Validating the Weights in Rule-Based Expert Systems: A Statistical Approach

DANIEL E. O'LEARY AND NILS A. KANDELIN
University of Southern California

ABSTRACT: Validating an expert system can be done by comparing decisions that an expert would make to those of the system. Unfortunately, that approach can be very costly and may be infeasible because there may be only a limited amount of an expert's time available to devote to validation. Accordingly, there is a need to develop other methods of validation that take advantage of the limited time of an expert. An alternative approach is to examine the knowledge base and characteristics of it to determine those parts that do not behave in a manner similar to the rest of the knowledge base or as expected. The limited resources of the validation process could then be concentrated on, e.g., validating those weights that stand apart from the other weights. In a rule-based expert system this means examining the rules and the weights on the rules for any exceptions. This article focuses on using statistical methodologies to determine those *individual* weights and those *sets* of weights that do not behave as would be expected. Outlier analysis, determining underlying probability distributions for the weights, statistical significance, bootstrap resampling, and other computer-intensive statistical methods are used in the analysis. A case study is used to demonstrate the methods developed throughout this article.

KEYWORDS: Validation, Expert Systems, Weights on Rules, Rule-Based Expert Systems

INTRODUCTION

A key issue in expert systems is the validation of expert systems. Recently, for example, frameworks have been provided for the validation of expert systems.[16,17] These frameworks are designed to guide the validation effort by

International Journal of Expert Systems Pp. 253–279. Volume 1, Number 3, 1988, ISSN 0894-9077.
Copyright ©1988 by JAI PRESS INC. All rights of reproduction in any form reserved.

establishing those issues that need to be considered and providing general methods that can be used in the validation of expert systems. This is particularly apparent,[16] where research methods are used to design the validation effort.

However, there has been limited research in developing methods for the actual process of validating expert systems. The purpose of this article is to provide one such approach that can be useful as *part* of the validation process for expert systems (ES).

Validating ES can be done by viewing the ES as a black box and determining the quality of the decisions by comparing the decisions to human decision makers or other models. Validating the ES may be taken a step further by opening up the black box and having the expert examine the knowledge base to determine why the ES made a particular decision.

However, such processes of validating ES can be expensive or infeasible. Experts' time is expensive. Further, experts' time for such validation efforts may be limited. A substantial time and effort investment would be required to test all aspects of, say, a 2500-rule system.

Accordingly, other approaches toward validating expert systems are desirable. Because part of the validation process is aimed at ascertaining that what the system "knows" is correct, part of the validation process can be aimed at ensuring that the knowledge base is correct.

There are multiple forms of knowledge representation[21] used in knowledge bases. A substantial portion of the ES built to date are rule-based, in part, for example because much of the expert system software (shells) is rule-based. Rule-based expert systems (RESs) are expert systems that have a knowledge base of "if...then..," rules. Often such systems have weights on the rules indicating, for example, the strength of belief in the rule.

Accordingly, a portion of the validation process is concerned with ensuring that the rules are consistent and complete and that the weights on the rules are correct. Methods have been presented to aid in analyzing the logical structure of the *rules*.[15] This article is concerned with developing methods to help guide the efforts of the validator in assessing the possibilities that a *weight* or weights may be incorrect.

Sources and Consequences of Incorrect Weights

Typically, the knowledge engineer solicits weights from the expert and has the weights put in the knowledge base of the system. Accordingly, the weights on the rules in an expert system may be incorrect for a number of reasons. First, the wrong weight may have been recorded by the knowledge engineer, or the wrong weight might have been keyed into the system by data entry personnel.

Second, the expert is likely to satisfice rather than optimize.[18] As a result, the quality of the estimates generated by the expert are just "good enough" to meet

the expert's time limitations and quality expectations. However, weights developed in this manner may not be satisfactory for an ES, since the model depends entirely on the rules and the set of parameters generated for the rules rather than on other knowledge available to the human. A weight that is "good enough" when treated as a weight on an individual rule may not be "good enough" when queries to the system lead to that weight being combined with other weights (that also are just "good enough") to determine the system's judgment. Cascading a number of weights on rules that are "good enough" can lead to a "*not* good enough" solution.

Third, the expert may find it difficult to express, for example, the strength of belief on a particular rule as a number. In those cases, we might anticipate that the weights are likely to be incorrect.

Fourth, the expert may not supply an accurate estimate for a weight because the expert does not want to be "replaced by a machine." Although the expert may be cooperative in an initial prototype, as the potential capabilities of the system are seen the expert may become less cooperative. If the expert supplies inappropriate weights, the system may yield decisions that are not as good as those of the expert.

In any case, if the weights on the rules are incorrect, then the system likely will find solutions to user queries that do not represent the best or even the better solutions. This can lead to a lack of confidence in and use of the particular system and expert systems in general. Accordingly, it is important to validate those weights to ensure that they are correct.

APPROACHES TO VALIDATING THE WEIGHTS

One approach to validation of the weights is to have the expert examine each of the weights to ensure that they are correct. Unfortunately, this approach may not be feasible because of time or resource constraints, or it may not be cost-beneficial for large systems. In addition, this approach may suffer from the same satisficing problem discussed above. Further, if the expert supplied incorrect rules in an effort to undermine the system, then there is little reason to assume that the behavior would change. Thus, it is important to develop alternative cost-beneficial validation methods.

If particular weights can be identified as possibly incorrect, then the validator can concentrate on those weights in the validation effort—i.e., concentrate on the exception. There are at least three approaches. First, the validator can choose a selected subset of weights for further detailed examination. This subset likely would consist of those weights that appear to stand apart from the rest of the weights—i.e., the exceptions. Second, the validator can examine characteristics of a set of the weights to determine if the weights' behavior is as the underlying theory—e.g., measurement theory and probability theory—would predict. Third, the validator can analyze the weights' behavior to deter-

mine if the weights behave as would be expected on the expert system development process—e.g., the weights may be expected to be from the same distribution in the initial prototype and in a second version of the system. If the weights do not behave as anticipated, then the entire set of weights in the second version of the system may warrant further examination. This article uses statistical techniques to test the relationships between the data and the mathematical theory on which the weights are based to identify those weights that stand apart from the rest of the weights and those sets of weights that deserve further investigation because of measurably different behavior in different versions of the system.

Plan of this Article

This article proceeds as follows. The next section provides a background discussion on the weights in a rule-based expert system. The following section develops a case study that is used throughout the article to demonstrate the various approaches that are developed to help validate the weights. Then reasons for the weights coming from a single distribution or discernible groups of distributions are analyzed. The next section uses analysis of outliers and Chebyschev's inequality to ascertain those weights that do not belong to the same distribution as the rest of the weights and thus may warrant further examination. The following section discusses estimating the type of distribution from which the weights derive. The next section uses that distributional information as the basis to estimate statistical significance of particular weights. The following section analyzes the distribution of the means of the weights to determine if the mean of the weights behaves as anticipated. For example, the measurement structure of the weights may require that the weights have a mean of zero. Then the next section investigates behavior of the correlational relationship between two types of weights used in a system. The following section discusses the relationship between the weights on the rules in successive versions of the systems development and implementation efforts. The next section discusses extension of the approaches in this article to other systems of weights and to other types of numbers that require validation. The last section summarizes the article.

WEIGHTS IN RULE-BASED EXPERT SYSTEMS

The weights in RESs have been implemented in two primary formats.[19] EMYCIN[1] uses "measures of belief" and "measures of disbelief" to develop "certainty factors" (CFs), and AL/X[2] uses positive weights (PWs) and negative weights (NWs). A number of RESs and expert system shells have been developed using these alternative approaches to representing uncertainty. Although this article discusses these two systems and later focuses on AL/X,

the general nature of the discussion does not rely on a particular format for the weights.

CFs provide one way to think about confirmation and quantification of degrees of belief. Given the rule "If E then H," the expert provides measures of belief (MB) and measures of disbelief (MD) in hypothesis ($0 =< MB, MD =< 1$), where MB and MD are formally as follows:

$$MB(H,E) = \begin{cases} 1 & \text{if } Pr(H) = 1 \\ \dfrac{\text{Max}[Pr(H|E), Pr(H)] - Pr(H)}{\text{Max}[1,0] - Pr(H)} & \text{otherwise} \end{cases} \quad (1)$$

$$MD(H,E) = \begin{cases} 1 & \text{if } PR(H) = 0 \\ \dfrac{\text{Min}[Pr(H|E), Pr(H)] - Pr(H)}{\text{Min}[1,0] - Pr(H)} & \text{otherwise} \end{cases} \quad (2)$$

An examination of equations (1) and (2) indicates that one of MB or MD is always zero.

The CF combines those two measures as $CF = MB - MD$. The assumption is that the numbers developed by experts are "adequate" approximations to the numbers that would be calculated if the requisite probabilities in equations (1) and (2) were known.

PWs and NWs provide an alternative method of generating weights on the rules. For the rule "If E then H," let

$$\text{PW} = \text{Log}\,(Pr(E'|H) / Pr(E'|H')) \quad (3)$$

$$\text{NW} = \text{Log}\,(Pr(E|H) / Pr(E|H')), \quad (4)$$

where E' and H' correspond to "not E" and "not H." Thus, in theory, each of these weights is the logarithm of a likelihood ratio.

The PWs are "necessity factors" because a small value for PW means that a high probability for E is necessary to produce a high probability of H. The NWs are "sufficiency factors" because a large value of NW means that a high probability for E is sufficient to produce a high probability of H.

The development of the weights is done by soliciting either the weights or the probabilities. If the weights are gathered, then generally the PWs and the NWs are scaled, as, for example, in the case in the next section, between -30 and 30. Alternatively, if the probabilities are gathered, then the probability of errors in the relationship ($Pr(E'|H)$ and $Pr(E|H')$) can be gathered. In either case, in order to generate equations (3) and (4) the expert must generate two pieces of information.

Table 1: Weights on Rules in AUDITOR—Initial Version*

Rule	Positive Weights	Negative Weights
1	-6.0	2.0
2	6.0	0.0
3	-6.0	0.0
4	-6.0	2.0
5	-3.0	1.5
6	3.0	0.0
7	3.0	0.0
8	2.0	-1.0
9	6.0	-1.0
10	-2.5	1.0
11	-2.0	1.0
12	-6.0	2.0
13	2.5	-1.0

Note: *The order of the weights is different than in Table 2.
Source: Dungan.[3]

CASE STUDY: AL/X

There are few complete RES sets of rules and corresponding weights available in the literature. However, Dungan[3] lists all the rules and weights for the expert system AUDITOR. Accordingly, the weights in AUDITOR are used to illustrate the validation of the weights in a rule-based system. Although the AL/X method of placing weights on the rules is used, this same validation approach can be used on weights developed for EMYCIN.

AUDITOR was written in AL/X. Development of AUDITOR apparently had two primary stages, First a set of thirteen rules and their corresponding weights was developed. Second, in an effort to improve performance, the original set of rules was expanded and changed to thirty-eight rules. The initial set of weights is in Table 1. The final set of weights is summarized in Table 2. Histograms of the final weights are given in Figure 1.

Figure 1: Histograms of Weights

```
Histogram of PW  N = 38

Midpoint  Count
  -30       1  *
  -25       0
  -20       0
  -15       0
  -10       0
   -5       4  ****
    0      20  ********************
    5      11  ***********
   10       2  **
```

```
Histogram of NW  N=38

Midpoint  Count
  -2.0      1
  -1.5      0
  -1.0      9  ********
  -0.5      4  ****
   0.0     15  ***************
   0.5      0
   1.0      1  *
   1.5      0
   2.0      2  **
   2.5      0
   3.0      5  *****
   3.5      0
   4.0      1  *
```

Table 2: Weights on Rules in AUDITOR—Final Version

Rule	Positive Weights	Negative Weights
1	-3.0	3.0
2	0.5	0.0
3	0.5	-0.5
4	3.0	-1.0
5	2.0	-1.0
6	-30.0	1.0
7	-2.0	4.0
8	-1.0	2.0
9	1.0	-0.5
10	3.0	0.0
11	2.0	0.0
12	-2.0	2.0
13	0.0	-0.5
14	5.0	-1.0
15	3.0	-1.0
16	1.5	0.0
17	4.5	0.0
18	1.0	-1.0
19	2.0	-0.5
20	3.0	-1.0
21	1.0	0.0
22	8.0	0.0
23	1.0	0.0
24	-3.0	3.0
25	2.0	0.0
26	2.0	0.0
27	7.0	-1.0
29	-3.0	3.0
30	-3.0	3.0
31	-3.0	3.0
32	5.0	0.0
33	3.0	0.0
34	6.0	-1.0
35	-2.0	0.0
36	5.0	0.0
37	2.0	-1.0
38	1.0	0.0

	N	Mean	Median	Std Dev
PW	38	0.868	1.750	5.987
PW (Rule #6 omitted)	37	1.703	2.000	3.108
NW	38	0.289	0.000	1.478
NW (Rule #6 omitted)	37	0.270	0.000	1.493

Source: Dungan.[3]

The PWs were developed from a composite of relative strength judgments of four auditors. The four sets of relative strengths were translated into one numerical estimate for each rule using an averaging approach.[3] The NWs were assigned by Dungan after the PWs were developed.

ASSUMPTION OF SINGLE OR DISCERNIBLE DISTRIBUTIONS

This article assumes that the weights come from a single distribution or that the weights come from multiple distributions, yet the weights can be associated with discernible distributions. Such an assumption is not unusual, and is necessary to employ statistical-based analysis. Once this assumption is made, then both distribution-free and particular-distribution assumptions can guide further investigations.

There are several reasons to assume that the weights come from a single discernibly or different distributions. First, if an expert develops the weights, then each of the weights comes from the *same source*. In the example, the NWs came from the same person. Second, the weights may have massaged or generated using the *same process*. In the example, the PWs were developed using the same averaging process. Third, if the knowledge base refers to knowledge about a *single problem*, then there is reason to assume that the weights come from a single distribution. In the case study, a single auditing problem was analyzed. Alternatively, the knowledge base may be separated into a number of loosely connected segments, where each segment is concerned with a different problem. In that case, the weights on the rules in each of the individual segments may be expected to derive from a different, yet discernible distribution. Fourth, *empirically* the weights may appear to be from a single distribution. Empirical evidence indicates that it appears that the PWs (with one weight removed) come from the same distribution. Fifth, the mathematical theory on which the weights are based may indicate that the weights are drawn from a *particular distribution*. That is the case with AL/X and the example, as discussed later in the article.

DETERMINATION OF OUTLIERS—NO DISTRIBUTION FORM SPECIFIED

The validator may not be able to specify the form of the particular distribution. In that case, one approach that can be used to ascertain the existence of those weights that stand apart from the rest of the weights is to find those weights that are outliers from the others. Thus, validation of the weights would then be concerned with determining whether any of the weights do not come from the same distribution as the other weights. If it appears that a weight does not come from the same distribution as the other weights, then that weight would warrant investigation.

Exploratory data analysis (EDA)[20] can be used for general output analysis. EDA is distribution-free. Rather than summarizing the data using the mean and standard deviation, for example, the median and heuristically based distances from the median are used to analyze the data. In addition to showing trends and patterns, EDA techniques "reveal surprising, unexpected or amusing features of the data that otherwise might go unnoticed."[20] Using current computer technology, these displays can be quickly generated. They can be repeated

under a variety of different assumptions to observe sensitivities or changes in the display. In addition, more traditional methods, such as regression analysis and Chebyschev's inequality, can be used to estimate which weights stand apart from the other weights.

Outliers can be identified from the data using either univariate or bivariate methods. We can investigate the PWs and the NWs either separately or as pairs.

Outliers—Univariate

The *stem-and-leaf display* is a univariate EDA tool that contains all the information of a histogram. It breaks the display of the data into groups (such as 0.0 to 4.9, 5.0 to 9.9, etc., depending on the scaling), like the histogram. The stem-and-leaf display uses three columns to summarize the data display. The first column displays a cumulative count of values in each of the groups as the groups approach the median, from both above and below the median. The line (or stem) that contains the median shows a count of values at the median and is denoted with parentheses. The second column of numbers holds the stem. This corresponds to the leftmost digit of the numbers in the data set. Stems will range from the leftmost digit of the smallest number in the set to the largest. The third column—the right-hand portion of the display—holds the leaves. Each leaf digit represents an individual value. The initial digits of that value are the stem digits. This is followed by the leaf digit. Thus, a stem of 4 and a leaf of 2 could represent the number 42. The position of the decimal point is indicated by the unit of the leaf digit printed at the top of the display.

The *boxplot* is a univariate EDA tool that focuses attention on extreme values by showing outliers and providing a way of determining those values that are a "measurable" distance from the median. The median is marked with a plus (+). The median is used to split the data into two sets of data, one on each side of the median. The middle half of each of those batches is referred to as a hinge. The difference between the two hinges is the *H*-spread. The hinges are used to form the vertical edges of the "box." Thus, the box that is formed (hinges on each side and the median in the middle) contains 50% of the data set. The remaining 50% of the data set is equally split into two groups of 25% that lie either below the lower hinge or above the upper hinge.

The hinges are then used to calculate the inner and outer fences, which are used for outliers identification. They are calculated as follows:

$$
\begin{aligned}
H\text{-spread} &= (\text{upper hinge} - \text{lower hinge}) \\
\text{Lower Inner Fence} &= (\text{lower hinge}) - 1.5 * (H\text{-spread}) \\
\text{Upper Inner Fence} &= (\text{upper hinge}) + 1.5 * (H\text{-spread}) \\
\text{Lower Outer Fence} &= (\text{lower hinge}) - 3 * (H\text{-spread}) \\
\text{Upper Outer Fence} &= (\text{upper hinge}) + 3 * (H\text{-spread})
\end{aligned}
$$

In a boxplot, dashed "whiskers" run from the hinges to the adjacent values on each side. Values between the inner and outer fences are "possible outliers" and are plotted with an asterisk (*). Values beyond the outer fences are "possible far outliers" and are plotted with a zero (0). Thus, boxplots provide both outliers and a means of assessing groups of values (e.g., in the whiskers) that may require further investigation.

Outliers—Bivariate

The *X–Y plot* is a bivariate EDA tool that displays the relationships between the X and Y variables. Unusual points, such as outliers, do not fit whatever pattern (e.g., a straight line) that might be present, making them more noticeable.

One pattern that is accessible from a statistical perspective is a straight line. Regression analysis can be used to fit a line to the data in the X–Y plot. Once regression analysis is used, then supporting tests of regression analysis can be analyzed. At least three different methods have been developed to aid in the identification of outliers in regression data: leverage matrix analysis, residuals analysis, or a combination of the two.

The **leverage values** (h_{ii}), which are the diagonal elements are of the hat matrix, can be used to find outlying X (independent variable) observations. The element h_{ii} can be obtained from the product of $(X'_i)\ (X'X)^{-1}\ (X_i)$, where X_i corresponds to the *ith*-observation and X is the data matrix. Leverage values exceeding $2p/n$ or $3p/n$ are considered outliers, where p is the number of variables. **Standardized residuals**, e_i ($(Y_{\text{actual}} - Y_{\text{predicted}})$ divided by the square root of *MSE*, where *MSE* is the mean squared error) also can be used to assist in the detection of Y (the dependent variable) outliers. There are several ways of identifying those Y values that may be outliers,[13] including direct examination of the standardized residuals for unusually large values.

Cook's distance, D_i[13] combines both leverages and standardized residuals to provide a measure of the impact of the *ith*-observation on the estimated regression coefficients. The D_i values are calculated using the formula $D_i = [(e_i^2) / (\text{p} * MSE)] * [h_{ii} / (1 - h_{ii})^2]$. While the D_i does not follow the F-distribution, researchers[13] have suggested that if a D_i value is at the 50 percentile level (or more) for $F_{(p/n-p)}$ distribution (where n is the number of observations), then the observation should be considered as having substantial influence on the regression (i.e., an outlier).

Case Study—Outliers

Analysis of the stem-and-leaf plots (Figure 2) shows that the −30.0 PW in rule 6 is the largest (absolute) value and the most "unusual" value of the entire rule set. No other values are as extreme. Both stem-and-leaf plots show evi-

Figure 2: Stem-and-Leaf Plots of Weights

```
Stem-and-leaf of PW

N=38
Leaf Unit = 1.0

1      -3 0
1      -2
1      -2
1      -1
1      -1
1      -0
10     -0 333322221
(21)    0 000111111222222333334
7       0 5556789
```

```
Stem-and-leaf of NW

N=38
Leaf Unit = 0.10

1    -2  0
1    -1
10   -1  000000000
14   -0  5555
(7)  -0  0000000
17    0  00000000
9     0
9     1  0
8     1
8     2  00
6     2
6     3  00000
1     3
1     4  0
```

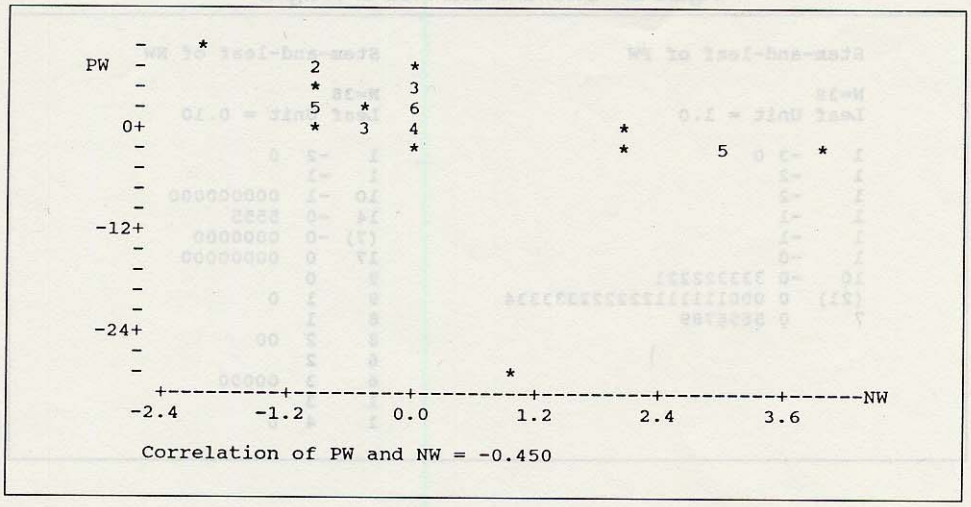
dence of some central tendency in the distributions of the positive and negative weights; however, this is a tentative conclusion because of the relatively small ($N = 38$) sample size.

The boxplot (Figure 3a) for the PWs substantiates the −30.0 value of rule 6 (noted previously) as a possible far outlier. The boxplot (Figure 3b) for the NWs indicates that the 4 value of rule 7 is a possible far outlier, and those rules with a value of 2 or 3 are possible outliers. Each of these weights warrant investigation.

Figure 3a: Boxplot for PWs
Figure 3b: Boxplot for NWs

```
Boxplot for PW's
                                               ------
           O                                ---I  + I-------
                                               ------
      --+---------+---------+---------+---------+---------+-------PW
     -32.0     -24.0     -16.0      -8.0       0.0       8.0

                                 Exhibit 3-a

Boxplot for NW's
                   ---------
         ----------I       +--------        *       *       O
                   ---------

      +---------+---------+---------+---------+---------+------NW
   -2.4      -1.2        0.0       1.2      2.4        3.6

                                 Exhibit 3-b

     Note:
           * = possible outliers
           O = possible far outliers
           + = median
```

Figure 4: **X–Y** Plot for PWs and NWs



The X–Y plot of the PW values against the NW values (Figure 4) provides further evidence of the strong influence of the −30.0 positive weight of rule 6.

In the development of AUDITOR, the PWs were developed by a team of experts. Then, given those weights, the NWs were selected so that the expert system performed correctly. Thus, there is reason to assume that the PWs can be treated as the independent variable. Accordingly, it is reasonable to use a bivariate outlier test. Further, the X–Y plot suggests a straight-line pattern. However, the treatment of the NWs as an independent variable is also included for illustrative purposes.

Leverages over $2p/n$ or $3p/n$ merit checking. The leverage for rule 6 when the PWs are treated as the independent variable greatly exceeds $3p/n$, indicating that this observation is a possible outlier. No other PW has a significant leverage. When NW is used as the independent variable, several leverages barely exceed $2p/n$. The leverages are summarized in Table 3.

The standardized residuals indicate that rule 6 still dominates even if NW is treated as the independent variable. As shown in Table 3, the standardized residuals are large in both cases, but especially so when NW is used as the independent variable. This is consistent with the results using leverages.

The results for the test of Cook's D also are summarized in Table 3. The test found that treating the PW as either the dependent or independent variable yields rule 6 as an outlier. With PW as the dependent variable for rule 6, Cook's D-statistic was significant at the .999 level. With PW as an independent variable for rule 6, Cook's D statistic was significant at about the .500 level.

This discussion indicates that observation 6 is definitely an outlier from the rest of the observations. If the observation is not deleted, then it will have a large influence and probably distort any further correlational or distribution

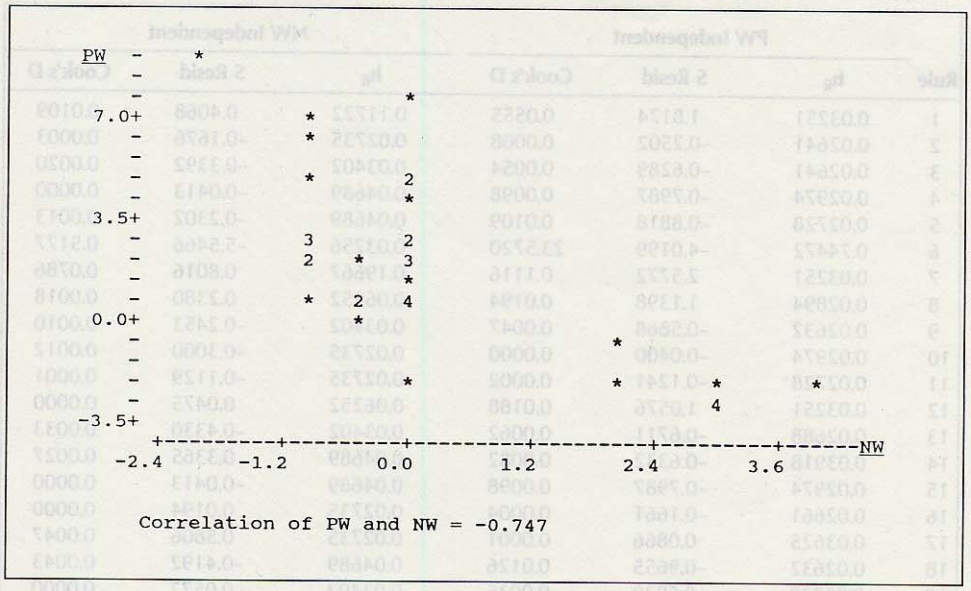
Table 3: Outlier Detection*

	PW Independent			NW Independent		
Rule	h_{ii}	**S Resid**	**Cook's D**	h_{ii}	**S Resid**	**Cook's D**
1	0.03251	1.8174	0.0555	0.11722	0.4068	0.0109
2	0.02641	-0.2502	0.0008	0.02735	-0.1676	0.0003
3	0.02641	-0.6289	0.0054	0.03402	-0.3392	0.0020
4	0.02974	-0.7987	0.0098	0.04689	-0.0413	0.0000
5	0.02728	-0.8818	0.0109	0.04689	-0.2302	0.0013
6	0.74472	-4.0199	23.5720	0.03256	-5.5466	0.5177
7	0.03251	2.5772	0.1116	0.19667	0.8016	0.0786
8	0.02894	1.1398	0.0194	0.06252	0.2380	0.0018
9	0.02632	-0.5868	0.0047	0.03402	-0.2453	0.0010
10	0.02974	-0.0400	0.0000	0.02735	-0.3000	0.0012
11	0.02728	-0.1241	0.0002	0.02735	-0.1129	0.0001
12	0.03251	1.0576	0.0188	0.06252	0.0475	0.0000
13	0.02688	-0.6711	0.0062	0.03402	-0.4330	0.0033
14	0.03918	-0.6332	0.0082	0.04689	0.3365	0.0027
15	0.02974	-0.7987	0.0098	0.04689	-0.0413	0.0000
16	0.02661	-0.1661	0.0004	0.02735	0.0194	0.0000
17	0.03625	0.0866	0.0001	0.02735	0.5806	0.0047
18	0.02632	-0.9655	0.0126	0.04689	-0.4192	0.0043
19	0.02728	-0.5029	0.0035	0.03402	-0.0577	0.0000
20	0.02974	-0.7987	0.0098	0.04689	-0.0413	0.0000
21	0.02632	-0.2081	0.0006	0.02735	-0.0740	0.0000
22	0.06466	0.3883	0.0052	0.02735	1.2352	0.0214
23	0.02632	-0.2081	0.0006	0.02735	-0.0740	0.0000
24	0.03759	1.7375	0.0590	0.11722	0.2104	0.0029
25	0.02728	-0.1241	0.0002	0.02735	0.1129	0.0001
26	0.02728	-0.1241	0.0002	0.02735	0.1129	0.0001
27	0.05466	-0.4677	0.0063	0.04689	0.7145	0.0125
28	0.07616	-1.0779	0.0479	0.09117	0.7660	0.0294
29	0.03759	1.7375	0.0590	0.11722	0.2104	0.0029
30	0.03759	1.7375	0.0590	0.11722	0.2104	0.0029
31	0.03759	1.7375	0.0590	0.11722	0.2104	0.0029
32	0.03918	0.1291	0.0003	0.02735	0.6741	0.0063
33	0.02974	-0.0400	0.0000	0.02735	0.3000	0.0012
34	0.04617	-0.5506	0.0073	0.04689	0.5255	0.0067
35	0.03251	-0.4619	0.0036	0.02735	-0.6352	0.0056
36	0.03918	0.1291	0.0003	0.02735	0.6741	0.0063
37	0.02728	-0.8818	0.0109	0.04689	-0.2302	0.0013
38	0.02632	-0.2081	0.0006	0.02735	-0.0740	0.0000

Notes: *$h_{ii} = h_{ii}$, S Resid = standardized residual.

analysis. Accordingly, for the remainder of the paper observation 6 will be deleted from the sample. The X–Y plot and the stem-and-leaf diagram of the revised data set are included as Figures 5 and 6.

Figure 5: X–Y Plot with Observation 6 Deleted



There is conflicting evidence on observation 7. Univariate tests indicate that it is a far outlier. However, this is not confirmed in bivariate analysis. Accordingly, observation 7 is not deleted, particularly because this analysis is done for illustrative purposes.

Figure 6: Stem-and-Leaf of PWs with Observation 6 Deleted

```
N  = 37
Leaf Unit = 0.10

 4    -3 0000
 8    -2 0000
 9    -1 0
 9    -0
12     0 055
18     1 000005
(6)    2 000000
13     3 00000
 8     4 5
 7     5 000
 4     6 0
 3     7 0
 2     8 0
 1     9 0
```

This section has presented some ways of determining those weights that appear to be outliers. It also presented ways of choosing weights for investigation that may deserve further consideration, rather than arbitrarily choosing the weights (e.g., choose those weights in the whiskers of the boxplots). Tests of this type likely are useful in ascertaining those weights in error, for example, rather than a "2" being used, a "20" is entered in the system.

Chebyschev's Inequality

Unfortunately, the analysis of outliers, discussed above, does not provide a quantitative estimate of the probability that a particular weight is not part of the same distribution as the rest of the weights. A bound on an estimate of this

probability can be obtained from Chebyschev's inequality. That bound can then be used to choose for examination the set of observations that meet certain requirements.

Chebyschev's inequality has been used in a number of applications[10] to isolate those members of a data set that may not be from the same distribution, without making a distribution assumption. The inequality can be used to compute an upper bound on the probability of a weight coming from the same distribution as the rest of the weights.

However, the cost of getting an estimate of a probability is that other information about the set of weights must be used. In particular, for Chebyschev's inequality we assume that the mean, u, and standard deviation, s, are known. Still, there is no distribution assumption. The inequality gives an upper bound on the probability of the variable being greater than a specified distance from the mean of the distribution. The limitation of Chebyschev's inequality is that it is very conservative, reducing the power of the estimate.

Formally,[10] if d is the distance of the random variable from the mean of the distribution, Chebyschev's inequality states that

$$Pr(|x-u|>=d) = < s^2 / d^2$$

It is difficult to compare outlier analysis and Chebyschev's inequality, in general, because both have different information requirements. However, the case study can provide a specific example that allows some comparison.

Case Study—Chebyschev's Inequality

In our example, for the NWs, $u = .289$ and $s = 1.478$. Consider observation 7 with a value of 4. Using Chebyschev's inequality (Alternatively, the mean and standard estimates could be made while withholding the weight to be tested),

$$Pr(|x-.289|>=3.711) = < 2.184/13.772 = .158$$

This indicates that the probability is less than 0.158 that the weight has a value of 4. As in the boxplot, this small probability is suggestive that the observation stands apart.

DISTRIBUTION ESTIMATES OF THE WEIGHTS

The mathematical theory on which the weights are based may indicate that the weights are drawn from a particular distribution (e.g., a normal distribution). For example, the weights generated for AL/X are logarithms of likelihood ratios. Under certain conditions the log of a likelihood ratio tends toward being a chi-square distribution.[9,11] Under the assumption that the *conditional proba-*

bilities are normally distributed, the likelihood ratios *are* chi-square distributions, with one degree of freedom.[11]

Alternatively, the behavior of experts in developing weights may suggest that weights are drawn from a particular distribution (e.g., normal) or that the underlying decision process draws from a particular distribution. For example, statisticians have found that the normal distribution is very robust for representing many underlying decision processes.

In addition, if a process is roughly normal, researchers have noted[8] that the logarithm transformation often reduces the violations from normality. Accordingly, for example, if the *ratios of the probabilities* used in calculating the PWs and NWs are roughly normal, then we could expect that the logarithm transform would reduce violations from normality.

Determining the type of distribution can aid the analysis in at least two ways. First, it allows us to use distribution information to determine the likelihood that particular observations are from the same distribution. This permits detail outlier determination and analysis. Second, it allows us to determine of the overall set of weights is behaving as we would anticipate—e.g., as a normal distribution. This permits us to determine if the weights have been generated appropriately.

Testing for a Normal Distribution

There are several methods to test the normality of the data. This article used two different tests: normal probability plots and Kolmogorov-Smirnov (K–S) tests.

Roughly speaking, the normal probability plots reflect the correlation between the sample z scores and the theoretical z scores that would come from a standard normal distribution of the same size as the sample. The z scores for the values are computed as the difference between the sample observation and the sample mean, divided by the sample standard deviation. The linearity of the normal probability plots and the correlation of the sample data with the normal scores measures the degree of normality. A table of critical values of the correlations was generated in Filliben.[7]

The K–S test is a goodness-of-fit test based on the relationship of the sample data to the theoretical distribution on a cumulative observation-by-observation basis. The test compares the expected cumulative frequencies assuming the normal distribution with the actual cumulative frequency from the data. If the difference between the probability of the expected and the actual cumulative frequencies exceeds a specified level, then the normality hypothesis is rejected. The specified levels are summarized in specialized tables.[6] The K–S test requires the standard deviation and mean of the distribution, either actual or estimated. Specific procedures for calculating the K–S test statistic D and tables of critical values are detailed in Ewart, Ford, and Lin.[6]

Testing for a Chi-Square Distribution

Unfortunately, there are not many tests that can be used to determine if a set of data comes from a chi-square distribution. The same methods used in examining normality can be used to test for a chi-square; however, tables of critical values are not readily available. Further research in this area might include generation of critical value tables via simulation for testing for the presence of a chi-square distribution in the same fashion as the normality plots or the K–S test.

Case Study—Distribution Estimates of the Weights

The normal probability plots for the PWs and NWs (Figures 7 and 8, respectively) suggest that the PWs are normal, with an observed correlation of 0.992 that is between the 75th percentile and 90th percentile of the null distribution. This leads us to conclude that there is no evidence to reject the null hypothesis of normality of the PWs.[7] However, the NWs are not normal. The correlation for the NWs is 0.934, which falls between the 0th percentile and 5th percentile of the null distribution. When the far outlier for the NWs (observation 7) is removed, the correlation becomes .932. Thus, even when the far outlier is removed, the test suggests that the NWs are not normal.

Estimates of the mean and standard deviation were computed from the data for each set of weights and used to calculate D for the K–S test. The largest D was 0.1207 for PWs and 0.2374 for NWs, which compares to a critical value

Figure 7: Normal Probability Plot of PWs with Observation 6 Deleted

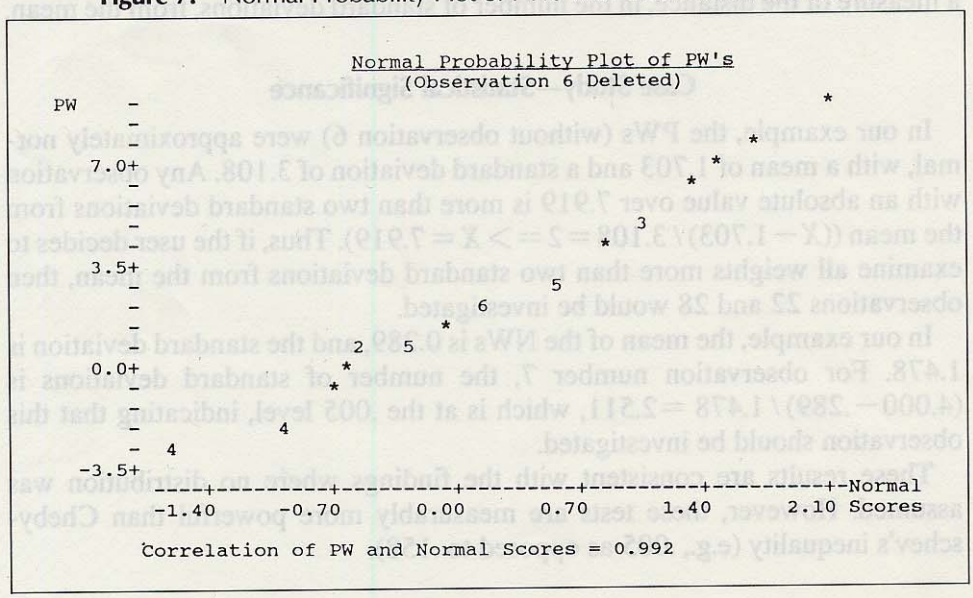
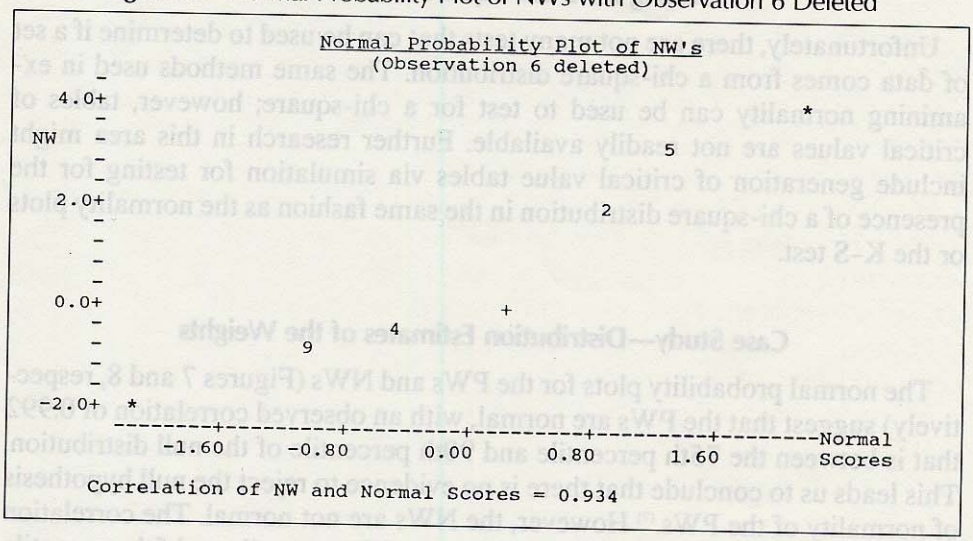


Figure 8: Normal Probability Plot of NWs with Observation 6 Deleted



(from the tables in [6]) of 0.1457. Therefore, the null hypothesis of normality is not rejected for the PWs but is rejected for the NWs.

STATISTICAL SIGNIFICANCE

If the distribution is normal, then normality tests of significance can be used to investigate the probability of a particular weight coming from the same distribution as the other weights. If we assume a normal distribution, then the deviation of a weight from the mean divided by the standard deviation provides a measure of the distance, in the number of standard deviations, from the mean.

Case Study—Statistical Significance

In our example, the PWs (without observation 6) were approximately normal, with a mean of 1.703 and a standard deviation of 3.108. Any observation with an absolute value over 7.919 is more than two standard deviations from the mean ($(X-1.703)/3.108=2 => X=7.919$). Thus, if the user decides to examine all weights more than two standard deviations from the mean, then observations 22 and 28 would be investigated.

In our example, the mean of the NWs is 0.289, and the standard deviation is 1.478. For observation number 7, the number of standard deviations is $(4.000-.289)/1.478=2.511$, which is at the .005 level, indicating that this observation should be investigated.

These results are consistent with the findings where no distribution was assumed. However, these tests are measurably more powerful than Chebyschev's inequality (e.g., .005 as opposed to .158).

DISTRIBUTION OF MEANS OF THE WEIGHTS

In some cases the weights of an expert system may have a mean that can be derived from a theory, due to measurement scaling. For example, the weight system may be constructed so that if the weights have been generated correctly, then the mean is 0 (that is *not* the case with either EMYCIN or AL/X). If that is the case, then the mean of a particular data set can be tested to determine if it meets those assumptions.

Bootstrap resampling[4,5,14] can be used to generate distributions of sample means from the original data. These distributions can then be used to make statements about the actual mean from the sample data.

Bootstrap resampling uses repetitive resampling from the original data to generate a set of the same number of elements as the original data. Then the mean of that sample is computed, and the mean is placed with the distribution of the means. Then another sample of data is generated with resampling from the original data, and the process is repeated. The process continues until an appropriate number of means have been generated. The resulting set of means provides a distribution of the means that can be used to develop confidence intervals about the mean.

Case Study—Distribution of Means of the Weights

The weights in AL/X should not necessarily have a particular mean. However, in order to demonstrate bootstrap resampling a distribution of sample means for both the PWs and NWs was generated. Two hundred samples were taken from the data set of each weight, and for every sample the mean was recorded.

The plot of the sample means from the bootstrap samples for the PWs (Figure 9a) appears symmetric, with a mean of 1.6875 and a standard deviation of 0.4861. The plot of the distribution of the sample means from the bootstrap samples for the NWs (Figure 9b) displays a symmetric distribution, with a mean of 0.3116 and a standard deviation of 0.2574. This bootstrap sample distribution can provide interesting information when compared to the actual sample means.

In general, we would expect that the bootstrap sample means would be relatively close to the actual sample means because of the sampling approach. For the NWs this appears true, as the actual sample means (0.289 including and 0.270 excluding rule 6, respectively) are close to the bootstrap distributions mean of 0.3116. Similarly, the actual sample mean of the PWs when rule 6 is excluded of 1.703 is close to the distribution of bootstrap samples mean of 1.6875. However, when rule 6 is included the PWs sample mean becomes 0.868, which demonstrates that the influence of the −30 positive weight of rule 6. This provides further evidence that the −30 weight is from a different distribution than the rest of the weights.

Figure 9a: Distribution of Sample Means for PWs

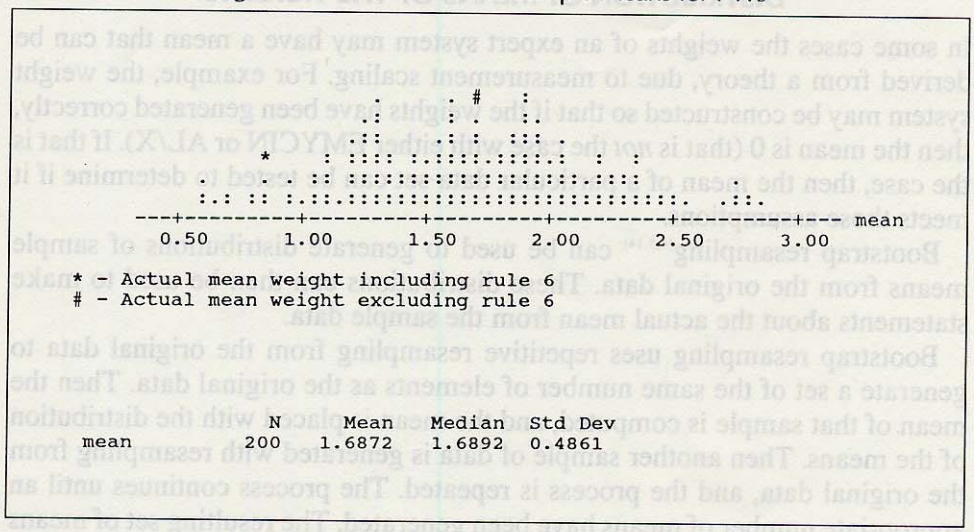
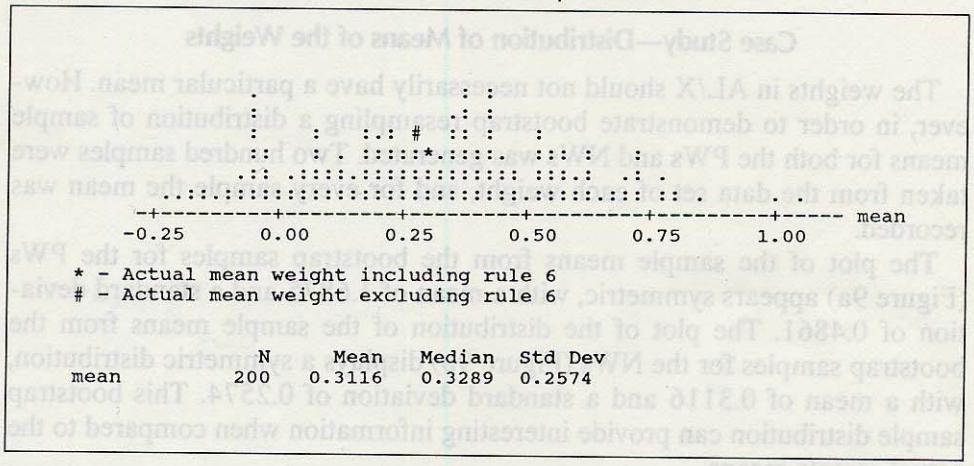


Figure 9b: Distribution of Sample Means for NWs



CORRELATION BETWEEN THE TWO SETS OF WEIGHTS

The sets of weights in an RES may be mathematically or theoretically related. If that is the case, then that relationship would manifest itself in the correlation between the actual values implemented in a system; otherwise, it is likely that the weights were not generated appropriately.

Simulation can be used to generate a distribution of correlations. That distribution can then be used to check the likelihood of the correlation of the actual data. If there is a small likelihood of the sample correlation, then the validator

likely would devote more resources to the validation of the weights than if the likelihood of the correlation was larger.

Distributions were developed for both AL/X weights and EMYCIN weights to demonstrate the similar behavior of the two systems of weights. These distributions illustrate the importance of accounting for the relationship between the two sets of equations—i.e., PWs and NWs or the MBs and MDs, respectively.

Distributions of Correlations for PWs and NWs

The simulation was developed by randomly choosing probabilities, P_1 and P_2, and using those probabilities to calculate PW and NW according to:

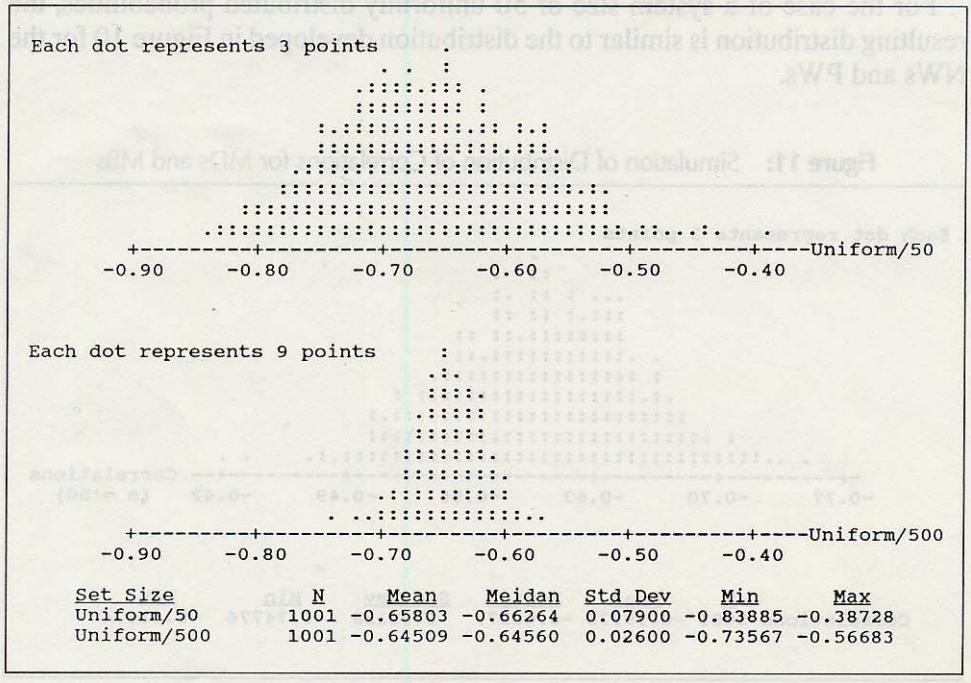
$$\mathrm{PW} = \mathrm{Log}\,\{P_1 / P_2\} \tag{5}$$

$$\mathrm{NW} = \mathrm{Log}\,\{(1 - P_1) / (1 - P_2)\} \tag{6}$$

Probabilities were chosen in pairs of size 50 and 500 from the uniform distribution. In each of the two cases, 1001 pairs were generated. These probabilities were then used to develop distributions of correlations of the PWs and NWs.

The findings of this simulation (see Figure 10) indicate that the correlation between the PWs and the NWs are always negative and on the average large.

Figure 10: Simulations of Distribution of Correlations for PWs and NWs with Varying Set Size



Set Size	N	Mean	Meidan	Std Dev	Min	Max
Uniform/50	1001	-0.65803	-0.66264	0.07900	-0.83885	-0.38728
Uniform/500	1001	-0.64509	-0.64560	0.02600	-0.73567	-0.56683

Accordingly, any set of weights generated for use by AL/X would be expected to have a large negative correlation.

In addition, the distribution of correlations changes as the number of pairs of weights per system (set) changes. In particular, the dispersion of the distribution of correlations decreases as the number of pairs increases from 50 to 500 elements per system.

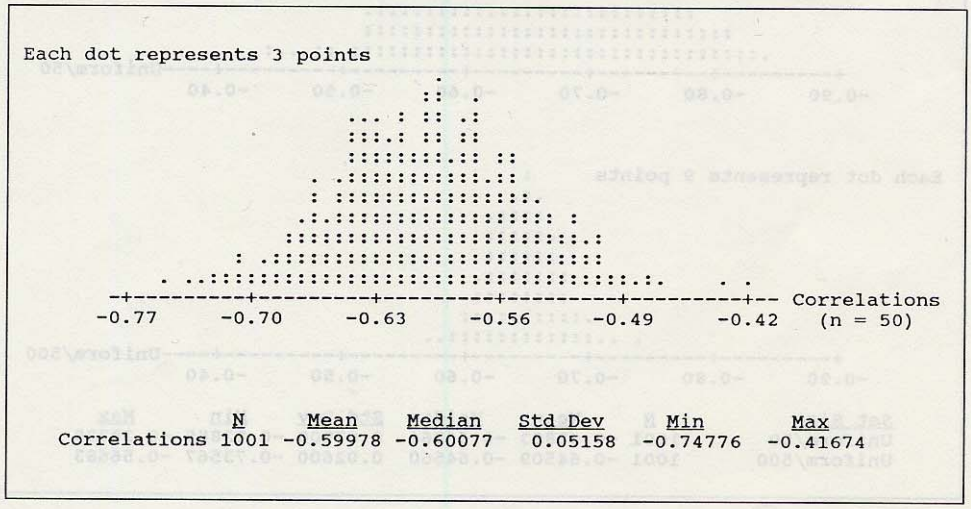
The simulated distribution can be used to develop an estimate of the probability of attaining a particular correlation from the system's set of weights. However, because the distribution changes as the number of elements in the set increases, a different distribution of correlations must be used, depending on the size of the system.

Distribution of Correlations for MBs and MDs

This same approach can be used to analyze other types of weights. In a manner similar to the approach used for the PWs and NWs, a distribution of correlations can be developed for the MBs and MDs. Because the approach is similar to that of the PWs and NWs, only one distribution was developed, and a slightly different approach was used for ease of computation. First, 1,000 pairs of probabilities, $Pr(H|E)$ and $Pr(H)$, were randomly generated from a uniform distribution. Second, 50 pairs of probabilities were randomly drawn from that set of 1,000. Third, those values were placed in (1) and (2), and the correlation between the MB and MD was computed. The second and third steps were done 1,001 times to develop the distribution in Figure 11.

For the case of a system size of 50 uniformly distributed probabilities, the resulting distribution is similar to the distribution developed in Figure 10 for the NWs and PWs.

Figure 11: Simulation of Distribution of Correlations for MDs and MBs



Case Study—Correlation Between the PWs and NWs

The results of the previous sections show that rule 6 is an outlier, and that there is a negative relationship between positive and negative weight values. The negative correlation of −0.450 is not readily apparent in the X–Y plot in Figure 4 (where rule 6 is included). However, when rule 6 is omitted, the correlation decreases to −0.747. The question is then whether or not this correlation is unusual.

The correlation between the PWs and the NWs was simulated using sets of 37 pairs of uniform distribution-based weights. Assume that the simulated correlations are arranged left to right, with increasing values (1,...,1,000). The correlation of −0.747 lies between observations 199 and 200. The correlation of −0.450, however, lies between observations 989 and 990. Thus the −0.747 does not appear to be an unusual correlation when compared to a distribution of likely (or possible) correlations between PWs and NWs. The correlation of −0.450 (resulting from the inclusion of rule 6) does appear to be highly unusual.

RELATIONSHIP BETWEEN PRELIMINARY ESTIMATES OF WEIGHTS AND FINAL VERSION OF THE WEIGHTS

The development of an expert system may require multiple revisions of the weights. For example, in the development of AUDITOR,[3] a major revision of the first set of weights was done in order to get the final set of weights. The relationship between the weights in different versions of the system can be investigated statistically.

There are several reasons to expect that the weights in the original version are or are not drawn from the same distribution as weights in the revised version. If one version is substantially different from the other version (e.g., with a number of rules or major changes in the understanding of the problem), then it is likely that the weights in the original version are not drawn from the same distribution as those in the revised version. However, if the versions of the system are very similar, we might expect that the weights are drawn from the same distribution.

In addition, if the expert sees that the initial prototype can perform certain tasks in an appropriate manner, then the expert may not be as "cooperative" in later versions of the process. This would likely yield a set of weights that comes from a different distribution. Thus it is important to be able to monitor the relationship between successive sets of weights.

Further, for systems with two sets of interlocking weights, such as AL/X and EMYCIN, we most likely would expect that if one of the sets of weights (e.g., PWs) is drawn from the same (different) distribution in both versions of the system, then the other sets of weights (e.g., NWs) are also drawn from the same

(different) distribution in both versions. In either case, because of the interlocking nature of equations (1) and (2) or (3) and (4), we would not expect to find that one of the types of weights (e.g., PWs) would be drawn from the same distribution for each version, while the other type of weights (e.g., NWs) was not.

Ascertaining the relationship between the different versions of the system can be accomplished by testing the difference in the means of the weights of the different versions to see if the difference is significant. Randomization tests[14] can be used to test the null hypothesis that the weights are independent of the particular version of the system. The experiment would be designed so that the n_1 weights of the first version and the n_2 weights of the second version were placed in one list. The list would then be shuffled, and the mean of the n_1 and the n_2 weights would be computed. Then the absolute value of the difference between those weights would be computed. This process would be performed a number of times. Then the percentage of time that the original difference was exceeded could be computed and would be used to estimate the probability of the hypothesis.

Case Study—Relationship Between Preliminary Estimates of Weights and Final Version of the Weights

In the case study, the number of rules was increased from 13 to 38. In addition, virtually all the weights for the original 13 rules were changed. As a result, there is no reason to expect that the weights for the first version of the system would come from the same distribution as the second version of the system. In order to test this, we test the null hypothesis that the absolute value of the difference in the means between the first and second version is not significant.

The results of the randomization for absolute mean differences for the PWs, based on 200 shuffles, are contained in Figure 12. The actual absolute value of

Figure 12: Absolute Differences Between Means—PWs

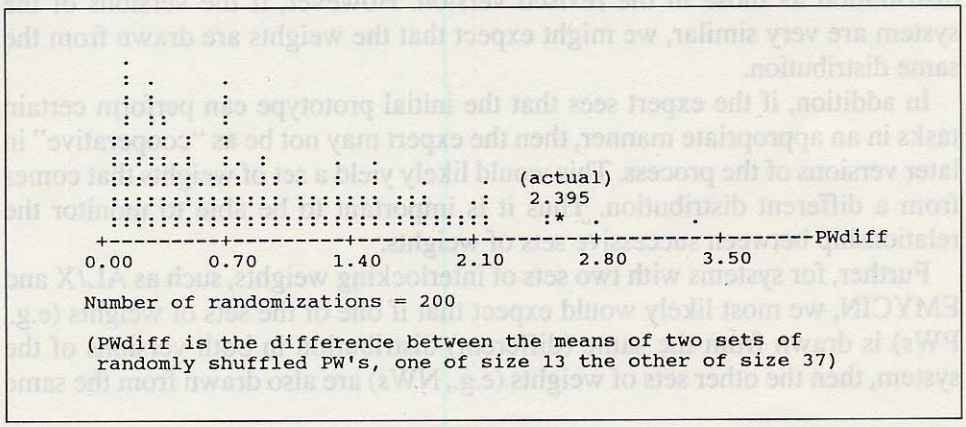
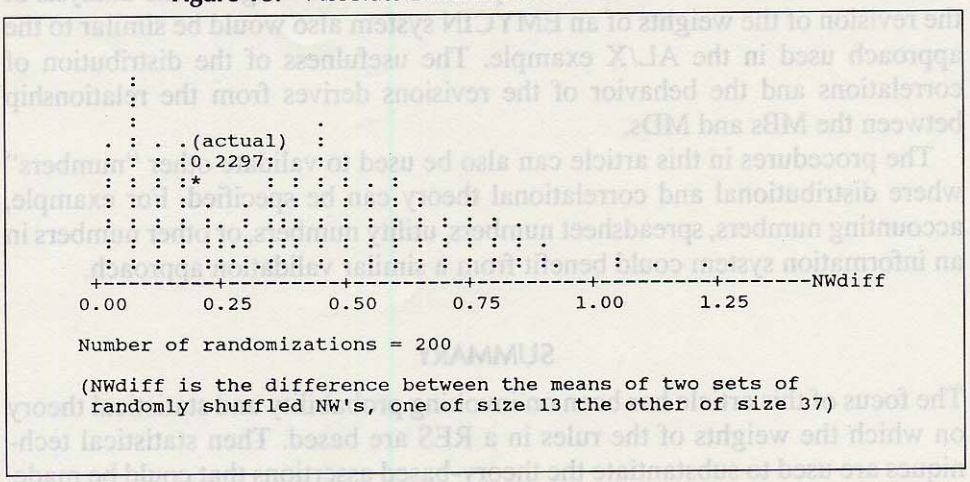


Figure 13: Absolute Differences Between Means—NWs



the difference in mean between versions was 2.395 for PWs. This was at the 2% level (between the ordered observations 196 and 197) and the generated distribution. This is highly unlikely; thus we conclude that the PWs are from different distributions.

The results of the randomization for absolute mean differences for the NWs, based on 200 shuffles, are contained in Figure 13. The actual absolute value of the difference in mean between versions was 0.2297 for NWs. This was at the 70% level (between the ordered observations 59 and 60) in the generated distribution. This test indicates that there is no reason to reject the null hypothesis that the two sets of NWs are not from different distributions.

These findings could result from the process by which the weights were assigned. The PWs were assigned a composite average of judgments of various experts (other than the developer), whereas the NWs were assigned exclusively by the developer.

EXTENSION TO EMYCIN AND OTHER SYSTEMS

The procedures developed in this article are not limited to the approach used in AL/X. The same outlier and distribution identification methods can be used in EMYCIN weights (either CFs or the nonzero MBs or MDs). The existence of a single expert or single process for the development of the weights could lead to the assumption of a single distribution. However, the mathematical structure does not provide a distribution of the MBs and MDs. The analysis of the distribution of the means is not likely to find use in analysis of EMYCIN, because there is no a priori basis to determine the expected mean. The correlation analysis of the MBs and MDs is exactly the same as that one with AL/X, as illustrated above. A simulated distribution can be used to provide estimates of

the likelihood of the correlation of a particular set of weights. The analysis of the revision of the weights of an EMYCIN system also would be similar to the approach used in the AL/X example. The usefulness of the distribution of correlations and the behavior of the revisions derives from the relationship between the MBs and MDs.

The procedures in this article can also be used to validate other "numbers" where distributional and correlational theory can be specified. For example, accounting numbers, spreadsheet numbers, utility numbers, or other numbers in an information system could benefit from a similar validation approach.

SUMMARY

The focus of this article has been on invoking probability and statistical theory on which the weights of the rules in a RES are based. Then statistical techniques are used to substantiate the theory-based assertions that could be made.

This article used some of the newer exploratory data techniques and computer-intensive techniques, in conjunction with more traditional statistical tools, such as Chebyschev's inequality and normal distribution confidence intervals, to analyze the assertions about membership in a distribution. In addition, the analysis took advantage of the distributional theory supporting the PWs and NWs.

The analysis in this article also took advantage of the relationship between the sets of weights in assessing the correlation. The analysis indicated that the impact of the process of revising the weights could be analyzed by determining if the weights are from the same distribution in successive revisions of the weights.

Techniques as developed here can provide relatively low-cost validation methods that could supplement or possibly replace some of the other, more costly tests for expert system validation that require the expert's involvement throughout.

The techniques in this article can be used to aid in accomplishing a number of validation goals for the weights. Outlier analysis and statistical significance can be used to assess if particular values are part of the same distribution as the other weights (e.g., all far outliers). Boxplots and statistical significance can also be useful in determining an objective measure of choosing other weights that should be examined (e.g., all those observations outside two standard deviations or two *H*-spreads). Determining the correlation and the position of the sample correlation in a simulated distribution of correlations can help evaluate the quality of the weight-generation process. Sets of weights with correlations in the fringes of the correlations distribution are likely to indicate either outliers or an inappropriate development of the weights. Finally, analyzing the relationship between successive sets of weights can help the validator determine if there has been a change in the process of weight generation. This

may prove particularly useful in identifying those cases where the expert provides incorrect estimates of the weights because of concern that a machine will take over the expert's job.

REFERENCES

(1) Buchanan, B., and E. Shortliffe. *Rule-Based Expert Systems* (Reading, MA: Addison Publishing Company, 1985).

(2) Duda, R., J. Gaschnig, and P. Hart. "Model Design in the Prospector Consultant System for Mineral Exploration." in *Expert Systems for the Micro-electronic Age*, edited by D. Mitchie (Edinburgh: Edinburgh University Press, 1979).

(3) Dungan, C. *A Model of Audit Judgement in the Form of an Expert System*, Ph.D. Dissertation, University of Illinois, 1983.

(4) Efron, B. "Bootstrap Methods: Another Look at the Jackknife." *Annals of Statistics* 7(1979): 1-26.

(5) Efron, B., and G. Gong. "A Leisurely Look at the Bookstrap, Jackknife and Cross-Validation." *The American Statistician* 37 (February 1983): 36-48.

(6) Ewart, P., J. Ford, and C. Y. Lin. *Applied Managerial Statistics* (Englewood Cliffs, NJ: Prentice-Hall, 1982).

(7) Filliben, J. "The Probability Plot Correlation Coefficient Test for Normality." *Technometrics* 17(1): 111-117.

(8) Foster, G. *Financial Statement Analysis* (Englewood Cliffs, NJ: Prentice-Hall, 1978).

(9) Graybill, F. *An Introduction to Linear Statistical Models* (New York: McGraw-Hill, 1961).

(10) Kaplan, R. S. *Advanced Management Accounting* (Englewood Cliffs, NJ: Prentice-Hall, 1982).

(11) Lindgren, B. W. *Statistical Theory*, third edition (New York: Macmillan, 1976).

(12) *Minitab Reference Manual* (State College, PA: Minitab, Inc., 1985).

(13) Neter, J., W. Wasserman, and M. Kutner. *Applied Linear Statistical Models*, second edition (Homewood, IL: Irwin, 1985).

(14) Nguyen, T., W. Perkins, T. Laffey, and D. Pecora. "Knowledge Base Verification." *AI Magazine* (Summer 1987): 69-75.

(15) Noreen, E. *An Introduction to Testing Hypotheses Using Computer-Intensive Methods*, Unpublished Manuscript, Graduate School of Business Administration, University of Washington, Seattle November, 1986.

(16) O'Leary, D. "Validation of Expert Systems." *Decision Sciences* (3): 468-486.

(17) O'Keefe, R., O. Balci, and E. Smith. "Validation of Expert System Performance." Unpublished Working Paper, Computer Science Department, Virginia Tech, TR 86-37.

(18) Simon, H. *The Sciences of the Artificial* (Cambridge, MA: The MIT Press, 1981).

(19) Spiegelhalter, D. "A Statistical View of Uncertainty in Expert Systems," in *Artificial Intelligence and Statistics*, edited by William Gale (Reading, MA: Addison-Wesley, 1986).

(20) Velleman, P., and D. Hoaglin. *Applications, Basics and Computing of Exploratory Data Analysis* (Belmont, CA: Duxbury Press, 1981).

(21) Winston, P. *Artificial Intelligence* (Reading, MA: Addison-Wesley, 1984).

Manuscript received October 1987; Revised February 1988; Accepted February 1988.

Address correspondence to Daniel E. O'Leary, Graduate School of Business, University of Southern California, Los Angeles, California 90089-1421.