C4 Understanding and Applying "*Daubert's* Error Rate" - Part Two

Mark I. Marpet, PhD, PE, 14 Cowie Road, Chester, NJ 07930-9715; Marcus P. Besser, PhD, Thomas Jefferson University, 130 South Ninth Street, Philadelphia, PA 19107-5233; and Howard P. Medoff, PhD, Pennsylvania State University, 1600 Woodland Road, Abington, PA 19001*

After attending this presentation, attendees will be in a better position to qualify or cross-examine an expert with respect to this area. Error Rate is used as an indicator of the reliability of a scientific- hypothesis, technical test, procedure, or process, and is often required for admissibility of expert testimony. Technical attendees will learn two forms that an error rate can take and how to develop uncertainty intervals, derived from the Cumulative Function and from a confidence interval derived from a sign test on the median, which can serve to characterize the error rate for a measurement process.

This presentation will impact the forensic science community by providing practitioners with the resources to improve their own practice of forensic science and, for the legal community, to give insight into how an uncertainty interval is developed.

Rule 702 of the Federal Rules of Evidence require that the evidence in a forensic proceeding be reliable. One of the criteria that the court set forth in interpreting Rule 702 is the error rate of techniques being proffered. This paper will impact the forensic science community by offering two methods of determining an error rate.

One classic metric of measurement uncertainty is the confidence interval on the mean (the arithmetic average), a function of that mean and the sample standard deviation, which connects the sample and the population using a *Student's-t* probability distribution.

This paper (and Part 1) together, looks at alternative ways of developing uncertainty estimates, which vary in their simplicity and their efficiency:

1. The Chebychev inequality (discussed in part 1);
2. The sign-test derived confidence interval on the median (discussed in this paper).
3. The uncertainty interval based upon the cumulative function (discussed in this paper).

Example 2: A confidence interval based upon the sign test on the median: This method requires only that the samples be independently drawn from the same continuous distribution; it makes no assumption concerning the form of that distribution. The sample values are placed in order from smallest to largest. For the laser-rangefinder- settling-time data used in example 1, rearranged in ordinal, rather than tabular form, the results obtained are:

4, 7, 8, 8, 9, 9, 9, 9, 10, 10, 10, 11, 11, 11, 11, 11, 11, 11, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 13, 13, 13, 13, 13, 13, 13, 13, 14, 14, 14, 14, 14, 14, 15, 15, 15, 16, 17, 18, 20, 20, 22, and 25.

Without going into theory, the confidence interval rests upon the assumption that the median lies either (a) before the first point (4) or after the last point (25), (b) between the first and second data points (4 & 7) or between the next-to-last and last data points (22 & 25), (c) between the second and third data points (7 & 8) or the third-from-last and-second from last data points (20 & 22) or (d) between the third and fourth

The widest possible confidence interval, (4, 25) fails to capture the median only if the true median is less than four or greater than 25; the next widest confidence interval (7, 22) fails to capture the true median only if it is not within the data range (except for the end points). The probability calculation turns put to be binomial with *p(success)* = ½. Thus, the confidence level for a confidence interval that extends k points from the extreme values of the data set is:

1 - (2 x *b(k; n, ½)),* where

b(k; n, ½) is the binomial distribution probability of *k* successes in *n* trials with a probability of success at each trial of ½.[1] The upper limit results only show:

 ** Presenting Author*

Upper Limit	Distance from (either) end	Confidence Level
25	0	100.0%
22	1	100.0%
20	2	100.0%
20	3	100.0%
18	4	100.0%
17	5	100.0%
16	6	100.0%
15	7	100.0%
15	8	100.0%
15	9	100.0%
14	10	100.0%
14	11	100.0%
14	12	100.0%
14	13	100.0%
14	14	100.0%
14	15	99.9%
13	16	99.8%
13	17	99.6%
13	18	99.2%
13	19	98.4%
13	20	97.2%
13	21	95.3%
13	22	92.8%
13	23	89.6%
12	24	86.2%
12	25	82.9%
12	26	80.2%
Median	27	

Discussion of the sign-test-based confidence Interval: One thing is clear: one cannot in general specify a specific confidence level, e.g., 95%, because the confidence interval is based upon discrete data points. (Here, you can get *near* that 95% value: The 95.3% confidence interval is between 13 and 21 milliseconds; the 92.8% confidence interval is between 13 and 22 milliseconds.) This discreteness is a function of the discrete nature of the underlying (binomial) distribution, which has non- zero probability only at the integers. That said, (a) this confidence interval is exact and not subject to any limitations on the underlying distribution; and, (b) only tradition lies behind using 95% confidence limits, or any other specific confidence level.

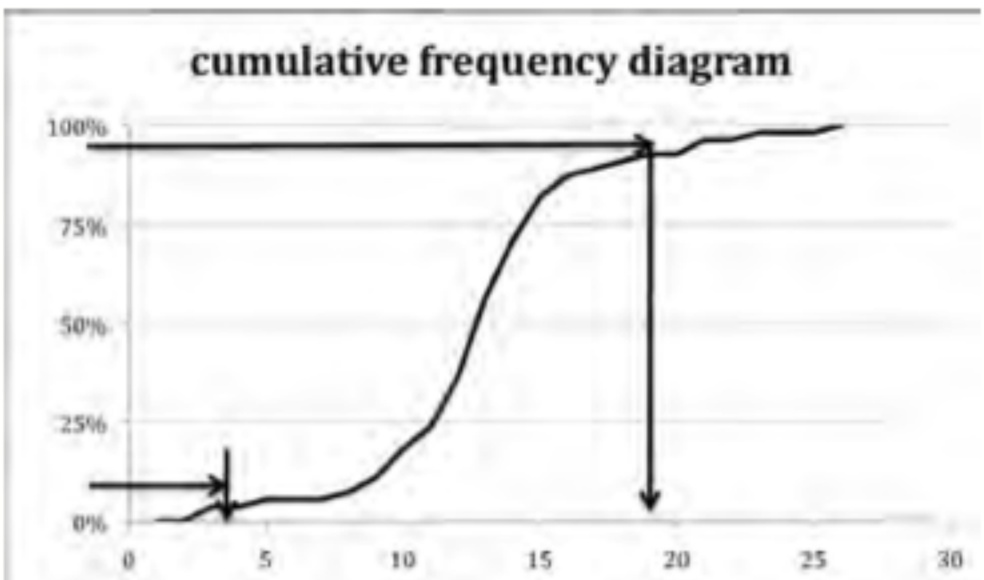
Example 3a: An Uncertainty Interval From Raw Data

The data here relates to the same 55 observations[2] of the time (in milliseconds) to stabilize a laser rangefinder. The first column is the millisecond value; the second column is the number of readings (of the 55) in each millisecond "bin"; the third column is the cumulative number of occurrences, and fourth column is the cumulative frequency. More specifically, **Col. (1)** represents time in milliseconds, and is graphed on the *x*-axis. The *x*-values, here, represent how many milliseconds it takes to stabilize the rangefinder, rounded to the nearest second. The values in column one are from a continuous distribution, e.g., a value of four seconds is recorded for any actual value: $3.5 < t \leq 4.5$ (the shaded row is an example within an example, with $x = 4$.) **Col. (2)** represents the number of occurrences: How many observations, out of the 55, does the settling time take on the value *x*, e.g., for *x* = four seconds, one value out of the 55 values is recorded. **Col. (3)** represents the cumulative number of occurrences: How many times in the experiment was a value of *x or fewer* seconds recorded. On the *x* = 4 line, the four-or-fewer seconds occurred three times (none for 0 seconds, none for one second, twice for two seconds, none for three seconds, and once for four seconds. The shaded portion of **col. (2)** contains these values.) **Col. (4)** represents the Cumulative frequency**:** and is graphed on the *y*-axis. **Col. 4** is but **Col. 3** divided by 55, the sample size, expressing the cumulative results in column three, as a decimal. For the shaded row, representing four or

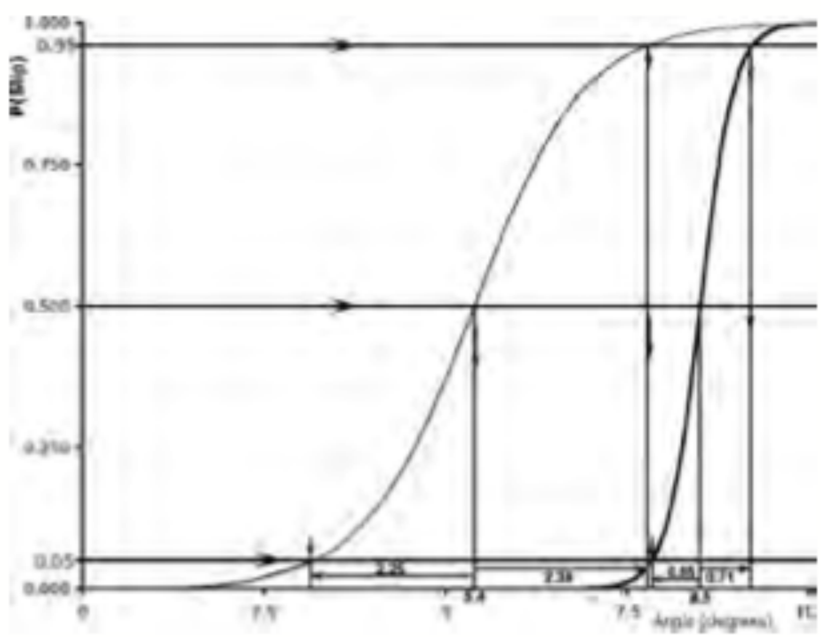
Col. (1)	Col. (2)	Col. (3)	Col. (4)
0	0	0	0.00
1	0	0	0.00
2	2	2	0.04
3	0	2	0.04
4	1	3	0.05
5	0	3	0.05
6	0	3	0.05
7	1	4	0.07
8	2	6	0.11
9	4	10	0.18
10	3	13	0.24
11	7	20	0.36
12	11	31	0.56
13	8	39	0.71
14	6	45	0.82
15	3	48	0.87
16	1	49	0.89
17	1	50	0.91
18	1	51	0.93
19	0	51	0.93
20	2	53	0.96
21	0	53	0.96
22	1	54	0.98
23	0	54	0.98
24	0	54	0.98
25	1	55	1.00

fewer seconds (actually a time of 4.5 seconds or less), there are three or fewer occurrences, giving a cumulative frequency of 5%.

Graphing the first column as the independent variable on the *x*-axis and the fourth column as the dependent variable on the *y*-axis, and drawing lines horizontally from the 5% and 95% ordinate values to the CF, and then down to the *x*-axis, the 90% CI is seen for the stabilization time for the laser rangefinder ranges from 4 to 19.6 milliseconds. (This can, of course, also be seen on the chart, just above.)



Example 3b: The confidence interval from a regression- developed cumulative function: The following diagram illustrates the uncertainty interval from the cumulative function by depicting two logistic regressions, developed from real data collected using a novel tribometric device:



Here, two datasets are compared by comparing the two curves, again drawing horizontals at 5% and 95% probabilities. Where each of the two horizontal lines cross the CFs, verticals have been drawn down to the x-axis. For the leftmost, shallower-sloping curve, the P(slip) = 5% is 2.25° away from the 5.4°, the median value. The P(Slip) = 95% is 2.38° from the median. Thus, the 90% CI can be expressed as (2.15°, 7.78°). Similarly, the rightmost logistics regression curve yields the CI: 8.5°± 0.7°. Obviously the rightmost curve represents an over three times more accurate instrument-and-process than the leftmost curve.

Discussion of the Cumulative Funtion methodology: Given the analysis, above, it appears that measurement uncertainty-the metrological equivalent of "error rate" in categorical classification-can be reasonably characterized by the methodology above. Moreover, by drawing the horizontals across at different ordinal values, we can obtain other CIs. For example, drawing horizontals at 25% and 75%, we obtain a 50% CI; by drawing across at 2½% and 97½%, we obtain the commonly-used 95% CI.

This is *not* a a true confidence interval, meeting the definition:

$$P(\underline{L} \leq parameter \leq \overline{L}) = (1-\alpha)$$

, where L and L are lower and upper bounds, respectively, and (1-〈) is the level of confidence, i.e., the probability that the *true parameter* lies between the lower and upper limits. There is no 'true parameter' coming into the analysis. This is not really a legal handicap, as none of the error rate explications require a confidence interval. Rather, what is described here is an uncertainty interval.

Discussion—General: The techniques developed in this paper provide tools to meet the error-rate criterion of *Daubert*.

An appropriate experimental design under which data is collected is essential to prevent the parameter under analysis from being confounded by some unanticipated factor. Given an appropriate experimental design, more data provides more reliable results than less.

The methodologies described above should allow forensic practitioners to estimate their own processes' or tests' uncertainty *via* calculation of the error rate expressed as a CI or UI. This is one valuable component of meeting the requirements of Federal Rule of Evidence 702 as explicated by the *Daubert* Trilogy. Accomplishing a data-specific error rate will make your tests and analyses better able to withstand scrutiny. Familiarity with the general principles described above will assist those in the legal community in proffering competent (or rejecting problematic) technical evidence.

References:

1. The formula for calculating the binomial probability of x successes in n trials with *p(success)* = ½ is =binomdist(x, n, ½)

The mathematical formula is: 1.

$$b(x,n,\tfrac{1}{2}) = \frac{n!}{x!(n-x)!}\left(\tfrac{1}{2}\right)^{x}\left(\tfrac{1}{2}\right)^{(n-x)}$$

2. While the appropriate sample size always depends upon the characteristics of that which is being characterized, a sample size of 55, is in our opinion, on the small side of reasonable for developing *Daubert* –purposed Confidence Intervals.

***Daubert* Error Rate, Forensic Science, Uncertainty Interval**