Approximate confidence intervals in calibration using the bootstrap

ARTICLE *in* ANALYTICAL CHEMISTRY · MAY 1993

Impact Factor: 5.64 · DOI: 10.1021/ac00058a012

1 AUTHOR:

Peter L Bonate
Astellas Pharmaceutical
85 PUBLICATIONS **1,647** CITATIONS

SEE PROFILE

Reprinted from Analytical Chemistry, 1993, 65.

Approximate Confidence Intervals in Calibration Using the Bootstrap

Peter L. Bonate

Lilly Laboratory for Clinical Research, Eli Lilly and Company, Wishard Memorial Hospital, 1001 West 10th Street, Indianapolis, Indiana 46202

The goal of calibration is to estimate a sample's concentration and the error associated with that estimate when only its signal response is known. Simulations were run to test the accuracy and precision of various parametric confidence intervals to confidence intervals generated by the bias-corrected, nonparametric bootstrap approach. None of the methods studied reached their asymptotic coverage probability, although the exact parametric confidence interval came the closest. The ranges of exact parametric confidence intervals were significantly larger than the other methods. Approximate parametric confidence intervals and bootstrap confidence intervals were both dependent on the number of replicates analyzed and on the coefficient of variation of the assay. When only a single replicate was available for analysis, the bootstrap method was dismal in containing the true sample concentration. As the number of replicates increases, the width of the bootstrap confidence interval converged to the exact parametric confidence interval, whereas the approximate parametric confidence interval width increased and began to diverge from the exact parametric confidence interval. Bootstrap confidence intervals produce maximal coverage with minimal range when 2–4 replicate samples are available for analysis.

INTRODUCTION

Assume we observe n pairs of data (x_i, Y_i), where $x_1 \ldots x_n$ are called the calibrators or standards and are known with certainty, and $Y_i \ldots Y_n$ are measurable random variables, often instrument signal. It is assumed that instrument signal is related to concentration by some model, $Y = f(x)$, called a calibration curve. A typical model is the simple linear model

$$Y_i = Mx_i + B + E_i \qquad (1)$$

where E_i are independent, normally distributed random errors with mean 0 and variance σ^2. M and B are least-squares estimates for the population parameters m and b. We now have a new observation, Y_o, with an associated unknown x value, x_o, called an "unknown". Assume observations, $Y_{o1} \ldots Y_{oq}$, are replicate measurements from x_o and the model used to describe the calibrators, $Y = f(x)$, applies. The objective of the calibration problem is to estimate x_o and the error associated with the estimate.

One solution to estimate x_o is the "classical predictor model" (CPM). The CPM rearranges the calibration curve solving for X.[1-4] For example, given the model in eq 1 and an observed value, Y_o, the true value of x_o can be estimated by

$$\hat{X}_o = (Y_o - B)/M \qquad (2)$$

where $\hat{X}_o$ is the estimator for x_o. Note that x_o is a constant and has no variance while $\hat{X}_o$ is a random variable which has a sampling distribution associated with it. Using the CPM, the mean and variance of $\hat{X}_o$ does not exist because after taking the expected value of $\hat{X}_o$, the ratio of normal random variables is related to a Cauchy distribution, which in spite of its symmetry about the mode tails toward zero too slowly for moments to exist.[5]

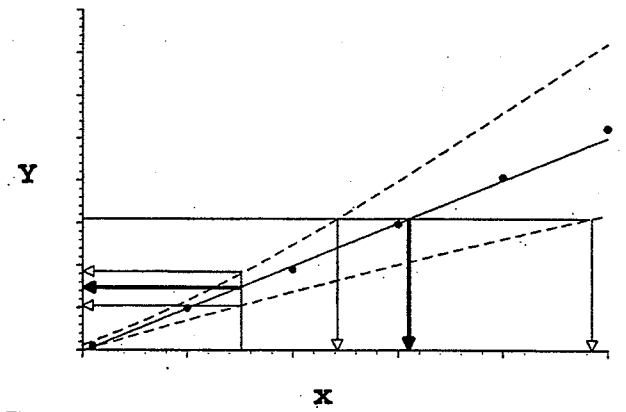

Figure 1. Plot of a theoretical calibration curve and its corresponding confidence interval demonstrating that while the confidence interval for a predicted Y value given x ($\hat{Y}|x$) is symmetric, the confidence interval for a predicted X given Y ($\hat{X}|Y$) may be asymmetric. Direction of arrows indicates whether the confidence interval is for $\hat{Y}|x$ or $\hat{X}|Y$. Open arrows indicate confidence region. Closed arrows indicate predicted value.

Development of an interval in which the true concentration of x_o is contained with a probability of $(1 - \alpha)$, called a confidence interval for x_o, is problematic. Under least-squares assumptions, the confidence interval for the true value of y given x is symmetrical around the predicted value, $\hat{Y}$. However, the confidence interval for the true x given an observed Y may lead to a nonsymmetric distribution. When a linear model with an intercept, i.e., eq 1, is used to describe the relationship between X and Y, the degree of asymmetry increases as $\hat{X}$ is removed from the mean of X. When a linear model without an intercept is used to describe the relationship between X and Y, the degree of asymmetry increases as $\hat{X}$ increases (Figure 1). Since the degree of asymmetry is changing across calibrators the problem is that the sampling distribution of X_o is unknown. When the variance of observed responses is constant across all levels of standards, i.e., homoscedastic data, Fieller's theorem may be applied to develop the confidence interval for the true concentration,

(1) Zar, J. H. *Biostatistical Analysis*, 2nd ed.; Prentice-Hall: Englewood Cliffs, NJ, 1984; pp 276–7.

(2) Bonate, P. *J. Chromatogr. Sci.* **1990**, *28*, 559–62.

(3) Osborne, C. *Int. Stat. Rev.* **1991**, *59*, 309–36.

(4) Buonaccorsi, J. P. *Technometrics* **1986**, *28*, 149–55.

(5) Rice, J. A. *Mathematical Statistics and Data Analysis*; Wadsworth and Brooks: Pacific Grove, CA 1988.

0003-2700/93/0365-1367$04.00/0

x_o, given only Y_o.[6-8] However, when the variance of the responses across all levels of standards is not constant, i.e., heteroscedastic data, as is often the case in instrumental analyses, Fieller's theorem is no longer valid. Some authors[9,10] propose using an approximate symmetrical $(1 - \alpha)$ % confidence interval for x_o

$$(X_L, X_U) = \bar{X}_o \pm \frac{ts}{M}\sqrt{\frac{\bar{X}_o^{\,P}}{q} + \frac{1}{\sum w_j} + \frac{(\bar{X}_o - \bar{X}_w)^2}{SSx_w}} \quad j = 1, 2 \ldots, n \quad (3)$$

where X_L and X_U are the lower and upper confidence values for x_o, $\bar{X}_o$ is the average back-calculated concentration determined from eq 2, t is Student's two-tailed t-distribution with $n + q - 3$ degrees of freedom, n is the number of standards in the calibration curve, P is the power of the weighting factor for the calibration curve (for example, if $1/X^2$ is the weighting factor, then $P = 2$), q is the number of "unknowns" used to determined $\bar{X}_o$, s is the square root of the mean square error from the calibration curve, $\bar{X}_w$ is the weighted average of n standards,

$$\bar{X}_w = \sum_j w_j x_j \Big/ \sum_j w_j \quad j = 1, 2, \ldots n \quad (4)$$

SSx_w is the weighted sum of squares of the standards,

$$SSx_w \equiv \sum_j w_j (X_j - \bar{X}_w)^2 \quad j = 1, 2, \ldots n \quad (5)$$

and w_j is the jth weighting factor used in the calibration curve. The weighting function, w_j, is arbitrary; usually $1/x$ or $1/x^2$ is used. This type of confidence interval will be referred to as the approximate parametric method (APM).

An exact asymmetric parametric solution to the confidence interval problem may be developed using the equation for the confidence interval of y_o given x_o. Iterating $\bar{X}_L$ and $\bar{X}_U$ until the relative difference function is minimized will generate the $(1 - \alpha)$% confidence interval for x_o

$$\min\{(|\overline{Y_o} - [M\bar{X}_L + B + ts(\bar{X}_L^{\,P}/q + 1/\sum w_j + (\bar{X}_L - \bar{X}_w)^2/SSx_w)^{1/2}]|)/\overline{Y_o} \times 100\%\} \quad (6)$$

$$\min\{(|\bar{Y}_o - [M\bar{X}_U + B - ts(\bar{X}_U^{\,P}/q + 1/\sum w_j + (\bar{X}_U - \bar{X}_w)^2/SSx_w)^{1/2}]|)/\overline{Y_o} \times 100\%\} \quad (7)$$

where $\bar{Y}_o$ is the average observed response from the unknown, and the other variables are as described in eqs 2 and 3. All the variables in eqs 6 and 7 are known, except for $\bar{X}_L$ and $\bar{X}_U$. This method is the mathematical equivalent to the graphical approach taken in Figure 1. Equations 6 and 7 will produce an exact confidence interval for the true concentration and will be referred to as the exact parametric method (EPM). The EPM will be the gold standard by which the other methods will be compared.

A statistic is formally defined as a function of a variable that has uncertainty in its measurement. For example, the mean, variance, and range are statistics. The bootstrap is a computer-intensive, nonparametric technique used to estimate the standard errors of statistics with unknown distributions and is ideally suited to those statistics where parametric theory does not provide a solution. The bootstrap has been used to estimate correlation coefficients,[11-13] parameter accuracy in analysis of variance with repeated measures,[14,15] and errors associated in parameter estimates of pharmacokinetic data.[16] The purpose of these simulations was to examine how nonparametric confidence intervals using the bootstrap compare to parametric estimates in the calibration problem.

NONPARAMETRIC BOOTSTRAP METHOD

Freedman[14,15] presented the theory behind the bootstrap for the general linear model, showing that the bootstrap approximation to the distribution of the least-squares estimators is valid under normality of residuals. The algorithm is as follows: If eq 1 is the calibration curve for an analysis using n standards, the bootstrap process first transforms the residual vector, **E**, into its weighted residual vector

$$E_j = [Y_j - (Mx_j + B)]/x_j \quad j = 1, 2, \ldots n \quad (8)$$

From **E**, randomly select a residual with replacement, E_j*, until a new residual vector of size n is created. Each E_j* is sampled from **E** with equal probability. Let

$$Y_j^* = Mx_j + B + (E_j^* x_j) \quad (9)$$

be the bootstrapped data set where M and B are defined in eq 1, E_j* is the jth bootstrapped residual, and the quantity in parentheses $(E_j^* x_j)$ is a correction factor for the weighted residuals. Having now generated a new data set (**Y***, **x**), let

$$\hat{Y}_j^* = M^* x_j + B^* \quad (10)$$

be the bootstrapped model where M* and B* are the bootstrap estimates of m and b and $\hat{Y}_j$* is the bootstrap predicted estimate of Y. Given that Y_o is the observed response from a sample with true concentration x_o, then the bootstrapped estimate of x_o, $\hat{X}_o$*, is given by eq 2 after substituting M* for M and B* for B. Note that $\hat{X}_o$* is a random variable with error associated with it, while x_o is a constant. Repeating this process d times generates the sampling distribution of X_o. The bootstrap and Monte Carlo simulation provide a mean to estimate both the variance and bias of any statistic of interest. In the case of $\bar{X}_o$

$$\text{bias}(\bar{X}_o) = \bar{X}_o^* - \bar{X}_o \quad (11)$$

$$\text{variance}(\bar{X}_o) = \sum_j (\hat{X}_{oj}^* - \bar{X}_o^*)^2/d \quad j = 1, 2, \ldots d \quad (12)$$

where $\bar{X}_o$* is the mean of all bootstrap simulations

$$\bar{X}_o^* = \sum_j \hat{X}_{oj}^*/d \quad j = 1, 2, \ldots d \quad (13)$$

If a sufficient number of bootstrap samples are drawn (d = 1000), then the confidence interval for the true concentration can be determined using the $(Z/2)$% tails of the distribution of X_o*, where Z is the desired critical value, usually 10%.

One assumption of the bootstrap is that bias$(\bar{X}_o) = 0$. When this is not the case, taking the Z% tails leads to a biased estimate of the confidence interval. Efron proposed a procedure, leading to more general confidence intervals and corrects for bias, that centers the bootstrap distribution such that the bias equals zero and takes the adjusted percentile tails for the confidence interval.[17] The bootstrap estimates of standard error and bias may be applied not only to the

(6) Fieller, E. C. *Biometrika* **1932**, *24*, 428–40.
(7) Fieller, E. C. *J. R. Stat. Soc.* **1940**, *Suppl. 7*, 1–64.
(8) Fieller, E. C. *J. R. Stat. Soc., Ser. B* **1954**, *16*, 175–85.
(9) Miller, J. C.; Miller, J. N. *Statistics for Analytical Chemists*; Ellis Horwood Ltd.: Chichester, 1984; pp 94–5.
(10) Miller, J. N. *Analyst* **1991**, *116*, 3–14.
(11) Efron, B.; Tibshirani, R. *Stat. Sci.* **1986**, *1*, 54–77.
(12) Lunneborg, C. E. *Psychol. Bull.* **1985**, *98*, 205–12.
(13) Rasmussen, J. *Psychol. Bull.* **1987**, *101*, 136–9.
(14) Freedman, D. A. *Annu. Stat.* **1981**, *9*, 1218–28.
(15) Freedman, D. A.; Peters, S. C. *J. Am. Stat. Assoc.* **1984**, *79*, 97–106.
(16) Niedzwiecki, D.; Simonoff, J. S. *J. Pharmacokinet. Biopharm.* **1990**, *18*, 361–77.
(17) Efron, B. *SIAM Rev.* **1979**, *21*, 460–480.

averages but also to any statistic, no matter how complicated. The reader is referred to refs 17–19 for a more generalized exposition of the bootstrap methodology.

SIMULATION

The standards used in each calibration curve were randomly selected to have a lower concentration of 10–100 arbitrary units with the highest concentration 100 times the lower concentration. In all cases six standards were used, with three at the lower level and three at the higher level. For example, in one simulation the concentrations used may have been 25, 25, 25, 2500, 2500, and 2500 ng/mL. Next, the true linear model which described the relationship between Y and x, $y = f(x;m,b)$, was developed by randomly selecting a slope, m in the range of [1, 10] from a uniform distribution and randomly selecting an intercept, b, in the range of [−1, 1] from a uniform distribution. Observed values, Y_i, were then calculated by eq 1 where E_i was simulated at ± (5, 10, and 20% coefficient of variation, CV) normally distributed random error with mean $mx_i + b$.

The observed data (Y, x) were then fitted using the method of weighted least-squares to calculate the slope, intercept, coefficient of determination, and mean square error of the calibration curve. In all simulations, the weighting function was $1/x_i^2$. Assumptions regarding the quality of the calibration curve were as follow: The coefficient of determination for each calibration curve was greater than 98% in all cases; the slope and intercept were both significantly greater than zero at 90% confidence as tested using a Student's two-tailed t-test, and no back-calculated concentration determined using eq 2 had a relative error of greater than 1.5 × CV% from the target value. If any of these assumptions were not met, the data set was discarded and a new one created. The CPM (eq 2) was used to estimate back-calculated concentrations given the response.

Given the proper set of standards, the true value of the unknown sample, x_o, was selected. Two data sets were examined. In one set the outermost ends of the calibration curve were studied, i.e., ±10% from either the lowest or highest standard, and in the other set the middle 80% of the calibration range was studied. "Unknown" Y values, Y_{oj}, were calculated by eq 1, where E_j was the corresponding coefficient of variation used in the calculation of the calibration curve. The simulation examined 1, 2, 3, 5, and 7 replicates identically and independently drawn from x_o.

The pseudoexperimental data set was now complete. All observed signals, Y_o, were treated as unknowns, and the corresponding estimated sample concentration, $\hat{X}_o$, was back-calculated using eq 2. The estimated sample concentrations were then used to calculate the confidence interval for the true sample concentration using the exact parametric (eqs 3–5), approximate parametric (eqs 6 and 7), and the bootstrap methods.

One thousand bootstrap replications were used in the development of each confidence interval. Efron's bias-correction bootstrap procedure was used to correct for resampling bias. The accuracy of the bootstrap and parametric methods was compared by using a running counter in which the counter was increased if the 90% confidence interval contained the true sample concentration. A total of 1000 independent simulations were run under these experimental conditions. Another set of simulations was performed to examine the precision of the confidence interval methods. Using identical conditions as above, with the exception that 750 iterations were done, a counter was set to increment if the range of the generated bootstrap confidence interval was smaller than the parametric methods. The precision of the bootstrap and APM was compared to the EPM by examining the ratio of the range of the confidence interval generated by the EPM to the range of the confidence interval generated by the bootstrap or APM. A value greater than 1 was indicative of the EPM confidence interval being larger than the comparator confidence interval. A paired t-test was utilized to test the hypothesis that the range of the confidence interval generated by the bootstrap was significantly smaller than the range of the confidence interval generated by the APM. Testing for trends was done by linear regression analysis. All simulations and analyses were written in GAUSS[20] and performed on a Compaq 386 personal computer. All data computations were verified using Number Cruncher Statistical System, version 5.03.[21]

Table I. Example of Algorithm for Equations 6 and 7

std concn (ng/mL)	response	back-calcd concn	weighted concn ($1/x^2$)	$w_i{*}x_i$	SSx_w
50	215.99	42	0.00040000000	2.00×10^{-2}	0
50	279.11	55	0.00040000000	2.00×10^{-2}	0.00
50	274.78	54	0.00040000000	2.00×10^{-2}	0.00
5000	24787.95	4870	0.00000004000	2.00×10^{-4}	0.98
5000	28625.55	5624	0.00000004000	2.00×10^{-4}	0.98
5000	22301.57	4381	0.00000004000	2.00×10^{-4}	0.98
	sum		0.00120012000	6.06×10^{-2}	2.94

Summary Statistics of Regression Equation

calibration curve	response = 5.09 concn
MSE	0.363
$S = \sqrt{\text{MSE}}$	0.603
coeff of determination (R^2)	98.84
student's t-value	$t_{\alpha/2}(7) = 1.895$
$\bar{X}_w$	50.495

Sample To Be Estimated

unknown response in triplicate	unknown back-calcd concn
372.80	73
428.10	84
410.80	81
av 403.9	av 79

Iteration of Equations 6 and 7

X_U	relative difference	X_L	relative difference
76	19.67	76	11.22
80	15.34	70	2.63
90	4.61	65	4.48
95	0.71	67	1.64
96	1.78	68	0.22
94	0.35	69	1.20
93	1.41		

EXAMPLE

As an example to the algorithm, assume the calibration data is as given in Table I. The true model was $Y = 5x$, and the observed Y values were calculated with 10% CV. The weighted least-squares fit (using $1/X^2$ as the weighting factor) was $Y = 5.09x$, $R^2 = 98.84$, and MSE = 0.363. Assume analysis of an unknown sample in triplicate whose true concentration was 75 ng/mL. The observed values are reported in Table I. The CPM estimate of the sample concentration was 73, 84, and 81 ng/mL, respectively, with an average of 79 ng/mL.

(18) Efron, B.; Gong, G. *Am. Stat.* **1983**, *37*, 36–48.
(19) Wasserman, S.; Bockenholt, U. *Psychophysiology* **1989**, *26*, 208–21.
(20) *GAUSS Users Manual*; Aptech Systems, Inc.: Kent, WA, 1991.
(21) *Number Cruncher Statistical System, Version 5.03, Users Manual*; NCSS: Kaysville, UT, 1990.

Table II. Comparison of Parametric to Nonparametric Confidence Intervals When True Concentration Is at Outer 10% Ends of the Calibration Curve and the Desired Level of Confidence Is 90%

	% simulations containing the true mean				
	no. of unknowns				
	1	2	3	5	7
	Coefficient of Variation = 5%				
exact parametric method (EPM)	87	87	87	82	80
approximate parametric method (APM)	58	61	69	67	67
nonparametric bootstrap method	34	68	69	76	75
comparison of variance estimators					
range of bootstrap < range of EPM	100	95	90	78	76
range of bootstrap < range of APM	100	81	58	6	0
range of EPM < range of APM	0	0	0	0	0
	Coefficient of Variation = 10%				
exact parametric method (EPM)	85	87	84	80	82
approximate parametric method (APM)	55	71	71	68	70
nonparametric bootstrap method	32	61	69	74	78
comparison of variance estimators					
range of bootstrap < range of EPM	100	95	87	76	78
range of bootstrap < range of APM	100	80	57	6	0
range of EPM < range of APM	0	0	0	0	0
	Coefficient of Variation = 20%				
exact parametric method (EPM)	79	78	79	76	73
approximate parametric method (APM)	57	66	68	67	61
nonparametric bootstrap method	31	58	68	75	72
comparison of variance estimators					
range of bootstrap < range of EPM	100	88	77	64	62
range of bootstrap < range of APM	100	72	44	2	0
range of EPM < range of APM	0	1	1	0	0

Table III. Comparison of Parametric to Nonparametric Confidence Intervals When True Concentration Is In Middle 80% of the Calibration Curve and the Desired Level of Confidence Is 90%

	% simulations containing true mean				
	no. of unknowns				
	1	2	3	5	7
	Coefficient of Variation = 5%				
exact parametric method (EPM)	83	81	84	87	85
approximate parametric method (APM)	55	66	70	73	66
nonparametric bootstrap method	34	60	69	77	74
comparison of variance estimators					
range of bootstrap < range of EPM	100	92	85	86	85
range of bootstrap < range of APM	100	78	52	23	12
range of EPM < range of APM	0	0	0	0	0
	Coefficient of Variation = 10%				
exact parametric method (EPM)	85	82	82	85	86
approximate parametric method (APM)	61	70	70	73	75
nonparametric bootstrap method	37	64	69	77	80
comparison of variance estimators					
range of bootstrap < range of EPM	100	92	82	82	79
range of bootstrap < range of APM	100	77	46	23	14
range of EPM < range of APM	0	0	0	0	0
	Coefficient of Variation = 20%				
exact parametric method (EPM)	77	74	76	77	75
approximate parametric method (APM)	56	61	66	67	67
nonparametric bootstrap method	32	55	67	77	73
comparison of variance estimators					
range of bootstrap < range of EPM	100	86	75	68	64
range of bootstrap < range of APM	100	66	36	15	9
range of EPM < range of APM	0	0	0	1	1

The bootstrap distribution (using $d = 1000$) of the mean distribution is given in Figure 2. The distribution is decidely nonnormal (Martinez–Iglewicz Test for Normality = 1.1177, $p < 0.05$[22]). The bootstrap 90% confidence interval, after correcting for bias, was (73, 90 ng/mL). The exact 90% parametric confidence interval was (68, 94 ng/mL), while the approximate confidence interval was (66, 92 ng/mL). In this instance, even though the distribution of the unknown sample is non-Gaussian, all calculated confidence intervals contain the true concentration.

(22) Martinez, J.; Iglewicz, B. *Biometrika* 1981, *68*, 331–3.

RESULTS AND DISCUSSION

The results of the simulation examining the accuracy of the calculated confidence intervals are presented in Tables II and III. The most surprising result was that no matter where on the calibration curve the true concentration was, the percentage of simulations capturing the true concentration within a calculated confidence interval was the same. Analysis of variance indicated that there was no difference between whether the sample was in the middle of the calibration curve or was at the end of the calibration curve ($p > 0.1$); the percent of simulations containing the true concentration was the same.

Table IV. Comparison of Parametric to Nonparametric Confidence Intervals When True Concentration Is at Outer 10% Ends of the Calibration Curve and the Desired Level of Confidence Is 90%[a]

	no. of unknowns				
	1	2	3	5	7
Coefficient of Variation = 5%					
mean range of EPM to mean range of NBM	4.59	2.62	1.68	1.52	1.42
standard deviation	2.13	1.42	0.88	0.63	0.54
minimum	2.32	0.37	0.44	0.41	0.38
maximum	17.70	11.56	11.82	6.55	6.27
mean range of EPM to mean range of APM	1.43	1.52	1.56	1.78	1.92
standard deviation	0.52	0.49	0.40	0.45	0.43
minimum	1.10	1.09	1.11	1.22	1.15
maximum	5.14	4.62	5.98	4.99	5.92
paired *t*-test testing range of NBM is < range of APM	**52.73**	**26.40**	**4.67**	−13.23	−25.68
Coefficient of Variation = 10%					
mean range of EPM to mean range of NBM	3.74	1.97	1.57	1.36	1.28
standard deviation	0.69	0.69	0.50	0.42	0.35
minimum	2.29	0.34	0.30	0.40	0.42
maximum	8.29	5.59	3.80	4.05	2.98
mean range of EPM to mean range of APM	1.22	1.33	1.42	1.62	1.76
standard deviation	0.10	0.14	0.14	0.24	0.28
minimum	1.05	1.05	1.00	0.97	0.99
maximum	2.14	2.58	2.16	4.05	5.20
paired *t*-test testing range of NBM is < range of APM	**110.57**	**25.82**	**7.56**	−15.37	−28.23
Coefficient of Variation = 20%					
mean range of EPM to mean range of NBM	3.62	1.81	1.39	1.16	1.10
standard deviation	0.57	0.64	0.47	0.32	0.27
minimum	1.73	0.29	0.33	0.26	0.21
maximum	6.25	4.09	3.35	2.40	2.15
mean range of EPM to mean range of APM	1.26	1.32	1.41	1.55	1.69
standard deviation	0.07	0.10	0.14	0.20	0.27
minimum	1.01	0.98	0.95	0.93	0.90
maximum	1.69	2.17	2.32	2.31	2.59
paired *t*-test testing range of NBM is < range of APM	**113.81**	**20.71**	0.98	−27.28	−39.26

[a] EPM = exact parametric method; APM = approximate parametric method; NBM = nonparametric bootstrap method. Boldface values are significant at $p \leq 0.01$.

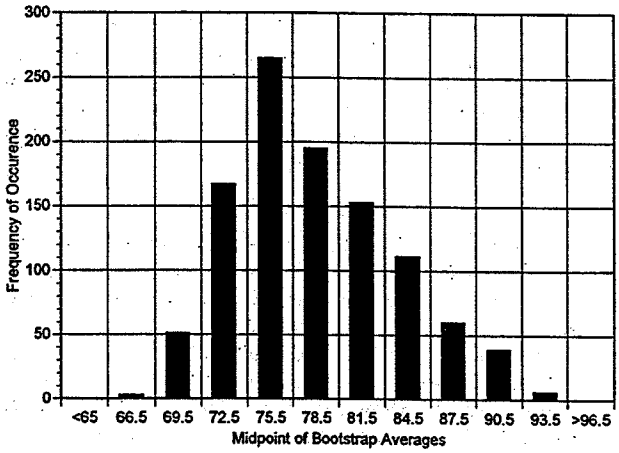

Figure 2. Histogram of the bootstrap sampling distribution of a sample whose concentration is unknown but whose response is known. The back-calculated concentrations were calculated using the calibration data in Table I. Refer to Example Section for details.

Without doubt, the superior method was the EPM. The EPM contained the true concentration of the same percentage independent of the number of replicates examined, while the APM ($p < 0.001$) and bootstrap ($p < 0.001$) were dependent on the number of replicates examined. As the CV of the assay was increased, the probability of containment for the EPM did decrease slightly but was still superior to the other methods examined. Both the APM and bootstrap were dependent on the CV of the assay ($p < 0.05$).

For both the APM and bootstrap the accuracy of containing the true concentration increased as the number of replicates increased. Neither method contained the true concentration as well as the EPM did. While this may seem to indicate that neither the APM or bootstrap is very accurate in containing the true concentration, Golberg[23] points out: "The true mean may not lie within the desired confidence interval. However, in the long run, for many tests 95% of the intervals constructed in this manner will contain the true value.". Taking many random samples and forming confidence intervals for each one, about 95% of these intervals would contain the true value. Any data set that is collected on a continuous random variable is only one realization of an infinite number of outcomes. With this in mind, the accuracy of the bootstrap and the APM was really quite good.

When only one observation per sample was available, the bootstrap was dismal in containing the true concentration. Failure of the bootstrap to contain the true concentration was no surprise and was probably due to the inability of the bootstrap to generate a sampling distribution. When the number of replicates increased, the bootstrap rapidly reached its asymptotic coverage probability, which was significantly better than that of the APM.

The results of the simulation examining the precision of the calculated confidence intervals are presented in Tables IV and V. At every level of CV examined, the range of the confidence interval calculated by the EPM was significantly larger than any of the other methods studied ($p < 0.001$). Where the unknown sample was at on the calibration curve had a significant effect on the precision of the calculated confidence interval ($p < 0.001$). When the sample was at the outer ends of the calibration curve, the confidence interval was larger than if the sample was near the middle of the calibration curve. This result was expected for the EPM and APM, which are based on least-squares assumptions, but was

(23) Golberg, M. A. *An Introduction to Probability Theory with Statistical Applications*; Plenum Press: New York, 1984.

Table V. Comparison of Parametric to Nonparametric Confidence Intervals When True Concentration Is in Middle 80% of Calibration Curve and Desired Level of Confidence Is 90%[a]

	no. of unknowns				
	1	2	3	5	7
Coefficient of Variation = 5%					
mean range of EPM to mean range of NBM	3.25	1.80	1.41	1.21	1.18
standard deviation	0.28	0.54	0.41	0.29	0.26
minimum	2.28	0.44	0.36	0.35	0.29
maximum	4.48	3.23	2.35	2.05	1.75
mean range of EPM to mean range of APM	1.18	1.32	1.45	1.68	1.87
standard deviation	0.02	0.03	0.03	0.04	0.05
minimum	1.16	1.30	1.42	1.64	1.83
maximum	1.45	1.51	1.64	2.04	2.21
paired *t*-test testing range of NBM is < range of APM	**206.38**	**23.77**	**2.78**	−41.41	−70.34
Coefficient of Variation = 10%					
mean range of EPM to mean range of NBM	3.28	1.78	1.40	1.23	1.17
standard deviation	0.29	0.52	0.42	0.30	0.25
minimum	2.35	0.44	0.29	0.31	0.37
maximum	4.02	2.77	2.57	1.93	1.78
mean range of EPM to mean range of APM	1.20	1.32	1.44	1.66	1.86
standard deviation	0.02	0.01	0.02	0.02	0.03
minimum	1.16	1.30	1.42	1.63	1.83
maximum	1.28	1.39	1.55	1.75	2.14
paired *t*-test testing range of NBM is < range of APM	**200.64**	**23.89**	**2.67**	−39.34	−72.67
Coefficient of Variation = 20%					
mean range of EPM to mean range of NBM	3.26	1.68	1.30	1.08	1.06
standard deviation	0.37	0.55	0.40	0.29	0.24
minimum	1.76	0.35	0.38	0.26	0.26
maximum	4.72	2.79	2.30	1.87	1.65
mean range of EPM to mean range of APM	1.27	1.36	1.47	1.68	1.87
standard deviation	0.05	0.03	0.02	0.02	0.02
minimum	1.17	1.30	1.43	1.63	1.83
maximum	1.57	1.54	1.64	1.79	1.97
paired *t*-test testing range of NBM is < range of APM	**146.64**	**16.07**	**12.20**	−57.13	−95.45

[a] EPM = exact parametric method; APM = approximate parametric method; NBM = nonparametric bootstrap method. Boldface values are significant at $p \leq 0.01$.

unexpected for the bootstrap, which does not make use of least-squares assumptions. It is reassuring to know that the nonparametric method approximates the error structure of the parametric methods.

As the number of replicates was increased, the range of the bootstrap confidence interval tended to converge to the range of the EPM confidence interval ($p < 0.001$). This would be expected based on the central limit theorem. As the number of replicates increased, the nonparametric and parametric theories should converge to a normal distribution and be superimposable. However, as the number of replicates was increased, the range of the APM confidence interval did not converge to the EPM confidence interval but began to diverge ($p < 0.001$). This was also demonstrated by the observation that as the number of replicates was increased, an inverse relationship between the range of the bootstrap confidence interval and the range of the APM confidence interval was observed ($p < 0.05$). Paired *t*-tests reveal that when 1–3 replicate samples were analyzed the confidence interval generated by the bootstrap method was greater than the APM, but when more samples were analyzed, the confidence interval generated by the bootstrap was smaller than the APM. Assuming the range of the EPM to be constant, this would indicate that as the sample size was increased, the range of the APM confidence interval decreased, whereas the bootstrap confidence interval increased.

When the analyst has only one sample to analyze, as is often the case in routine analytical work, the confidence interval generated by the EPM may be 2–4 times as large as the bootstrap method and 1–2 times as large as the APM. When multiple samples are available, the bootstrap converges to the EPM in both coverage and precision.

CONCLUSION

In choosing which type of confidence interval methodology to use, the analyst must strike a balance between accuracy and precision. No method studied was both accurate and precise. As with most analytical assays, accuracy increases at the expense of precision. For the bootstrap method, the best balance between precision and accuracy was achieved when 2–4 replicates were analyzed. The EPM was still the gold standard by which all other confidence intervals are to be judged, but its large confidence intervals may be too liberal for some analysts. When only one sample was available for analysis, either of the parametric methods produced adequate results. When 2–5 samples were available for analysis, the APM and bootstrap were comparable. When a large number of replicates were available, the bootstrap and EPM were comparable in precision and accuracy.

RECEIVED for review September 4, 1992. Accepted February 9, 1993.