1980

On Increasing Confidence in Confidence Intervals

Peter J. Denning

Report Number:
80-324

ON INCREASING CONFIDENCE IN CONFIDENCE INTERVALS

Peter J. Denning

Computer Science Department
Purdue University
W. Lafayette, IN 47907

CSD-TR-324

January 1980

Abstract: A set of N data elements $x_1,\ldots,x_N$ has r^{th} moment $E(x^r) = (x_1^r + \cdots + x_N^r)/N$ and variance of the r^{th} moment $Var(x^r) = E(x^{2r}) - E^2(x^r)$. The r^{th} moment of the data elements in an arbitrary subset of k elements is used to estimate $E(x^r)$. Over all the choices of the sample, the mean error is zero and the mean squared error $\frac{N-k}{k(N-1)} Var(x^r)$. A little-known theorem by Madow shows that the frequency distribution of values of the estimator is approximately normal with mean $E(x^r)$ and variance $Var(x^r)$. All these results are proved without assuming statistical independence among sampled data elements. The conclusion is that confidence-interval calculators based on the normal distribution actually applies in more cases than is commonly believed.

Work reported herein supported in part by NSF Grant MCS78-01729 at Purdue University.

1. INTRODUCTION

Two kinds of error arise when an analytic model is used to predict the future behavior of a queueing system:

1. Estimation Errors: Differences between the forecasted and actual values of parameters used by the model; and
2. Modeling Errors: Differences between modeling assumptions and the real system, leading to discrepancies even if exact values of parameters are known.

Estimation theory, a branch of stochastic analysis, provides tools for quantifying the first type of error; the most familiar of these tools are unbiased estimators and confidence intervals. Because stochastic modeling assumptions typically introduce unmeasurable abstractions, the second type of error cannot be quantified; experimental validation is often the only tool available for gaining an understanding of modeling errors.

One way to deal with the second type of error is to weaken the modeling assumptions so that the models apply more widely. If the weakened assumptions can be experimentally tested, it becomes possible to quantify the errors between actual and calculated performance metrics in terms of errors between actual values of parameters and the values induced by modeling assumptions. This approach is taken by operational analysis ([1],[2],[3], and [4]). The analysis of a single queue, for example, is greatly simplified by the assumption of "homogeneous arrivals", which asserts that the arrival rate is a constant independent of queue length. It is possible to express the potential error of the model in terms of potential errors between the actual (queue-dependent) arrival rate and the constant assumed by homogeneous arrivals ([5],[6]). In contrast, it is not possible to quantify the error caused by potential violations of the Poisson arrival assumption in the analysis of an M/G/1 stochastic queue.

Many derivations of Estimation Errors are based on stochastic assumptions that make them subject to Modeling Errors as well. For example, in analyzing data from real systems [7] or from simulation models of systems [8], it is common to assume that the data samples are independent, identically-distributed (i.i.d.) random variables. This assumption leads to simple derivations of unbiased estimators for mean and variance, and also, through the central limit theorem, to confidence intervals containing the stochastic quantity being estimated. Interpreting such results is not always as easy as deriving them: the samples may not be i.i.d. or the real system may not satisfy all the assumptions of the stochastic model. One practical way to increase confidence in "confidence intervals" is to weaken the assumptions under which they are derived.

The purpose of this paper is to call attention to some results well known to sample-survey statisticians, but overlooked by many computer performance analysts. These results show that the i.i.d. assumption is unnecessary for many real sampling problems. Familiar sampling methods and the central limit theorem hold in many cases where the conventional i.i.d. assumptions cannot be justified. Operational assumptions lead to greater confidence in confidence intervals.

2. STOCHASTIC ANALYSIS OF DATA SAMPLING

The familiar problem statement is this: A series of k observations of a variable yields values (samples) $x_1,\ldots,x_k$. It is assumed that the series can be extended indefinitely (k can be increased without bound) and that the samples are i.i.d. random variables. The objective is to estimate the value of the r^{th} moment of the common, underlying distribution.

The attraction of the i.i.d. assumption is the simplicity of the ensuing mathematics. The difficulty is that one often cannot tell whether the assumption is justified. For example, the successive samples x_i of the CPU's queue length may be statistically dependent because the CPU speed is comparable to the sampling rate. Regenerative simulation attempts to avoid this difficulty by choosing samples from successive intervals between returns to a predetermined system state [8]; however, the user of the simulation does not know whether the returns to the given state define a renewal process in the real system.

Given that the samples $x_1,\ldots,x_k$ are i.i.d., one proceeds as follows to estimate the r^{th} moment of the underlying distribution and to characterize the accuracy of the estimate. The estimate of the r^{th} moment is

$$(2.1) \qquad m_r = \frac{1}{k} \sum_{i=1}^{k} x_i^{\,r}$$

The statistic m_r is itself a random variable (it depends on the choice of the k elements constituting the sample). It is unbiased because its mean is $E(x^r)$, the r^{th} moment of the distribution underlying the data. Its variance is

$$(2.2) \qquad \mathrm{Var}(m_r) = \frac{\mathrm{Var}(x^r)}{k} = \frac{E(x^{2r}) - E^2(x^r)}{k} .$$

Eq 2.2 is not directly useful because it expresses the variance in the estimate in terms of (unknown) moments of the (unknown) distribution underlying the data. Therefore the variance must also be estimated. The usual formula is

$$(2.3) \qquad V_r = \frac{1}{k-1} \sum_{i=1}^{k} (x_i^{\,r} - m_r)^2 .$$

This estimator is also a random variable. The divisor $1/(k-1)$ is chosen rather than $1/k$ so that this estimator is unbiased -- that is, $E(V_r) = \mathrm{Var}(x^r)$. Using Eq. 2.2, the analyst estimates the variance of m_r as V_r/k.

Because m_r is the sum of i.i.d. random variables, the central limit theorem implies that m_r is (approximately) Normal with mean $\mu = E(x^r)$ and standard deviation $\sigma = [\mathrm{Var}(x^r)/k]^{1/2}$. This allows one to associate a confidence interval with an estimate -- a range within which the estimate must fall with a given probability. For example, a Normal random variable falls within 2σ of its mean with probability 95%. Because $\mu - 2\sigma \leq m_r \leq \mu + 2\sigma$ if and only if $m_r - 2\sigma \leq \mu \leq m_r + 2\sigma$, one can state that "the true mean is within 2σ of m_r with probability 95%." If one estimates σ as V_r/k, then one can be "95% sure" that the observed value m_r is within 2σ of its true mean μ.

A typical proof of the central limit theorem is based on generating functions [9]:

$$\phi_x(u) = E(\exp(iux)) \tag{2.4}$$

where x is the random variable, u is a parameter of the generating function, and i is $\sqrt{-1}$. The generating function for the distribution of m_r is

$$(2.5) \qquad \phi_{m_r}(u) = E(\exp(\frac{iu}{k}\sum_{i=1}^{k} x_i))$$

$$= E(\prod_{i=1}^{k} \exp(\frac{iu}{k} x_i))$$

$$= \prod_{i=1}^{k} E(\exp(\frac{iux}{k}))$$

$$= [\phi_x(\frac{u}{k})]^k .$$

The key step in deriving the relation between $\phi_{m_r}(u)$ and $\phi_x(u)$ is the interchange of expectation and product in the third line; this interchange depends crucially on the i.i.d. assumption. The rest of the proof is based on showing that, for large k,

$$(2.6) \qquad [\phi_x(\frac{u}{k})]^k \cong \exp(-(\frac{u-\mu}{\sigma})^2/2)$$

which is the generating function for the Normal distribution with mean $\mu = E(x^r)$ and variance $\sigma^2 = Var(x^r)/k$.

The central point is that the familiar derivations for the variance of m_r and for confidence intervals rest heavily on the i.i.d. modeling assumption. If one is unsure whether the real data satisfy this assumption, one is also unsure how much confidence to have in statements about the accuracy of the estimates.

3. OPERATIONAL ANALYSIS OF DATA SAMPLING

The operational (or "phenomemological" [10]) view of the estimation problem deals only with measurable quantities. Rather than attempt to estimate the values of unobservable parameters, we focus instead on the accuracy with which we can infer the moments of the empirical distribution of all the data, given only a sample of k of the elements. By replacing the abstract "underlying distribution" with the concrete "empirical distribution of all the data", we can eliminate the concept of i.i.d. random variables and the difficulties of interpretation that attend it.

3.1 Estimating Moments

Suppose that $x_1, \ldots, x_N$ are the data values associated with a given population of N individuals. The r^{th} moment of all the data is defined to be

$$E(x^r) = \frac{1}{N} \sum_{i=1}^{N} x_i^{\,r}. \tag{3.1}$$

Although $E(x^r)$ is by this definition precisely measureable, it may in practice be too expensive to measure the whole population. Therefore, we try to estimate $E(x^r)$ from an easily-measured sample.

Let A denote a sample, an arbitrary subset of k of the indices $1, \ldots, N$; there are $\binom{N}{k}$ distinct samples. The estimate of the r^{th} moment based on A is

$$(3.2) \qquad m_r(A) = \frac{1}{k} \sum_{i \epsilon A} x_i^{\,r}.$$

There is an error between this estimate and the true value:

$$(3.3) \qquad e_r(A) = E(x^r) - m_r(A).$$

The mean error over all the possible choices of the sample is

$$(3.4) \qquad E(e_r(A)) = \frac{1}{\binom{N}{k}} \sum_{\text{all} A} e_r(A) \ .$$

The mean squared error over all the choices of the sample is

$$(3.5) \qquad E(e_r^{\,2}(A)) = \frac{1}{\binom{N}{K}} \sum_{\text{all } A} e_r^{\,2}\,(A).$$

In Appendices 1 and 2 we show that expressions 3.4 and 3.5 reduce respectively to

$$(3.6) \qquad E(e_r(A)) = 0$$

$$E(e_r^{\,2}(A)) = \frac{N-k}{k(N-1)} \; Var(x^r) \qquad (3.7)$$

where $Var(x^r) = E(x^{2r}) - E^2(x^r)$. That the mean error (Eq. 3.6) is zero implies that the estimate $\hat{m}_r(A)$ is unbiased, as anticipated.

Formulae 3.6 and 3.7 are not new. Cochran [11] gives derivations like those in Appendices 1 and 2, which do not use any concept of independence among the x_i. Birnbaum [12] gives a simpler argument for the same formulae, but makes use of independence among the x_i. The important point is that, at the price of a slightly more complicated derivation, the estimate can be proved unbiased and its error related to the variance of all the data -- without assuming statistical independence of the data samples.

If N is very large and k is small compared to N, the factor (N-k)/(N-1) in formula 3.7 is approximately 1, and the mean squared error in the estimate has the same mathematical form as the variance of the stochastic m_r of Eq. 2.2. The factor (N-k)/(k-1) is sometimes called the "finite population connection" [11]. Formula 3.7 shows an operational sense in which variances are additive in the absence of an independence assumption about the (random) variables.

In a stochastic context, the concept of choosing the elements in set A is called sampling without replacement.

In this case, the various sample elements cannot be considered independent (for example, x_1 and x_2 cannot be from the same individual). In contrast, the i.i.d. concept is closely related to sampling with replacement [12]. It is often not easy to tell whether sampling is done with or without replacement. Suppose $x_1,\ldots,x_k$ are observed in the first k cycles of a regenerative simulation; shall we regard this as sample of k distinct elements from a longer sequence of distinct observations, or as k independent instances drawn with replacement from some underlying set of values? Operationally, ambiguities of this type do not arise because there is no "underlying" distribution.

Formula 3.7 suffers from a difficulty similar to 2.2, namely that the error in the estimate is expressed in terms of a quantity that is not known. (To obtain the exact value of $\mathrm{Var}(x^r)$, we would need to measure the entire population, in which case the estimate $m_r(A)$ is of no interest.) We show Appendix 3, using arguments like those in [11], that

$$(3.8) \qquad V_r(A) = \frac{N-1}{N(k-1)} \left(\sum_{i \in A} x_i^{2r} - k\, m_r^2(A) \right)$$

is an unbiased estimator for $\mathrm{Var}(x^r)$. The mean squared error in $m_r(A)$ can be estimated as

$$E(e_r^{\,2}(A)) = \frac{N-k}{k(N-1)} V_r(A), \tag{3.9}$$

by applying 3.7 with the variance estimator.

3.2 Estimating Confidence Intervals

To calculate confidence intervals operationally, we need to know the frequency distribution of $m_r(A)$ as the sample A is varied over all the $\binom{N}{k}$ distinct choices. The generating function approach noted earlier cannot be applied to this problem because we are not willing to introduce an i.i.d. assumption for the elements of the sample A.

In 1948, Madow published a remarkable theorem about samples drawn from finite universes [13]. When applied to the case at hand, Madow's theorem says that $m_r(A)$ will be approximately normally distributed (as A varies over all possibilities) with mean $E(x^r)$ and variance $E(e_r^{\,2}(A))$ provided that N is large, $k < N$, and that

$$\frac{\frac{1}{N} \sum_{i=1}^{N} (x_i^{\,r} - E(x^r))^t}{(\mathrm{Var}(x^r))^{t/2}} < c \tag{3.10}$$

for some constant $c > 0$ and all even integers $t > 0$. Condition (3.10) requires that the higher-order central moments of the data do not increase too rapidly as compared with the variance. The sampling ratio, k/N, need not be small -- as long as 3.10 holds, values of the estimate $m_r(A)$ will be normally distributed even though k/N be near 100%.

Madow's proof shows that the moments of the distribution of $m_r(A)$ approximate those of the Normal distribution arbitrarily closely for large enough N. He does this by algebraic manipulation of expressions for these moments; no assumption is made about whether the elements $x_1,\ldots,x_N$ are independent or not. Therefore, Madow's theorem shows that we may use the central limit theorems to estimate confidence intervals for operational data sampling even when the sampling ratio may be large.

Condition 3.10 usually fails only for populations over which $m_r(A)$ is essentially constant as A varies. An example is the population $(x_1, x_2,\ldots,x_N) = (1, 0,\ldots,0)$. In this case, $m_r(A)$ is either 0 or $1/k$, so the frequency histogram has only two points. In the case $r = 1$, the left side of 3.10 works out to be

$$\frac{\frac{1}{N}\left(\frac{N-1}{N}\right)^t + \frac{N-1}{N^{t+1}}}{\left(\frac{N-1}{N^2}\right)^{t/2}} = \frac{(N-1)^t + N-1}{N(N-1)^{t/2}}$$

which diverges at t becomes large.

Armed with Madow's theorem, one can calculate confidence intervals in the usual way. Given a sample A from the population, one estimates $m_r(A)$ from Eq. 3.2 and $E(e_r^2(A))$ from Eq. 3.9. The standard derivation is estimated as $\sigma = [E(e_r^2(A))]^{1/2}$. For the desired confidence level p, one finds a value b such that

$$p = \Phi(b) - \Phi(-b)$$

where $\Phi(.)$ is the cumulative distribution of the unit Normal. The confidence interval is $[m_r(A) - b\sigma, m_r(A) + b\sigma]$. We interpret this to mean that the true value of $E(x^r)$ is contained in the confidence interval for $p\binom{N}{k}$ of the possible choices of A.

4. CONCLUSION

We have studied the problem of estimating a moment of a set of data. Operationally, we calculate from a sample an estimate of a moment that could be calculated exactly from all the data, were measuring all the data feasible or desirable. The uncertainty in the operational estimate can be expressed as the mean squared error over all the distinct choices of the sample. This expression does not rely on any assumption of statistical independence among the sampled data elements, as does the corresponding expression in stochastic analysis. The uncertainty of the operational estimate can also be expressed in terms of confidence intervals calculated on the assumption that the estimator is distributed normally about the true value. Madow's theorem shows that the normal approximation for the distribution of estimator values has mean $E(x^r)$ and variance $(N-k)Var(x^r)/k(N-1)$. This approximation is valid under very general conditions; no assumption of statistical independence among sampled data elements is required. In fact, it is not relevant whether the data elements are independent.

These results help explain why assumptions like additivity of variance or normally distributed estimator values work well even when the i.i.d. assumptions cannot be justified. Familiar estimation techniques are more general than is commonly believed.

APPENDIX 1 -- The Mean Error

The mean error over all choices of the sample is (cf. 3.4):

$$E(e_r(A)) = \frac{1}{\binom{N}{k}} \sum_{\text{all } A} (E(x^r) - m_r(A))$$

$$= \frac{1}{\binom{N}{k}} \sum_{\text{all } A} E(x^r) - \frac{1}{\binom{N}{k}} \frac{1}{k} \sum_{\text{all } A} \sum_{i \epsilon A} x_i^{\,r}$$

The term $E(x^r)$ appears $\binom{N}{k}$ times in the first sum, which thereby evaluates to $E(x^r)$. Each index i appears in $\binom{N-1}{k-1}$ distinct choices of the sample because, given that iϵA, there are $\binom{N-1}{k-1}$ ways to choose the remainder of A. The second sum therefore reduces to

$$\frac{1}{\binom{N}{k}} \frac{1}{k} \binom{N-1}{k-1} \sum_{i=1}^{N} x_i^{\,r} = \frac{1}{N} \sum_{i=1}^{N} x_i^{\,r} = E(x^r),$$

where we used the identity,

$$\binom{N}{k} = \frac{N!}{(N-k)!\ k!} = \frac{N}{k} \frac{(N-1)!}{(N-k)!\ (k-1)!} = \frac{N}{k} \binom{N-1}{k-1} .$$

Therefore $E(e_r(A)) = E(x^r) - E(x^r) = 0$; $m_r(A)$ is an unbiased estimator of the true mean $E(x^r)$.

APPENDIX 2 -- The Mean Squared Error

The mean squared error over all choices of the sample is (cf. 3.5):

$$E(e_r^{\,2}(A)) = \frac{1}{\binom{N}{k}} \sum_{\text{all } A} (E(x^r) - m_r(A))^2 \;.$$

$$= \frac{1}{\binom{N}{k}} \sum_{\text{all } A} E^2(x^r) - 2E(x^r)\,\frac{1}{\binom{N}{k}} \sum_{\text{all } A} m_r(A) + \frac{1}{\binom{N}{k}} \sum_{\text{all } A} m_r^{\,2}(A)$$

The first sum reduces to $E^2(x^r)$. Since $m_r(A)$ is an unbiased estimator, the second sum reduces to $2E^2(x^r)$. Thus the mean squared error reduces to

$$\text{(i)} \quad E(e_r^{\,2}(A)) = \frac{1}{\binom{N}{k}} \sum_{\text{all } A} m_r^{\,2}(A) - E^2(x^r).$$

When $k = 1$ each sample A contains exactly one of the i. Thus $m_r^{\,2}(A) = x_i^{\,2r}$ for some i and the sum over "all A" reduces to $E(x^{2r})$, so that

$$\text{(ii)} \quad E(e_r^{\,2}(A)) = E(x^{2r}) - E^2(x^r) = \text{Var}(x^r), \qquad \text{for } k = 1.$$

To evaluate $E(e_r^{\,2}(A))$ for $1 < k \leq N$, we need to evaluate the sum

$$\text{(iii)} \quad \sum_{\text{all } A} \Big(\frac{1}{k} \sum_{i \in A} x_i^{\,r} \Big)^2 = \frac{1}{k^2} \sum_{\text{all } A} \Big(\sum_{i \in A} x_i^{\,r} \Big) \Big(\sum_{j \in A} x_j^{\,r} \Big)$$

$$= \frac{1}{k^2} \sum_{\text{all } A} \sum_{i \in A} x_i^{\,2r} + \frac{1}{k^2} \sum_{\text{all } A} \sum_{\substack{(i,j) \in A^2 \\ i \neq j}} x_i^{\,r} x_j^{\,r}$$

In the first sum, each index i appears in $\binom{N-1}{k-1} = \frac{N-1}{k-1}\binom{N-2}{k-1}$ distinct choices of A; this sum therefore reduces to

$$\frac{1}{k^2} \frac{N-1}{k-1} \binom{N-2}{k-2} \sum_{i=1}^{N} x_i^{\,2r} = \frac{N(N-1)}{k^2(k-1)} \binom{N-2}{k-2} E(x^{2r}).$$

In the second sum, each pair of distinct indices (i,j) appears in $\binom{N-2}{k-2}$ distinct choices of A; the second sum reduces to

$$\frac{1}{k^2} \binom{N-2}{k-2} \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} x_i^{\,r} x_j^{\,r} = \frac{1}{k^2} \binom{N-2}{k-2} \Big(\sum_{i=1}^{N} \sum_{j=1}^{N} x_i^{\,r} x_j^{\,r} - \sum_{i=1}^{N} x_i^{\,2r} \Big)$$

$$= \frac{1}{k^2} \binom{N-2}{k-2} \big(N^2 E^2(x^r) - N E(x^{2r}) \big)$$

Adding the results for the two sums, we find that the left side of (iii) reduces to

$$\text{(iv)} \quad \frac{1}{k^2} \binom{N-2}{k-2} \Big(E(x^{2r}) \frac{N(N-k)}{k-1} + N^2 E^2(x^r) \Big)$$

Dividing this by $\binom{N}{k} = \frac{N(N-1)}{k(k-1)}\binom{N-2}{k-2}$ and subtracting $E^2(x^r)$ as required in (i), we find.

$$\text{(v)} \qquad E(e_r^{\,2}(A)) = E(x^{2r})\left(\frac{N(N-k)}{k^2(k-1)}\;\frac{k(k-1)}{N(N-1)}\right) - \left(E^2(x^r)\;\frac{N^2}{k^2}\;\frac{k(k-1)}{N(N-1)} - 1\right)$$

$$= \frac{N-k}{k(N-1)}\left(E(x^{2r}) - E^2(x^r)\right)$$

$$= \frac{N-k}{k(N-1)}\;\mathrm{Var}(x^r),$$

which was to be shown. Note that this expression reduces to (ii) when $k = 1$.

APPENDIX 3 -- Estimating Variance

Eq. (3.8) states that an unbiased estimator for $Var(x^r)$ is

$$V_r(A) = \frac{N-1}{N(k-1)} \left(\sum_{i\epsilon A} x_i^{2r} - km_r^2(A)\right).$$

To demonstrate this we must reduce

$$\text{(i)} \qquad \frac{1}{\binom{N}{k}} \sum_{\text{all } A} V_r(A) = \frac{1}{\binom{N}{k}} \sum_{\text{all } A} \sum_{i\epsilon A} x_i^{2r} - \frac{1}{\binom{N}{k}} \frac{N-1}{N} \frac{k}{k-1} \sum_{\text{all } A} m_r^2(A)$$

to $Var(x^r)$. The index i appears in $\binom{N-1}{k-1}$ distinct choices of A in the first sum on the right; hence it reduces to

$$\text{(ii)} \qquad \frac{1}{\binom{N}{k}} \frac{N-1}{N(k-1)} \binom{N-1}{k-1} \sum_{i=1}^{N} x_i^{2r} = \frac{k}{k-1} \frac{N-1}{N} E(x^{2r}).$$

Eq. A.2(iv) shows that the second sum reduces to

$$\text{(iii)} \qquad \frac{1}{\binom{N}{k}} \frac{N-1}{N} \frac{k}{k-1} \frac{1}{k^2} \binom{N-2}{k-2} E(x^{2r}) \frac{N(N-k)}{k-1} + N^2E^2(x^r)$$

$$= E(x^{2r}) \frac{N-k}{N(k-1)} + E^2(x^r)$$

Substituting the results (ii) and (iii) into (i) gives the desired result:

$$\text{(iv)} \qquad \frac{1}{\binom{N}{k}} \sum_{\text{all } A} V_r(A) = E(x^{2r})\left(\frac{k}{k-1}\,\frac{N-1}{N} - \frac{N-k}{N(k-1)}\right) - E^2(x^r)$$

$$= E(x^{2r}) - E^2(x^r)$$

$$= \operatorname{Var}(x^r).$$

References

[1] Buzen, J. P., "Operational analysis: the key to the new generation of performance prediction tools," Proc. IEEE COMPCON (1976), IEEE, New York.

[2] Buzen, J. P., "Fundamental operational laws of computer system performance," Acta Informatica 7, 2(1976), 167-182.

[3] Denning, P. J. and J. P. Buzen, "The operational analysis of queueing network models," Computing Surveys 10, 3 (September 1978), 225-261.

[4] Buzen, J. P. and P. J. Denning, "Operational analysis: questions and answers," Technical Report CSD-TR-318, Computer Science Dept., Purdue University, W. Lafayette, IN 47907, (November 1979).

[5] Buzen, J. P. and P. J. Denning, "An operational treatment of queueing distributions and mean value analysis," Technical Report CSD-TR-309, Computer Science Dept., Purdue University, W. Lafayette, IN 47907 (November 1979).

[6] Buzen, J. P. and P. J. Denning, "Measuring and calculating queue length distributions," IEEE Computer (February 1980).

[7] Bard, Y. and M. Schatzoff, "Statistical methods in Computer Performance Analysis," in Current Trends in Programming Methodology III (K. M. Chandy and R. Yeh, Eds.) Prentice-Hall (1978), 1-51.

[8] Iglehart, D. L., "The regenerative method for simulation analysis," in Current Trends in Programming Methodology III (K. M. Chandy and R. Yeh, Eds.) Prentice-Hall (1978), 52-71.

[9] Wozencraft, J. and I. Jacobs, Principles of Communications Engineering, Wiley (1965), 106-111.

[10] Rubin, D. B., "The phenomenological Bayesian perspective in sample surveys from finite populations: Foundations," presented at Inst. for Math. Statistics, 1978; available from Dept. Statistics, Harvard University, Cambridge, MA 02138.

[11] Cochran, W. G., Sampling Techniques, New York: Wiley (1953), Chapter 2.

[12] Birnbaum, Z. W., Introduction to Probability and Mathematical Statistics, Harper & Brothers (1962), pp 175-179.

[13] Madow, W. G., "On the limiting distributions of estimates based on samples from finite universes," Annals of Mathematical Statistics 19 (1948), 535-545.