ARMY RESEARCH LABORATORY

Statistical Resampling Plans

Malcolm S. Taylor

ARL-TR-1794 September 1998

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.

Army Research Laboratory
Aberdeen Proving Ground, MD 21005-5067

ARL-TR-1794 **September 1998**

Statistical Resampling Plans

Malcolm S. Taylor
Information Science and Technology Directorate, ARL

Abstract

Statistical resampling plans can be thought of as procedures for "recycling data." They commonly involve computer-intensive methods for the interrogation of data through repeated inspection. The procedures detailed in this document free the data analyst from two major restrictions underlying much of the development in mathematical statistics: the normal or Gaussian assumption, and the necessity to focus on statistics that are in closed form and mathematically tractable.

Acknowledgments

I imposed upon a greater-than-usual number of people in the preparation of this manuscript. Included among them are Tom Walker, Aberdeen Test Center, who provided the impact data in Table 1 and Ted Muehl, Survivability/Lethality Analysis Directorate, the target description in Figure 4. Bill Cunningham instructed me on the use of graphics packages and, along with Tim Hanratty, performed the computations underlying Tables 2 and 5. Fred Brundick mercifully freed me from a host of LaTeX entanglements.

INTENTIONALLY LEFT BLANK.

Table of Contents

INTENTIONALLY LEFT BLANK.

List of Figures

List of Tables

INTENTIONALLY LEFT BLANK.

1. The Bootstrap

Let $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ be a random sample from a population with an unspecified distribution F and let $\mathrm{s}(\mathbf{x})$ be an arbitrary statistic. If the statistic estimates a parameter θ of the distribution F, then we will write $\hat{\theta} = s(\mathbf{x})$, where the hat notation denotes an estimate. To have value as an estimator, a statistic needs to estimate accurately; common measures of accuracy are standard error and bias. Almost without exception (the sample mean being a notable exception) an estimate of the standard error of a statistic cannot be determined from a single random sample. The bootstrap procedure attempts to overcome this impediment through intensive computation. Simply put, the bootstrap procedure attempts to provide an estimate of standard error of an arbitrary statistic $\hat{\theta}$ given only a single random sample. With this as its goal, the name "bootstrap" is not a misnomer.

A fundamental notion (or artifice) is that of a bootstrap sample. A bootstrap sample $\mathbf{x}^\star = (x_1^\star, x_2^\star, \ldots, x_n^\star)$ is a random sample of size n drawn with replacement from the class of n objects making up the random sample $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, and so we might have $x_1^\star = x_{18}, x_2^\star = x_7, x_3^\star = x_3, \ldots, x_n^\star = x_5$, with some of the $x_1, x_2, \ldots, x_n$ values appearing several times in the bootstrap sample and, others, not at all.

Why would anyone want to engage in such an exercise?

Kiefer and Wolfowitz (1956) established for a univariate random sample, $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, that the corresponding empirical distribution $\hat{F}$ is the nonparametric maximum likelihood estimate of the unknown distribution F. This result supports the construction of bootstrap samples in one dimension, but extension to a multivariate situation is not direct, and an alternative rationale for bootstrap sampling is needed.

Toward this end, consider the so-called Dirac-comb density estimator associated with a one-dimensional data set $\mathbf{x} = (x_1, x_2, \ldots, x_n)$. The Dirac-comb density estimator is given by

$$\hat{f}_\delta(x) = \frac{1}{n} \sum_{i=1}^{n} \delta(x - x_i), \tag{1}$$

where the expression $\delta(\cdot)$ is a Dirac δ (or unit impulse) function, which can in turn be expressed as

$$\delta(x) = \lim_{\tau \to 0} \frac{1}{\sqrt{2\pi}\tau} e^{-\frac{x^2}{2\tau^2}}. \tag{2}$$

Expression (1) is a function that is zero everywhere except at the data points.* At each data point the density is unbounded; the assignment of probability mass 1/n at that location is merely suggestive of the limiting process. As a nonparametric density estimator, $\hat{f}_\delta(x)$ is unappealing. It dictates that any subsequent experiment can only provide the data points already observed, and moreover, those points must have equal probability of occurrence.

The representation $\hat{f}_\delta(x)$ is not without value in some situations; for example, the mean of $\hat{f}_\delta(x)$ is

$$\mu_\delta = \int_{-\infty}^{\infty} x\hat{f}_\delta(x)dx = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x} \tag{3}$$

and

$$\sigma_\delta^2 = \int_{-\infty}^{\infty} (x-\bar{x})^2\hat{f}_\delta(x)dx = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2 = s^2. \tag{4}$$

The Dirac-comb density estimator may be easily extended to higher dimensions. For a sample $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ of p-component random vectors, $\hat{f}_\delta(x)$ becomes

$$\hat{f}_\delta(X) = \frac{1}{n}\sum_{i=1}^{n}\delta(X - X_i), \tag{5}$$

where

$$\delta(X) = \lim_{\tau\to 0}\left(\frac{1}{\sqrt{2\pi}\tau}\right)^p exp\left(-\frac{\sum_{j=1}^{p} x_j^2}{2\tau^2}\right) \tag{6}$$

and where x_j is the jth component of an observation $X = (x_1, x_2, \ldots, x_p)$.

---

*The unit impulse function (2) is more commonly expressed as

$$\delta(x - x_i) = \begin{cases} 0, & x \neq x_i \\ \infty, & x = x_i \end{cases}$$

and

$$\int_{-\infty}^{\infty}\delta(x-x_i)dx = 1.$$

Since in practice bootstrap sampling is carried out in both univariate and multivariate situations without regard to dimension, viewing the bootstrap procedure as based on a Dirac-comb estimator results in a more unified theory.

Through a straightforward Monte Carlo procedure in which many samples of size n are drawn with replacement from $\hat{F}$—bootstrap samples—an approximation to the sampling distribution of $\hat{\theta}$ is produced. In order for the statistical accuracy of an estimate to be determined, properties of its sampling distribution must be established. For each bootstrap sample $\mathbf{x}_1^\star, \mathbf{x}_2^\star, \ldots, \mathbf{x}_b^\star$ we determine a corresponding $\hat{\theta}_i^\star = s(\mathbf{x}_i^\star),\ i = 1, 2, \ldots, b$. This empirical approximation to the sampling distribution of $\hat{\theta}$ allows an assessment of standard error to be made.

Some attention to notation is in order. If the standard error of $\hat{\theta}$ is denoted as $\sigma_F(\hat{\theta})$, the bootstrap estimate of standard error should be written $\sigma_{\hat{F}}(\hat{\theta}^\star)$, where the star $(\star)$ notation continues to indicate an estimate of θ corresponding to a bootstrap sample from the empirical distribution $\hat{F}$. The number of distinct bootstrap samples from a data set of n distinct values can be shown to be $m = C_n^{2n-1}$. For a modest sample size, say n = 10, this means that 92,378 bootstrap samples must be evaluated, each with a probability of occurrence p_j given by a multinomial distribution, in order to determine $\sigma_{\hat{F}}(\hat{\theta}^\star)$. The computational burden imposes practical limitations on the sample size n, but formally

$$\sigma_{\hat{F}}(\hat{\theta}^\star) = \{\sum_{j=1}^{m} p_j [s(\mathbf{x}_j^\star) - s(\cdot)]^2\}^{1/2}, \tag{7}$$

where $s(\cdot) = \sum_{j=1}^{m} p_j s(\mathbf{x}_j^\star)$, and $\mathbf{x}_1^\star, \mathbf{x}_2^\star, \ldots, \mathbf{x}_m^\star$ are all distinct, is the appropriate expression.

Of course $\sigma_{\hat{F}}(\hat{\theta}^\star)$ is in general not equal to $\sigma_F(\hat{\theta})$ (the value we really seek), but it is the bootstrap approximation. Efron and Tibshirani (1993) refer to it as the ideal bootstrap estimate of standard error. As suggested, the ideal may not be easily attained because of the attendant combinatorial explosion, but an approximation to the ideal may be determined via the following algorithm.

The bootstrap algorithm.

1. Obtain b independent bootstrap samples $\mathbf{x}_1^\star, \mathbf{x}_2^\star, \ldots, \mathbf{x}_b^\star$ by sampling with replacement from a random sample $\mathbf{x} = (x_1, x_2, \ldots, x_n)$.

2. Evaluate $\hat{\theta}_i^\star = s(\mathbf{x}_i^\star), \quad i = 1, 2, \ldots, b.$

3. Calculate

$$est\, \sigma_{\hat{F}}(\hat{\theta}^\star) = \{\sum_{i=1}^{b} [\hat{\theta}_i^\star - \hat{\theta}^\star(\cdot)]^2/(b-1)\}^{1/2},$$

where $\hat{\theta}^\star(\cdot) = \sum_{i=1}^{b} \hat{\theta}_i^\star/b.$

Notice in step 3 of the algorithm use of the notation $est\, \sigma_{\hat{F}}(\hat{\theta}^\star)$; an alternate expression, $\hat{\sigma}_{\hat{F}}(\hat{\theta}^\star)$, could have been used.

An important asymptotic result, stated here without proof, is that in the limit as $b \to \infty$, the bootstrap estimate of standard error converges in probability to the ideal bootstrap estimate:

$$\lim_{b\to\infty} est\, \sigma_{\hat{F}}(\hat{\theta}^\star) = \sigma_{\hat{F}}(\hat{\theta}^\star). \tag{8}$$

This result provides relief from the tedius, and often impractical, attempt at enumeration of the distinct bootstrap configurations needed for evaluation of equation (7).

Before proceeding further, an example is in order.

Example 1. Measures of dispersion for bivariate data are often relevant in applied problems, including ballistic studies, where the ordered pairs (x_i, y_i) correspond to impact locations on a target. Measures of bivariate dispersion have been extensively studied (e.g., Grubbs [1964]), including an intriguing statistic known as the extreme spread or bivariate range. For a sample $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ the extreme spread may be expressed as

$$R = \max_{i,j} [(x_i - x_j)^2 + (y_i - y_j)^2]^{1/2} \quad i, j = 1, 2, \ldots, n. \tag{9}$$

The extreme spread is simply the maximum of the distances between all C_2^n pairs of points in the sample.

This statistic is intuitively appealing and simple to evaluate, and yet knowledge of its distribution remains largely confined to Monte Carlo studies. This holds true even under a normal

assumption for the variates (x_i, y_i), except for the trivial case, n=2, in which R can be shown to follow a chi distribution.

In Table 1, impact coordinates for 10 rounds of small arms ammunition are listed; a scatterplot of the data appears in Figure 1.

Table 1. Impact coordinates for 10 rounds of small arms ammunition fired against a vertical target (measurements are in inches)

(6.5, 33.0)	(11.5, 7.0)
(3.0, 17.0)	(13.0, 3.5)
(6.5, 18.5)	(10.0, 7.5)
(8.5, 16.0)	(15.5, 13.0)
(8.5, 10.5)	(21.5, 15.0)

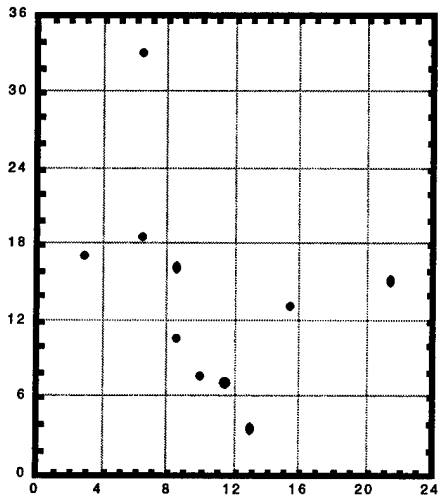

Figure 1. Location of impact for 10 rounds of small arms ammunition.

Estimates of standard error for the extreme spread from the impact data in Table 1 were computed, following the bootstrap algorithm. Those estimates of standard error for values of b=50(50)300 appear in Table 2.

A histogram of b = 300 bootstrap replications is shown in Figure 2. Although $C_2^{10} = 45$ distinct values of extreme spread are possible, after 300 replications, only 14 have been realized. This is not unexpected; the distribution will be negatively skewed under the influence of the max operator appearing in statistic (9).

Table 2. Bootstrap estimates of standard error (300 bootstrap replications gave the tabled values of $est\,\sigma_{\hat{F}}(\hat{\theta}^\star)$)

b:	50	100	150	200	250	300
est $\sigma_{\hat{F}}(\hat{\theta}^\star)$	5.74	5.72	5.84	5.97	6.07	6.06

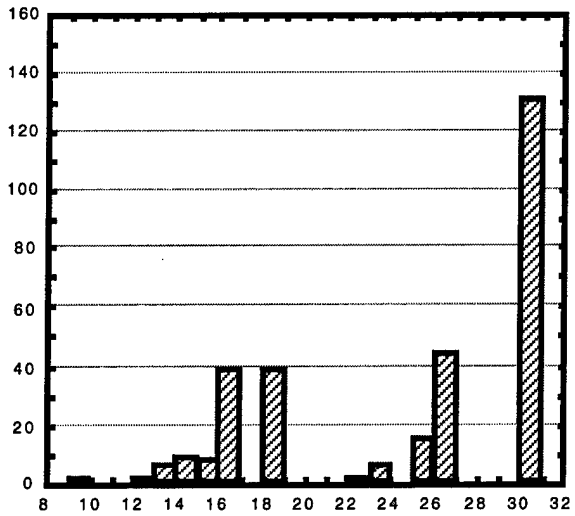

Figure 2. A histogram of b=300 bootstrap replications of the extreme spread for the impact data in Table 1.

Guidance regarding the number of bootstrap replications necessary to estimate a standard error with an acceptable degree of precision is limited mostly to empirical evidence. This evidence suggests that a few hundred bootstrap replications is usually adequate, although, on occasion, a smaller number may suffice. A definitive answer is strongly case-dependent, and involves the rate of convergence in expression (8).

2. The Parametric Bootstrap

There is no inherent requirement that the bootstrap approach be restricted to a nonparametric setting. If there is support for a specific form for the distribution F, then F can be estimated parametrically. We will adopt the notation $\hat{F}_f$ to indicate that a specific functional form, $f(\cdot)$, has been assumed for the random variable. Suppose for the impact data in section 1 the sample $\mathbf{z} = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$ is assumed to have come from a normal population with parameters

μ and Σ. Estimates of the mean vector and covariance matrix will be given by $(\overline{x}, \overline{y})'$ and

$$\frac{1}{n-1}\begin{bmatrix} \Sigma(x_i-\overline{x})^2 & \Sigma(x_i-\overline{x})(y_i-\overline{y}) \\ \Sigma(x_i-\overline{x})(y_i-\overline{y}) & \Sigma(y_i-\overline{y})^2 \end{bmatrix}.$$

From the impact data in Table 1, we obtain a mean vector $\hat{\mu} = (10.45, 14.10)'$ and covariance matrix $\hat{\Sigma} = \begin{bmatrix} 27.74 & -15.27 \\ -15.27 & 67.76 \end{bmatrix}$. The bootstrap procedure, after parameterizing F (the empirical distribution function $\hat{F}$ is replaced by a normal distribution $\hat{F}_f = N(\hat{\mu}, \hat{\Sigma})$) reduces to an ordinary Monte Carlo procedure. Instead of sampling with replacement from $\mathbf{z}$ to generate bootstrap samples $\mathbf{z}_1^\star, \mathbf{z}_2^\star, \ldots, \mathbf{z}_b^\star$, Monte Carlo samples are drawn directly from $\hat{F}_f$ to produce $\mathbf{z}_i^\star$, $i = 1, 2, \ldots, b$. The histogram of $\hat{R}^\star$ values for b=300 is shown in Figure 3. The estimate of standard error from these data, *est* $\sigma_{\hat{F}_f}(\hat{\theta}^\star) = 6.71$—compared to 6.06 for the nonparametric model.

Taylor and Grubbs (1975) carried out extensive Monte Carlo simulation studies of the extreme spread distribution for small sample sizes, under a standard normal assumption for the underlying distribution F, and concluded that the distribution is adequately described by a chi variate.

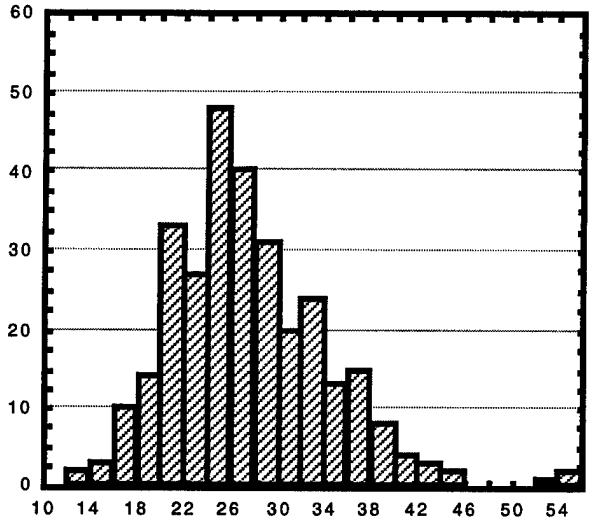

Figure 3. Histogram of $\hat{R}^\star$ values for b=300 replications under a normal assumption: $F = N(\mu, \Sigma)$.

3. Bias Estimation

The bootstrap procedure described in section 1 is not limited to estimation of the standard error of a statistic $\hat{\theta} = s(\mathbf{x})$; it can be used with only minor modification to provide an estimate of bias. The bias of an estimator $\hat{\theta}$ is the difference between the expectation of $\hat{\theta}$ and the parameter θ that it purports to estimate; formally,

$$bias = E_F(\hat{\theta}) - \theta. \tag{10}$$

An unbiased estimator (i.e., a statistic $\hat{\theta}$ satisfying $bias = E_F(\hat{\theta}) - \theta = 0$) is desirable inasmuch as it provides a measure of assurance that the estimator is not producing values that are either systematically too small or too large.

The fundamental problem remains unchanged. Only a single random sample $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ from an unspecified distribution F is available. Following the same rationale used in assessing the standard error, the ideal bootstrap estimate of bias becomes

$$\widehat{bias} = E_{\hat{F}}(\hat{\theta}^\star) - \hat{\theta}, \tag{11}$$

where, as before, $\hat{F}$ is the empirical distribution of F, $\hat{\theta}^\star$ is the bootstrap estimate of θ, and $\hat{\theta}$ is the estimate of θ based on the random sample $\mathbf{x}$. The same computational burden attendant to the evaluation of $\sigma_{\hat{F}}(\hat{\theta}^\star)$ in section 1 is now imposed on the determination of the expectation $E_{\hat{F}}(\hat{\theta}^\star)$, and is addressed in much the same manner. Namely, the further approximation

$$est\ \widehat{bias} = \frac{1}{b}\sum_{i=1}^{b}\hat{\theta}_i^\star - \hat{\theta} \tag{12}$$

is invoked. To calculate $est\ \widehat{bias}$, it is only necessary to incorporate expression (12) at the final step of the bootstrap algorithm, and, if desired, $est\ \widehat{bias}$ and $est\ \sigma_{\hat{F}}$ can be determined simultaneously.

The bias of $\hat{R}$ based on b = 300 bootstrap replications from the impact data in Table 1 was determined to be $est\ \widehat{bias} = -5.55$.

Should we have much confidence in this estimate of bias?... No.

The sign of the resultant estimate, if not the magnitude, was never in doubt. This is an example in which the bootstrap fails.

To see why this is so, consider the estimate of bias (12). In that expression we have $\hat{\theta} = R(\mathbf{x})$, the extreme spread of the basic random sample $\mathbf{x}$. Clearly, no bootstrap sample can lead to a $\hat{\theta}^\star$ larger than $\hat{\theta}$. It will, however, produce estimates $\hat{\theta}^\star$ less than $\hat{\theta}$ whenever a pair of impact coordinates maximizing expression (9) fail to appear in the bootstrap sample $\mathbf{x}^\star$. Thus the bootstrap estimate of bias will always be negative, except in the rare event when a maximizing pair appears in *every* bootstrap sample.

The situation is closely related to the well-known problem of estimating the range of a uniform variate $U[0, \theta]$ from a random sample $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, in which the maximum likelihood estimate is $\hat{\theta} = x_{[n]}$, the largest order statistic.

A reality of the bootstrap is that the procedure is held captive in the convex hull of the original sample $\mathbf{x} = (x_1, x_2, \ldots, x_n)$. A modification known as the smoothed bootstrap has been advanced, in which a scaled random variate with small dispersion is added to each observation in the original data set, but the result is essentially a perturbed Dirac-comb density estimator.

This is not a contrived example. Instances in which Dirac-comb density estimator based techniques (viz., the bootstrap) produce unsatisfactory results are not difficult to find. Consider the situation where several rounds of ammunition impact nearby, but never directly on, an aim point. If the distribution of shots is assumed circular normal with mean located at the aim point—the most common assumption in ballistic analyses—then a neighborhood of the aim point will appear, with high probability, to be a particularly safe place to reside. Anomalies such as these gave impetus to an algorithm advanced by Taylor and Thompson (1982, 1986), which is discussed in section 5.

4. Bootstrap Confidence Intervals

Application of the bootstrap procedure to confidence interval construction has been a major research focus, and a substantial number of procedures toward this end have been advanced. In this section a single approach, the *percentile method*, for constructing approximate confidence intervals in small sample nonparametric situations is considered. The intent is not to suggest that this is in some sense a "best" method; it is rather a widely used approach whose motivation is reasonably accessible.

Under quite general conditions, the distribution of an estimate $\hat{\theta}$ of a parameter θ approaches a normal distribution as the sample size $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ increases without bound. This asymptotic result leads to the approximation

$$\hat{\theta} \sim N(\theta, \sigma^2) \tag{13}$$

from which a $100(1-2\alpha)$ confidence interval for θ,

$$[\hat{\theta} - z^{(1-\alpha)} \cdot \sigma, \hat{\theta} - z^{(\alpha)} \cdot \sigma] \tag{14}$$

commonly expressed as

$$\hat{\theta} \pm z^{(1-\alpha)} \cdot \sigma \tag{15}$$

is derived. This is sometimes referred to as the standard confidence interval for θ; $z^{(1-\alpha)}$ is the $100(1-\alpha)$th percentile of the standard normal distribution.

Now consider as an estimate of θ, the bootstrap estimate, $\hat{\theta}^\star$. Appealing to the same argument, we are led to the approximation

$$\hat{\theta}^\star \sim N(\hat{\theta}, \hat{\sigma}^2), \tag{16}$$

where $\hat{\theta}$ is determined from $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and $\hat{\sigma}^2$ from any appropriate estimation procedure (e.g., the bootstrap, jackknife, or historical evidence).

Proceeding formally,

$$\begin{aligned} Pr[z^{(\alpha)} \leq \frac{\hat{\theta}^\star - \hat{\theta}}{\hat{\sigma}} \leq z^{(1-\alpha)}] &= \\ Pr[z^{(\alpha)} \cdot \hat{\sigma} \leq \hat{\theta}^\star - \hat{\theta} \leq z^{(1-\alpha)} \cdot \hat{\sigma}) &= \\ Pr[\hat{\theta} + z^{(\alpha)} \cdot \hat{\sigma} \leq \hat{\theta}^\star \leq \hat{\theta} + z^{(1-\alpha)} \cdot \hat{\sigma}] &= 1 - 2\alpha. \end{aligned} \tag{17}$$

Notice that the lower and upper confidence limits for θ appearing in (14) coincide with the 100αth and $100(1-\alpha)$th percentiles of the distribution of $\hat{\theta}^\star$ (since $z^{(\alpha)} \equiv -z^{(1-\alpha)}$). This observation is important because, when the distribution of $\hat{\theta}^\star$ is approximately normal, one might entertain as an approximate $1-2\alpha$ central confidence interval for θ the interval $[\hat{\theta}^{*(\alpha)}, \hat{\theta}^{*(1-\alpha)}]$—the 100αth and $100(1-\alpha)$th percentiles of $\hat{\theta}^\star$.

And this is precisely what the percentile method does.

Further investigation, beyond the scope of this exposition, suggests that this procedure, or a modification of this procedure, may provide acceptable results in those situations for which there exists a transformation to normality for the estimator $\hat{\theta}^\star$. In that eventuality, the percentile method exploits this circumstance automatically, and without the necessity of formally determining the transformation.

Consider the following example.

Example 2. An historic approach to vulnerability assessment of military systems commences with the overlay of an m×n rectangular grid as illustrated in Figure 4. For each cell in the grid, a detailed computer model estimates the probability of incapacitation of the system that would result from an impact in that location under prescribed ballistic conditions. These conditional probabilities, commonly referred to as probability-of-kill, will be denoted by $\hat{P}_{k|h_i}$, $i = 1, 2, \ldots, mn$. The cell-level $\hat{P}_{k|h_i}$s are combined under a weighting scheme over the entire grid to produce a global estimate of incapacitation, $\hat{P}_k$.

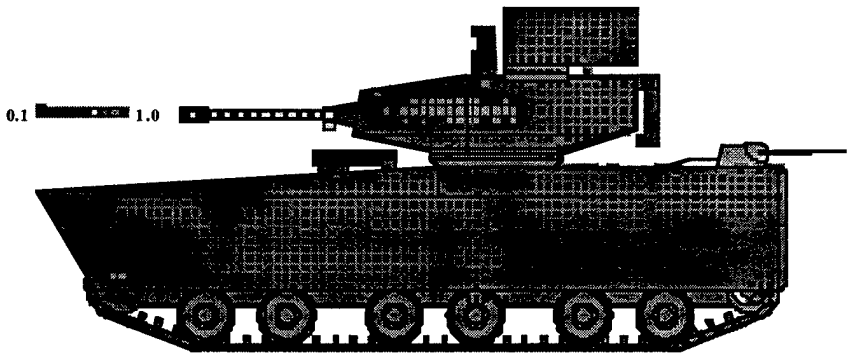

Figure 4. Armored vehicle with an overlaid rectangular grid.

Taylor and Bodt (1984) constructed approximate confidence intervals for P_k, using the percentile method, for several choices of grid size. Coarse grid templates were chosen, but with the finest level of detail retained, to produce within-cell variation. A set of $\hat{P}_{k|h_i}$ values was obtained, one value

from each grid cell, to form the base sample from which bootstrap samples would be drawn. Each confidence interval, $[\hat{P}_k^{\star(\alpha)}, \hat{P}_k^{\star(1-\alpha)}]$, was based on b = 1,000 bootstrap estimates of P_k.

This approach differs from the standard bootstrap procedure outlined in section 1 in that a stratified sample, rather than a simple random sample, was collected. The modification is appropriate for this particular application, and, since $\hat{P}_k$ is essentially a weighted sum, the Central Limit Theorem supports the anticipation of an approximate normal distribution for the bootstrap estimates $\hat{P}_k^{\star}$.

A total of 100 confidence intervals were constructed for each grid size by returning to the coarsened grid 100 times to establish a new stratified base sample, and then repeating the entire bootstrap procedure. This exercise is again outside the usual bootstrap venue; the intent was to see if one hundred $(1-2\alpha)$-level confidence intervals covered the parameter $\hat{P}_k$ approximately $100(1-2\alpha)$ times. Empirical coverage results are shown in Table 3 and the mean interval width in Table 4.

Table 3. Number of bootstrap confidence intervals covering the finest-level parameter estimate, $\hat{P}_k$ (100 confidence intervals were constructed for each table entry)

grid	.99	.95	.90	.80
24×60	100	100	100	100
16×40	100	100	100	99
12×30	100	100	100	100
8×20	100	99	99	95

Table 4. Mean width of the confidence intervals corresponding to the entries in Table 3

grid	.99	.95	.90	.80
24×60	.056	.042	.035	.027
16×40	.084	.063	.053	.041
12×30	.112	.084	.071	.055
8×20	.167	.127	.106	.083

Here the percentile method appears overly conservative. Efron and Tibshirani (1993) suggest modifications intended to increase its efficacy. DiCiccio and Romano (1988), Hall (1988), and Loh (1991) provide a more complete discussion of confidence interval construction.

5. A Generalized Bootstrap

The nonparametric bootstrap is not concerned with the construction of a nonparametric density estimator beyond the rudimentary Dirac-comb: its focus is on the generation of values that appear to conform to the same structure as that underlying the authentic data set. Given the ubiquity of fast computing, consideration of resampling schemes based on estimators more complex than the Dirac-comb appears warranted, and that is the focus of this section. A resampling plan that is useful for simulating data (SimDat) from an unspecified multivariate distribution is developed. The procedure is shown to behave in some respects like a normal kernel estimator; i.e., the density may be approximated by

$$\hat{f}_\delta(X) = \frac{1}{n}\sum_{i=1}^{n} K(X - X_i, \Sigma_i), \tag{18}$$

where $K(\cdot)$ is a normal distribution centered at zero with a locally estimated covariance matrix Σ_i. The bootstrap procedure arises as a limiting case, and, in that sense, SimDat may be considered a generalized bootstrap.

The algorithm operates as follows.

Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ be n observations of a p-component random vector X from a population with an unspecified distribution F. Suppose that a pseudo-sample of size N, where n and N are not necessarily equal, is to be generated. The data are rescaled so that the marginal sample variances coincide and, for each of the n observations, a specified number (m-1) of nearest neighbors is determined. The nearest neighbors are retained in an n×(m-1) array for use as a reference set.

To commence, a single point is selected at random from the data set $\mathbf{X}$ and its m-1 nearest neighbors recalled from the reference array. The mean of the resultant set of m points,

$$\bar{X} = \frac{1}{m}\sum_{i=1}^{m} X_i, \tag{19}$$

is determined, and each of the m points coded about $\bar{X}$ to obtain

$$X_i' = X_i - \bar{X}, \quad i = 1, 2, \ldots, m. \tag{20}$$

This calculation, like the determination of the m-1 nearest neighbors for each data point, will be required only once. The X_i' values along with $\bar{X}$ will be retained in an n×(m+1) array.

Next, a random sample $u_1, u_2, \ldots, u_m$ is generated from the uniform distribution

$$U\left(\frac{1}{m} - \sqrt{\frac{3(m-1)}{m^2}}, \frac{1}{m} + \sqrt{\frac{3(m-1)}{m^2}}\right) \tag{21}$$

and the linear combination

$$X' = \sum_{i=1}^{m} u_i X_i' \tag{22}$$

formed. Finally, with the addition of $\bar{X}$, a pseudo-data-point X is produced:

$$X = X' + \bar{X}. \tag{23}$$

And the process repeats.

A second point is selected at random from $\mathbf{X}$ and *its* m-1 nearest neighbors recalled; continuing until at last N pseudo-values have been determined.

With increasing m and n, the procedure behaves very much like the normal kernel approach mentioned earlier. To see why this is so, consider the sampled vector X_i and its m-1 nearest neighbors:

$$X_i = \begin{bmatrix} x_{1i} \\ x_{2i} \\ \vdots \\ x_{pi} \end{bmatrix}, \quad i = 1, 2, \ldots, m. \tag{24}$$

For the moment, consider this collection of m points as coming from a truncated distribution with mean vector μ and covariance matrix Σ. For the random sample $u_1, u_2, \ldots, u_m$ from the uniform distribution (21), we have

$$E(u_i) = \frac{1}{m}, \quad Var(u_i) = \frac{m-1}{m^2}, \tag{25}$$

and

$$Cov(u_i, u_j) = 0, \text{ for } i \neq j. \tag{26}$$

In the linear combination

$$Z = \sum_{i=1}^{m} u_i X_i, \tag{27}$$

the rth component of the vector Z, $z_r = u_1 x_{r1} + u_2 x_{r2} + \cdots + u_m x_{rm}$, satisfies

$$E(z_r) = \mu_r, \tag{28}$$

$$Var(z_r) = \sigma_r^2 + \frac{m-1}{m}\mu_r^2, \tag{29}$$

and

$$Cov(z_r, z_s) = \sigma_{rs} + \frac{m-1}{m}\mu_r\mu_s. \tag{30}$$

If the mean vector of the random variable X was $\mathbf{0} = (0, 0, \ldots, 0)'$, then the mean vector and covariance matrix of Z would be the same as that of X: $E(z_r) = 0$, $Var(z_r) = \sigma_r^2$, and $Cov(z_r, z_s) = \sigma_{rs}$. By translation to the local sample mean of the nearest-neighbor cloud—the motivation behind expression (23)—we do not achieve this result exactly, but we do come very close to generating an observation from the truncated distribution that models the points in the nearest-neighbor cloud. For moderately large m, the Central Limit Theorem becomes dominant, and SimDat comes close to sampling from n normal distributions with mean and covariance matrices corresponding to those of the n m-nearest-neighbor clouds.

Sufficient conditions for consistency of the nonparametric density estimator corresponding to SimDat is problematic. Mack and Rosenblatt [10] advance a formula for nearest-neighbor nonparametric density estimators:

$$m = Cn^{4/(p+4)}, \tag{31}$$

which seems promising, were it not for the fact that the parameter C is usually not available.

At the extremes, a choice of m=1 causes SimDat to collapse to the bootstrap; m=n produces an estimator that roughly samples from a multivariate normal distribution with the mean vector and covariance matrix computed from the original data set.

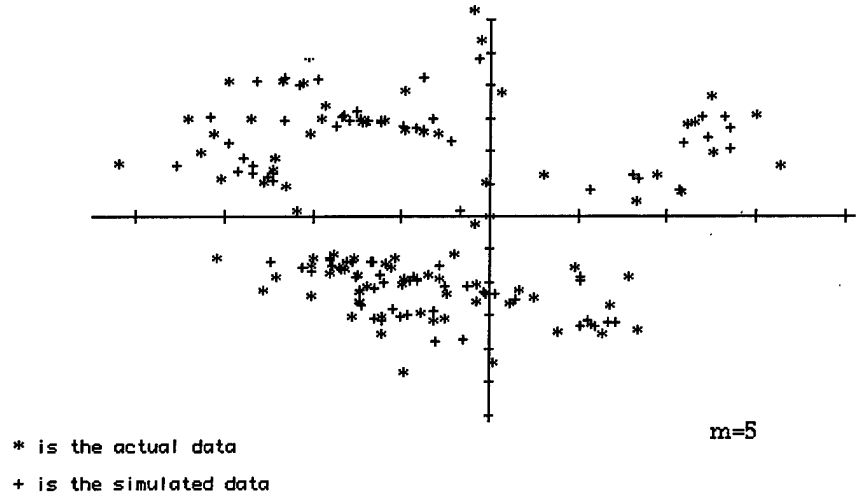

Figure 5. Samples from a normal mixture: $1/2N_1 + 1/3N_2 + 1/6N_3$ with m=5.

Example 3. In Figure 5, a sample of 85 values from a mixture of 3 normal distributions, $N_i(\mu_i, \Sigma_i)$, $i = 1, 2, 3$, with mean and covariance matrices

$$\mu_1 = \begin{pmatrix} -1 \\ -2 \end{pmatrix}, \quad \Sigma_1 = \begin{pmatrix} 1 & 1/2 \\ -1/2 & 1 \end{pmatrix}, \tag{32}$$

$$\mu_2 = \begin{pmatrix} -2 \\ 3 \end{pmatrix}, \quad \Sigma_2 = \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix}, \tag{33}$$

and

$$\mu_3 = \begin{pmatrix} 2 \\ 3/2 \end{pmatrix}, \quad \Sigma_3 = \begin{pmatrix} 1 & 1/10 \\ 1/10 & 1 \end{pmatrix}, \tag{34}$$

along with a pseudo-data-set of size 85 generated by SimDat, with m=5, is presented. The emulation of the data appears satisfactory, at least visually. In Figure 6, the same exercise, but with m=15, is shown. There, effects of a modest oversmoothing appear to be present.

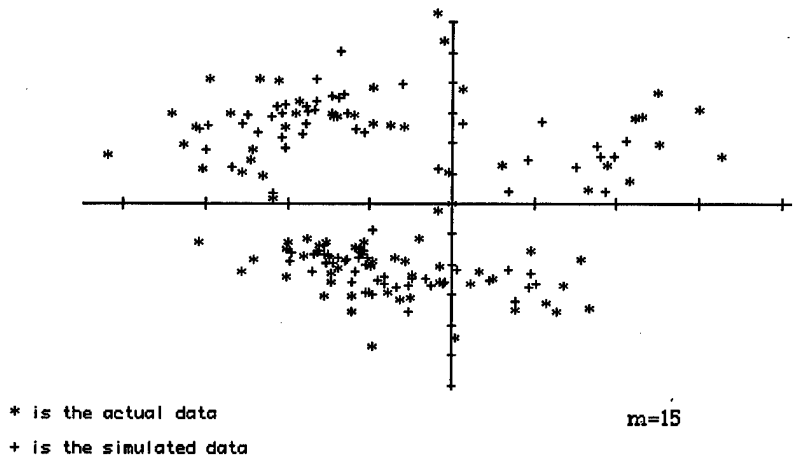

Figure 6. Samples from a normal mixture: $1/2N_1 + 1/3N_2 + 1/6N_3$ with m=15.

Example 4. The SimDat algorithm was applied to the impact data from Table 1.† Three hundred pseudo-data-sets of size 10 were generated, with a value of m=3, and their extreme spreads determined. A histogram of the results is shown in Figure 7.

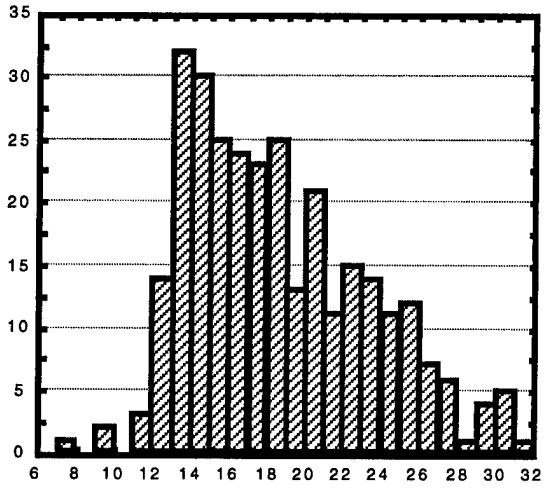

Figure 7. SimDat algorithm applied to the impact data, with m=3.

Empirical evidence suggests that if the data set is very large, say of size 1,000 or greater, good results are generally obtained with m≈.02n. For smaller values of n, m≈.05n often work well.

†A version of SimDat in the S language is available under the name "gendat" from the S Library; a Fortran version of SimDat entitled "RNDAT" is available from IMSL.

6. Summary

A comparison of the nonparametric bootstrap, parametric bootstrap, and SimDat procedure, all applied to estimation of extreme spread from the impact data presented in Table 1, is shown in Table 5. This application proves vexing to any approach, as the tabled values suggest. The nonparametric bootstrap applied to the impact data has been extensively discussed. It is worth remembering that the extreme spread of the original data set, $\hat{\theta} = 30.21$ (appearing as the maximum value for the nonparametric bootstrap) is the sole authentic estimate. The parametric bootstrap has the smallest estimate of bias, a direct consequence of the strong distribution assumption. Notice the maximum value of 54.64—far in excess of 30.21. SimDat has the smallest estimated standard error and the largest bias. It generated some extreme spreads larger than the authentic $\hat{\theta}$—notice the maximum, 31.18—but most pseudo-points are aggregated in the regions of highest population density by virtue of the randomly selected center points that commence the algorithm. In other words, SimDat is providing a modest amount of smoothing under control of the parameter m, as it was intended to do.

Table 5. Descriptive statistics for the nonparametric(n) and parametric(p) bootstrap and the SimDat procedure. Estimates of standard error, bias, and maximum and minimum values of the extreme spread are for 300 replications.

method	std err est	bias est	min	max
bootstrap(n)	6.06	-5.55	9.82	30.21
bootstrap(p)	6.71	-2.60	13.18	54.64
SimDat	4.71	-11.56	7.78	31.18

And finally, a note of caution. Both the nonparametric bootstrap and SimDat are model-free techniques for examining data. The nonparametric techniques use the power of the computer to circumvent detailed modeling, and such an approach is highly useful as an interpolative device. When the dimensionality of a data set becomes high–say, five or greater, these approaches should be used with caution, since a number of widely separated modes, with data-scarce regions in between, might easily be obscured. Dealing with such data sets nonparametrically and away from the modes becomes more of an extrapolation problem, and blindly using smoothed interpolation routines can be a precursor to faulty inference.

7. References

DiCiccio, T. J., and J. P. Romano. "A Review of Bootstrap Confidence Intervals" (with Discussion). *Journal of the Royal Statistical Society*, Series B, vol. 50, pp. 338–354, 1988.

Efron, B. *The Jackknife, the Bootstrap and Other Resampling Plans.* CBMS-NSF Regional Conference Series in Applied Mathematics, no. 38, Society for Industrial and Applied Mathematics, 1982.

Efron, B., and R. J. Tibshirani. *An Introduction to the Bootstrap.* Monographs on Statistics and Applied Probability, no. 57, New York: Chapman Hall, 1993.

Grubbs, F. E. *Statistical Measures of Accuracy for Riflemen and Missile Engineers.* Ann Arbor: Edwards Brothers, 1964.

Guerra, V., R. A. Tapia, and J. R. Thompson. "A Random Number Generator for Continuous Random Variables Based on an Interpolation Procedure of Akima's." *Computer Science and Statistics*, pp. 228–230, 1976.

Hall, P. "Theoretical Comparison of Bootstrap Confidence Intervals" (with Discussion). *Annals of Statistics*, vol. 16, pp. 927–985, 1988.

Kiefer, J., and J. Wolfowitz. "Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters." *Annals of Mathematical Statistics*, vol. 27, pp. 887–906, 1956.

LePage, R., and L. Billard. *Exploring the Limits of Bootstrap.* New York: Wiley, 1992.

Loh, W-Y. "Bootstrap Calibration for Confidence Interval Construction and Selection." *Statistica Sinica*, vol. 1, pp. 477–491, 1991.

Mack, Y. P., and M. Rosenblatt. "Multivariate k-Nearest-Neighbor Density Estimates." *Journal of Multivariate Analysis*, vol. 9, pp. 1–15, 1979.

Taylor, M. S., and B. A. Bodt. "Construction of Approximate Confidence Intervals for Probability-of-Kill via the Bootstrap." ARBRL-TR-0250, U.S. Army Ballistic Research Laboratory, Aberdeen Proving Ground, MD, 1984.

Taylor, M. S., and F. E. Grubbs. "Approximate Probability Distributions for the Extreme Spread." *Naval Research Logistics Quarterly*, vol. 22, no. 4, pp. 713–719, 1975.

Taylor, M. S., and J. R. Thompson. "A Data Based Random Number Generator for a Multivariate Distribution." *Proceedings of the NASA Workshop on Density Estimation and Function Smoothing*, pp. 214–225, 1982.

Taylor, M. S., and J. R. Thompson. "A Data Based Algorithm for the Generation of Random Vectors, *Computational Statistics and Data Analysis*, vol. 4, pp. 93–101, 1986.

INTENTIONALLY LEFT BLANK.

NO. OF
COPIES ORGANIZATION

2 DEFENSE TECHNICAL
INFORMATION CENTER
DTIC DDA
8725 JOHN J KINGMAN RD
STE 0944
FT BELVOIR VA 22060-6218

1 HQDA
DAMO FDQ
DENNIS SCHMIDT
400 ARMY PENTAGON
WASHINGTON DC 20310-0460

1 OSD
OUSD(A&T)/ODDDR&E(R)
R J TREW
THE PENTAGON
WASHINGTON DC 20301-7100

1 CECOM
SP & TRRSTRL COMMCTN DIV
AMSEL RD ST MC M
H SOICHER
FT MONMOUTH NJ 07703-5203

1 PRIN DPTY FOR TCHNLGY HQ
US ARMY MATCOM
AMCDCG T
M FISETTE
5001 EISENHOWER AVE
ALEXANDRIA VA 22333-0001

1 DPTY CG FOR RDE HQ
US ARMY MATCOM
AMCRD
MG BEAUCHAMP
5001 EISENHOWER AVE
ALEXANDRIA VA 22333-0001

1 INST FOR ADVNCD TCHNLGY
THE UNIV OF TEXAS AT AUSTIN
PO BOX 202797
AUSTIN TX 78720-2797

1 GPS JOINT PROG OFC DIR
COL J CLAY
2435 VELA WAY STE 1613
LOS ANGELES AFB CA 90245-5500

NO. OF
COPIES ORGANIZATION

1 DARPA
B KASPAR
3701 N FAIRFAX DR
ARLINGTON VA 22203-1714

1 NAVAL SURFACE WARFARE CTR
CODE B07 J PENNELLA
17320 DAHLGREN RD
BLDG 1470 RM 1101
DAHLGREN VA 22448-5100

1 US MILITARY ACADEMY
MATH SCI CTR OF EXCELLENCE
DEPT OF MATHEMATICAL SCI
MDN A MAJ DON ENGEN
THAYER HALL
WEST POINT NY 10996-1786

1 DIRECTOR
US ARMY RESEARCH LAB
AMSRL CS AL TA
2800 POWDER MILL RD
ADELPHI MD 20783-1145

3 DIRECTOR
US ARMY RESEARCH LAB
AMSRL CI LL
2800 POWDER MILL RD
ADELPHI MD 20783-1145

ABERDEEN PROVING GROUND

4 DIR USARL
AMSRL CI LP (305)

NO. OF
COPIES ORGANIZATION

1 BARNARD BISSINGER
281 WEST MAIN ST
MIDDLETOWN PA 17057

ABERDEEN PROVING GROUND

32 DIR USARL
AMSRL-IS,
J GANTT (ALC)
J CHANDRA (ALC)
AMSRL-IS-C,
A BRODEEN
M TAYLOR (20)
AMSRL-IS-TP, F BRUNDICK
AMSRL-IS-CI,
A CELMINS
J DUMER
T HANRATTY
AMSRL-SL-BE, W BAKER
AMSRL-SL-BG, T MUEHL
AMSRL-HR-S, J GRYNOVICKI
AMSRL-CI-HC, C ZOLTANI
AMSRL-WM-BC, D WEBB

1 DIR USATC
STEAC-PC T WALKER

2 DIR USAMSAA
AMXSY-CA
F WOFFORD
W YEAKEL

2 DIR USAARDEC
AMSTA-AR-FSF-T
BLDG 120
M ANDRIOLO
B LAURIE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	September 1998	Final, Mar 97- Dec 97

4. TITLE AND SUBTITLE

Statistical Resampling Plans

5. FUNDING NUMBERS

1L162618AH80

6. AUTHOR(S)

Malcolm S. Taylor

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

U.S. Army Research Laboratory
ATTN: AMSRL-IS-C
Aberdeen Proving Ground, MD 21005-5067

8. PERFORMING ORGANIZATION REPORT NUMBER

ARL-TR-1794

9. SPONSORING/MONITORING AGENCY NAMES(S) AND ADDRESS(ES)

10.SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Statistical resampling plans can be thought of as procedures for "recycling data." They commonly involve computer-intensive methods for the interrogation of data through repeated inspection. The procedures detailed in this document free the data analyst from two major restrictions underlying much of the development in mathematical statistics: the normal or Gaussian assumption, and the necessity to focus on statistics that are in closed form and mathematically tractable.

14. SUBJECT TERMS

data analysis, nonparametric statistics, density estimation, resampling, bootstrap

15. NUMBER OF PAGES

27

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

INTENTIONALLY LEFT BLANK.

USER EVALUATION SHEET/CHANGE OF ADDRESS

This Laboratory undertakes a continuing effort to improve the quality of the reports it publishes. Your comments/answers to the items/questions below will aid us in our efforts.

1. ARL Report Number/Author ARL-TR-1794 (Taylor) Date of Report September 1998

2. Date Report Received _____

3. Does this report satisfy a need? (Comment on purpose, related project, or other area of interest for which the report will be used.) _____

4. Specifically, how is the report being used? (Information source, design data, procedure, source of ideas, etc.) _____

5. Has the information in this report led to any quantitative savings as far as man-hours or dollars saved, operating costs avoided, or efficiencies achieved, etc? If so, please elaborate. _____

6. General Comments. What do you think should be changed to improve future reports? (Indicate changes to organization, technical content, format, etc.) _____

CURRENT ADDRESS

_____
Organization

_____
Name E-mail Name

_____
Street or P.O. Box No.

_____
City, State, Zip Code

7. If indicating a Change of Address or Address Correction, please provide the Current or Correct address above and the Old or Incorrect address below.

OLD ADDRESS

_____
Organization

_____
Name

_____
Street or P.O. Box No.

_____
City, State, Zip Code

(Remove this sheet, fold as indicated, tape closed, and mail.)
(DO NOT STAPLE)

DEPARTMENT OF THE ARMY

OFFICIAL BUSINESS

NO POSTAGE
NECESSARY
IF MAILED
IN THE
UNITED STATES

BUSINESS REPLY MAIL
FIRST CLASS PERMIT NO 0001,APG,MD

POSTAGE WILL BE PAID BY ADDRESSEE

DIRECTOR
US ARMY RESEARCH LABORATORY
ATTN AMSRL IS C
ABERDEEN PROVING GROUND MD 21005-5067