	2.3 - 7.0	3.2 - 7.0	3.4 - 7.0
n = 100	mushy region	3.8 - 4.8	3.8 - 4.8	3.8 - 4.8	3.8 - 4.8
	non-normal *d*'s range	2.7 - 7.0	2.3 - 7.0	3.1 - 7.0	3.4 - 7.0
n = 110	mushy region	3.1 - 5.0	3.1 - 5.0	3.1 - 5.0	3.1 - 5.0
	non-normal *d*'s range	2.6 - 7.0	2.3 - 7.0	2.9 - 7.0	2.9 - 7.0
n = 200	mushy region	4.2 -4.8	4.2 -4.8	4.2 -4.8	4.2 -4.8
	non-normal *d*'s range	2.7 - 7.0	N/A	3.4 - 7.0	3.4 - 7.0

Table 4. Range of d with a non-normal distribution ($p < 0.05$) for 3-colouring problems (sample size at each data point = 100)

Table 4 provides more evidence to our conjecture that the normality assumption should not be taken for granted when analysing CSP research results when the problems have low constraint density and domain size which are near to the crossover point.

5 Conclusions

The success of an empirical approach relies on a rigorous experimental design and statistical analysis of the results. Interpreting the results using an inappropriate statistical tool could lead to an inaccurate or even misleading conclusion. Many statistical tools require the fulfilment of certain assumptions, notably the normality assumption, for their application. We have provided evidence in this paper that the normality assumption should not be taken for granted in CSP research.

According to our research results based on random binary CSPs and 3-colouring problems, the normality assumption is often violated for a number of algorithm-heuristic combinations when applied to CSPs which have a low constraint density or small domain sizes within or close to mushy regions. This point deserves extra attention as many studies of CSP methods have used this type of problems, particularly those near to the crossover points, as benchmark problems. Although some parametric tests like *t*-test and ANOVA are quite robust, they could still result in making error-prone inference when normality assumption is seriously violated. One way to alleviate the problem is to adopt a more conservative level of confidence to avoid errors. However we believe that non-parametric statistics such as Wilcoxon's signed rank test are in general more appropriate for analysing CSP research results as they do not assume a normal distribution of data.

Acknowledgements

Thanks are given to James Borrett who helped to implement the Brélaz heuristic and commented on an early draft of this paper. We also thank Edward Tsang for suggesting improvements to this paper. This research was supported by the University of Essex research promotion fund ref. R7207 and by the EPSRC research grant ref. GR/J/42878.

References

[Brélaz 79]
Brélaz, D. 1979. New Methods to Color the Vertices of a Graph. *Communications of the ACM* 22(4), 251-256.

[Cheeseman *et. al.* 91]
Cheeseman, P., Kanefsky, B. & Taylor, W.M. 1991. Where the Really Hard Problems Are. In *Proceedings IJCAI-91*, 331-337.

[Crawford & Auton 93]
Crawford, J.M. & Auton, L.D. 1993. Experimental Results on the Crossover Point in Satisfiability Problems. In *Proceedings AAAI-93*, 21-27.

[Freuder 82]
Freuder, E. 1982. A Sufficient Condition for Backtrack-Free Search,. *J. ACM* 29, 24-32.

[Freuder & Wallace 91]
Freuder, E.C. & Wallace, R.J. 1991. Selective Relaxation for Constraint Satisfaction Problems. In *Proceedings 1991 IEEE International Conference on Tools for Artificial Intelligence*, 332-339.

[Frost & Dechter 95]
Frost, D. & Dechter, R. 1995. Look-ahead Value Ordering for Constraint Satisfaction Problems. In *Proceeding IJCAI-95*, 572-578.

[Gent & Walsh 94]
Gent, I.P. & Walsh, T. 1994. Easy Problems are Sometimes Hard. *Artificial Intelligence* 70, 335-345.

[Haralick & Elliott 80]
Haralick, R.M. & Elliott, G.L. 1980. Increasing Tree Search Efficiency for Constraint Satisfaction Problems. *Artificial Intelligence* 14, 263-314.

[Hogg & Williams 94]
Hogg, T. & Williams, C.P. 1994. The Hardest Constraint Problems: A Double Phase Transition. *Artificial Intelligence* 69, 359 - 377.

[Kondrak & van Beek 95]
Kondrak, G. & van Beek, P. 1995. *A Theoretical Evaluation of Selected Backtracking Algorithms*. In *Proceedings IJCAI-95*, 541-547.

[Kwan & Tsang 95]
Kwan, A.C.M. & Tsang, E.P.K. 1995. Minimal Forward Checking with Backmarking. Technical Report CSM-260, Department of Computer Science, University of Essex, U.K.

[Mitchell 94]
Mitchell, D.G. 1994. Respecting Your Data. In *Proceedings AAAI-94 Workshop on Experimental Evaluation of Reasoning and Search Methods*, 28-31.

[Prosser 94]
Prosser, P. 1994. Forward Checking with Backmarking. In *Proceedings ECAI-94 Workshop on Constraint Processing*, 59-68.

[Prosser 93]
Prosser, P. 1993. Hybrid Algorithms for the Constraint Satisfaction Problem. *Computational Intelligence* 9(3), 268-299.

[Smith 94]
Smith, B.M. 1994. Phase Transition and the Mushy Region in Constraint Satisfaction Problems. In *Proceedings ECAI-94*, 100-104.

[Snedecor 56]
Snedecor, G.W. 1956. *Statistical Methods* (5th edition). Iowa State University Press.

[Tsang 93]
Tsang, E. 1993. *Foundations of Constraint Satisfaction*. Academic Press, London.