Proceedings of the 1998 Winter Simulation Conference
D.J. Medeiros, E.F. Watson, J.S. Carson and M.S. Manivannan, eds.

LABATCH.2: SOFTWARE FOR STATISTICAL ANALYSIS OF SIMULATION SAMPLE PATH DATA

George S. Fishman

Department of Operations Research
CB#3180 Smith Building
University of North Carolina
Chapel Hill, NC 27599-3180, U.S.A.

ABSTRACT

LABATCH.2 is a collection of computer programs available in C, FORTRAN, and SIMSCRIPT II.5 by anonymous ftp, at http://www.or.unc.edu/∼gfish/labatch.2.html. It performs statistical analyses on sample sequences collected on strictly stationary stochastic processes. Designed to make its implementation easy for potential users, it may be invoked in-line or from a stored data file. For each sample sequence of length t, LABATCH.2 takes O(t) computing time and O($\log_2 t$) space.

For each series, LABATCH.2 provides an asymptotically valid confidence interval, based on the batch means method, for assessing how well its sample average approximates its true unknown mean. It also produces interim estimates of the variance of the sample average that enable a user to detect systematic error in the latest variance estimate, due to correlation between batches. It also allows a user to assess the extent to which the sample average is free of initial conditions.

LABATCH.2 has an interactive option that displays interim results on screen. Based on these quantities, a user may instruct LABATCH.2 to continue execution until the next update or to terminate statistical analysis and write the final results to a file.

1 INTRODUCTION

Since the typical simulation user's interest rarely lies in statistical analysis, only measures of assessment automatically generated during or at the completion of a simulation run can be expected to attract her/his attention. LABATCH.2 is a collection of computer programs designed to provide these measures. It performs statistical analyses on sample sequences collected on strictly stationary stochastic processes and offers two modes of implementation. One integrates LABATCH.2 into an executing data-generating program (e.g., simulation) to analyze the evolving data on repeated subroutine calls; the other takes data from an existing file as input. In addition to minimizing user effort, the first option considerably reduces space requirements. It also allows user interaction with the executing program via screen displays of interim estimates. The second option permits statistical analysis of stored data, regardless of source and date of generation, thereby making LABATCH.2 applicable in a considerably wider range of data-generating environments.

LABATCH.2 is a revision of LABATCH (Fishman 1996, Fishman and Yarberry 1997) that considerably simplifies its implementation and use. The simplifications are its most attractive feature. A user merely inserts a single subroutine call statement in her/his main program and assigns values to several control arguments of the subroutine. C, FORTRAN, and SIMSCRIPT II.5 implementations of LABATCH.2 are obtainable by anonoymous file transfer procedure (ftp) at http://www.or.unc.edu/∼gfish/labatch.2.html. Also available is report TR 97/04 (Fishman 1997), which contains complete details for implementation. The present account is an abridged version of the report.

2 OVERVIEW

For each sample sequence, $X_1, \ldots, X_t$, in its input, LABATCH.2 computes as part of its output a sample average, $\bar{X}_t$, as an estimate of its true unknown mean, μ, and an asymptotically valid $100 \times (1-\delta)$ percent confidence interval for assessing how well $\bar{X}_t$ approximates μ. The confidence interval relies crucially on an estimate, $BW(L,B)$, of the *asymptotic variance*, $\sigma^2_\infty := \lim_{t\to\infty} t \operatorname{var}\bar{X}_t$, computed by the *batch means method*, where B denotes batch size L, the number of batches, and *W(L,B)*, the sample variance of a batch average. Since this estimate of σ^2_∞ is also subject to *systematic error* as well as *random sampling error* and since the validity of the confidence interval depends on this systematic error being relatively negligible, LABATCH.2 also displays interim calculations of $BW(L,B)$ that allow a user to assess the extent to which systematic error remains in the final variance estimate used to compute the confidence interval for the mean μ. The ability to make this assessment with LABATCH.2 output is one of its most valuable assets.

In the present context, systematic error in the variance estimate can be present even for sample data free of bias due to initial conditions. It arises from neglecting all but the term proportional to t^{-1} in var$\bar{X}_t$ and from ignoring the correlation between batch averages. To provide a basis for systematic error assessment, LABATCH.2 computes a sequence of estimates of σ_∞^2 based on data subsequences of increasing lengths $t_1 < t_2 < \cdots < t_{J(t)} \leq t$, where $t_{i+1} = 2t_i$ for $i = 1, \ldots, J(t) - 1$, and where the user-specified path length, t, determines t_1 and $J(t)$. See Sec. 2.4.

As illustration, consider a simulation of the M/M/1 queueing model with .90 interarrival rate and unit service rate. The simulation began in the steady state and terminated when customer $t = 10^7$ entered service. Figure 1a displays LABATCH.2–computed 99 percent confidence intervals for the mean waiting time in queue (Series 1) and for the probability that a customer waits (Series 2). The true values are 9 and .90 respectively.

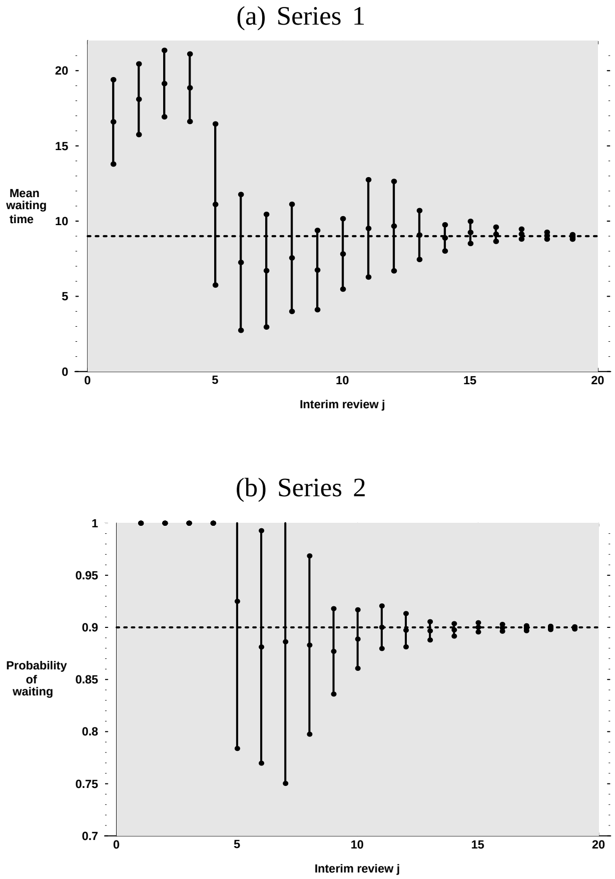
For each series, Figure 1b displays LABATCH.2–computed sequences of point and 99 percent interval estimates of μ and point estimates, $\sqrt{BW(L,B)}$, of σ_∞ as the batch size, B, and the number of batches, L, grow with sample path length according to the ABATCH rule (Sec. 3.3). As sample path length doubles $(t_{j+1} = 2t_j)$, the ABATCH rule doubles B if a test of the hypothesis H, batch averages are independent, detects systematic error. If no systematic error is detected (H is accepted), the rule increases B by a factor of $\sqrt{2}$ approximately. This implies that either L remains constant (H is rejected) on successive reviews or L also increases (H is accepted) approximately by a factor of $\sqrt{2}$. The $\sqrt{BW(L,B)}$ sequences in Figure 1b reveal that systematic error in the example has become negligible for Series 1 on reviews $j \geq 11$ for Series 2 on reviews $j \geq 7$. Note that the evidence for these assertions comes from interim review results subsequent to review $j = 11$ (Series 1) and $j = 7$ (Series 2), thus establishing the intrinisic value of the tableaus in Figure 1.b.

Figure 2 graphically displays the point and interval estimates (cols. 5, 6, and 7 in Figure 1b) for μ for each series. The graphs provide a convenient way of assessing the accuracy of the sample averages at a glance. These and all other graphs in the paper were created using Mathematica Ⓣ applied to the LABATCH.2 output after deleting the final tableau (Figure 1a) and all header and trailer information from the interim review tableaus

```
                                                       (a)
Final  Tableau
                                                 Mean  Estimation
                                                 ***************
                                                 (t  =  10000000  )

                                                                 99.0%
              _        Standard Error      Confidence  Interval                 _
Series        X       Sqrt[B*W(L,B)/t]      Lower       Upper     (Upper-Lower)/|X|

   1    0.8949D+01      0.5681D-01      0.8802D+01   0.9097D+01        0.3290D-01

   2    0.8995D+00      0.4200D-03      0.8984D+00   0.9006D+00        0.2416D-02
   _
   X  is  based  on  all  t  observations.
   W(L,B)  is  based  on  first   91.75%  of  the  t  observations.
                                                       (b)
Interim  Review  Tableau
                           ABATCH  Data  Analysis  for  Series   1

                                                                       99.0%
                                                   _           Confidence  Interval
Review    L*B      L        B      X        Lower       Upper Sqrt[B*W(L,B)] p-value

  1       35       7        5  0.1660D+02 0.1379D+02 0.1940D+02  0.4475D+01  0.3804
  2       70      10        7  0.1810D+02 0.1575D+02 0.2045D+02  0.6047D+01  0.1109
  3      140      14       10  0.1914D+02 0.1693D+02 0.2135D+02  0.8680D+01  0.1524
  4      280      20       14  0.1886D+02 0.1662D+02 0.2111D+02  0.1313D+02  0.0194
  5      560      20       28  0.1111D+02 0.5748D+01 0.1646D+02  0.4432D+02  0.0000
  6     1120      20       56  0.7254D+01 0.2741D+01 0.1177D+02  0.5278D+02  0.0001
  7     2240      20      112  0.6706D+01 0.2960D+01 0.1045D+02  0.6196D+02  0.0005
  8     4480      20      224  0.7556D+01 0.3996D+01 0.1112D+02  0.8328D+02  0.0358
  9     8960      20      448  0.6747D+01 0.4111D+01 0.9383D+01  0.8721D+02  0.7293
 10    17920      28      640  0.7817D+01 0.5478D+01 0.1016D+02  0.1130D+03  0.7095
 11    35840      40      896  0.9513D+01 0.6280D+01 0.1275D+02  0.2260D+03  0.0655
 12    71680      40     1792  0.9668D+01 0.6693D+01 0.1264D+02  0.2941D+03  0.8172
 13   143360      56     2560  0.9073D+01 0.7449D+01 0.1070D+02  0.2304D+03  0.7200
 14   286720      80     3584  0.8883D+01 0.8012D+01 0.9754D+01  0.1767D+03  0.2993
 15   573440     112     5120  0.9248D+01 0.8511D+01 0.9985D+01  0.2128D+03  0.7163
 16  1146880     160     7168  0.9126D+01 0.8653D+01 0.9600D+01  0.1945D+03  0.0993
 17  2293760     160    14336  0.9138D+01 0.8809D+01 0.9468D+01  0.1913D+03  0.1852
 18  4587520     224    20480  0.9032D+01 0.8805D+01 0.9259D+01  0.1874D+03  0.7882
 19  9175040     320    28672  0.8949D+01 0.8802D+01 0.9097D+01  0.1796D+03  0.2362

  If  data  are  independent:

   10000000 10000000       1  0.8949D+01 0.8941D+01 0.8957D+01  0.9821D+01  0.0000

  0.10  significance  level  for  independence  testing.
  Review  19  used  the  first   91.75%  of  the  t  observations  for  W(L,B).


Interim  Review  Tableau
                           ABATCH  Data  Analysis  for  Series   2

                                                                       99.0%
                                                   _           Confidence  Interval
Review    L*B      L        B      X        Lower       Upper Sqrt[B*W(L,B)] p-value

  1       35       7        5  0.1000D+01 0.1000D+01 0.1000D+01  0.0000D+00  0.0000
  2       70       7       10  0.1000D+01 0.1000D+01 0.1000D+01  0.0000D+00  0.0000
  3      140       7       20  0.1000D+01 0.1000D+01 0.1000D+01  0.0000D+00  0.0000
  4      280       7       40  0.1000D+01 0.1000D+01 0.1000D+01  0.0000D+00  0.0000
  5      560       7       80  0.9250D+00 0.7838D+00 0.1066D+01  0.9014D+00  0.0067
  6     1120       7      160  0.8812D+00 0.7697D+00 0.9928D+00  0.1007D+01  0.0140
  7     2240       7      320  0.8862D+00 0.7503D+00 0.1022D+01  0.1734D+01  0.1002
  8     4480      10      448  0.8830D+00 0.7975D+00 0.9686D+00  0.1762D+01  0.8214
  9     8960      14      640  0.8770D+00 0.8360D+00 0.9180D+00  0.1289D+01  0.9589
 10    17920      20      896  0.8888D+00 0.8607D+00 0.9169D+00  0.1316D+01  0.9648
 11    35840      28     1280  0.9001D+00 0.8797D+00 0.9206D+00  0.1398D+01  0.3029
 12    71680      40     1792  0.8973D+00 0.8813D+00 0.9133D+00  0.1581D+01  0.9235
 13   143360      56     2560  0.8967D+00 0.8879D+00 0.9055D+00  0.1250D+01  0.6376
 14   286720      80     3584  0.8976D+00 0.8916D+00 0.9036D+00  0.1220D+01  0.5008
 15   573440     112     5120  0.9001D+00 0.8956D+00 0.9045D+00  0.1280D+01  0.1352
 16  1146880     160     7168  0.8996D+00 0.8963D+00 0.9029D+00  0.1374D+01  0.5547
 17  2293760     224    10240  0.8992D+00 0.8968D+00 0.9016D+00  0.1381D+01  0.6289
 18  4587520     320    14336  0.8995D+00 0.8979D+00 0.9011D+00  0.1313D+01  0.7435
 19  9175040     448    20480  0.8995D+00 0.8984D+00 0.9006D+00  0.1328D+01  0.6766

  If  data  are  independent:

   10000000 10000000       1  0.8995D+00 0.8992D+00 0.8997D+00  0.3007D+00  0.0000

  0.10  significance  level  for  independence  testing.
  Review  19  used  the  first   91.75%  of  the  t  observations  for  W(L,B).
```

Figure 1: LABATCH.2 output for M/M/1 queueing simulation (Series 1: waiting time in queue, Series 2: 1:=wait, 0:= no wait)

(Figure 1b). EXCEL Ⓣ or similar software could have been used as alternatives.


(a) Series 1

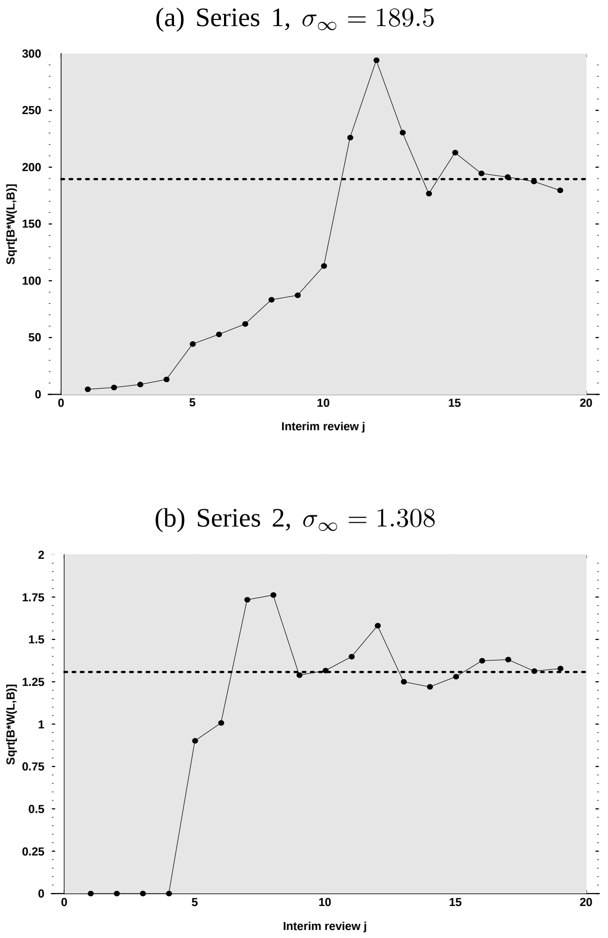
(b) Series 2

Figure 2: LABATCH.2 sample means and 99% confidence intervals for ABATCH rule; simulation starts in steady state

Figure 3 graphically displays $\sqrt{BW(L,B)}$ (col. 8 in Figure 2b) as estimates of σ_∞ for each series. For Series 1, $\sigma_\infty = 189.3$ and for Series 2, $\sigma_\infty = 1.308$ (Blomqvist 1967), giving us the luxury of a comparison based on theory.

For any path length, $t \geq 20$, LABATCH.2 automatically computes the number of batches, L, and the batch size, B, to be used in its first review. For example, it chose L=7 and B=5 for $t = 10^7$ for the M/M/1 example. This automation, described in Sec. 2.4, relieves the user of the need to specify initial values for L and B, as in the original LABATCH package.

A call from a C / FORTRAN / SIMSCRIPT II.5 main program to


(a) Series 1, $\sigma_\infty = 189.5$

(b) Series 2, $\sigma_\infty = 1.308$

Figure 3: LABATCH.2 estimates, $\sqrt{BW(L,B)}$, of σ_∞

$$\text{BATCH\_MEANS(IN\_UNIT,OUT\_UNIT,T,S\_NUM,PSI\_VECTOR,}\quad \text{DELTA,RULE, BETA,L\_UPPER,SCREEEN)} \tag{1}$$

implements LABATCH.2. As an example, suppose they arguments are:

$$\begin{aligned}
\text{IN\_UNIT} &= 0\\
\text{OUT\_UNIT} &= 15\\
\text{T} &= 1000\\
\text{S\_NUM} &= 2\\
\text{PSI\_VECTOR} &= \text{pointer to data vector with}\\
&\quad\ \text{S\_NUM entries}\\
\text{DELTA} &= .01\\
\text{RULE} &= 1\\
\text{BETA} &= .10\\
\text{L\_UPPER} &= 30\\
\text{SCREEN} &= 0 \qquad .
\end{aligned} \tag{2}$$

Then LABATCH.2 processes S_NUM=2 sequences in T=1000 iterative calls from the user's main program, computes $100 \times (1 -$ DELTA$)$=99 percent confidence intervals for the sample averages, and writes the output to a file called $\underset{\text{SIMU15}}{\overset{\text{c.15}}{\text{fort.15}}}$ (IN_UNIT=0, OUT_UNIT=15). RULE=1 causes LABATCH.2 to employ the ABATCH rule to determine batch size on each review and BETA=.10 causes it to test for independence of batch averages at the .10 significance level. LABATCH.2 begins its first iteration with the number of batches no greater than L_UPPER = 30. SCREEN=0 suppresses the screen interative feature (Sec. 4.2). In practice, all but the values of T and S_NUM can be set once and the subroutine used repeatedly in different settings.

Testing determines whether the batch size on review $j+1$ increases by a factor of 2 (rejection) or approximately as $\sqrt{2}$ (success). A doubling of batch size aims at reducing any residual systematic error detected on review j as fast as possible. A $\sqrt{2}$ increase signals that the number of batches are also increasing approximately by a $\sqrt{2}$ factor. Growth in both batch size and number batches as the sample path length grows is a necessary condition for obtaining a consistent estimator of σ^2_∞. Recall that σ^2_∞/t approximates the large-sample variance of $\bar{X}_t$.

3 RATIONALE

Let $\{X_i,\ i \geq 1\}$ denoted a strictly stationary stochastic process with unknown mean μ. Given a sample path, $X_1, \ldots, X_t$,

$$\bar{X}_t := \frac{1}{t}\sum_{i=1}^{t} X_i$$

provides a point estimate of μ. When data collection begins in the steady state, $\bar{X}_t$ is an unbiased estimator of μ.

Assumption 1. $\sigma^2_t := \text{var}\bar{X}_t$ satisfies $t\sigma^2_t \to \sigma^2_\infty$ as $t \to \infty$, where σ^2_∞ is a positive constant.

Assumption 2. There exist a constant $\lambda \in (0, 1/2)$ such that

$$t^{1/2}(\bar{X}_t - \mu)/\sigma_\infty = Z(t)/t^{1/2} + \text{O}(t^{-\lambda}) \text{ as } t \to \infty \text{ w.p.1},$$

where $\{Z(s), s \geq 0\}$ denotes standard Brownian motion.

Assumption 2 is the *Assumption of Strong Approximation* (ASA). A λ close to $1/2$ signifies a marginal distribution for the X_i close to the standard normal and low correlation between X_i and X_j for $\forall\ i \neq j$. Conversely, λ close to zero implies the absence of one or both of these properties. See Philipp and Stout (1975). Section 3.4 relies on Assumptions 1 and 2.

To assess how well $\bar{X}_t$ approximates μ, we need an estimate of σ^2_∞. The batch means method offers one option. Let $b(t)$ denote a positive integer $(< t)$, let $l(t) := \lfloor t/b(t) \rfloor$, and let $t'(t) := l(t)b(t)$. Our version of the batch means method partitions the subsequence, $X_1, \ldots, X_{t'(t)}$, into $l(t)$ nonoverlapping batches each of size $b(t)$, computes the batch averages,

$$Y_{jb(t)} := \frac{1}{b(t)} \sum_{i=1}^{b(t)} X_{(j-1)b(t)+i} \quad j = 1, \ldots, l(t), \tag{3}$$

and an estimate of var$Y_{jb(t)}$,

$$W_{l(t)b(t)} := \frac{1}{l(t)-1} \sum_{j=1}^{l(t)} (Y_{jb(t)} - \bar{X}_{t'(t)})^2, \tag{4}$$

and uses $b(t)W_{l(t)b(t)}$ as an estimte of σ^2_∞. Then

$$[\bar{X}_t \pm \tau_{l(t)-1}(1-\delta/2)\ \sqrt{b(t)W_{l(t)b(t)}/t}] \tag{5}$$

provides an approximating $100 \times (1-\delta)$ percent confidence interval for μ, where $\delta \in (0,1)$ and $\tau_{l(t)-1}(1-\delta/2)$ denotes the $1-\delta/2$ critical value of Student's t distribution with $l(t)-1$ degrees of freedom.

Damerdji (1994) indicates how $\{(l(t), b(t))\}$ must grow in order for (5) to be an asymptotically valid confidence interval for μ in the sense that it achieves the specified $1-\delta$ coverage rate as $t \to \infty$.

Unless clarity demands otherwise, we hereafter write $b := b(t)$ and $l := l(t)$ when batch size and number of batches are deterministic functions of t. Also, we assume that $t'(t) = t$ so that the batches in expression (3) use all the observations. Later, we consider the more general case of $t'(t) \leq t$.

Under relatively weak conditions, the choice, $l(t) \propto t^{1/2}$ and $b(t) \propto t^{1/2}$, induces the fastest convergence of the true coverage rate to the specified theoretical coverage rate, $1-\delta$ (Chien 1989). For other choices, see Fishman (1997).

Observe that for $t'(t) = t$

$$\text{E}V_t = \frac{t}{l-1}(\sigma^2_b - \sigma^2_t), \tag{6}$$

which, for positively autocorrelated sequences, is usually negative. If $\text{E}(X_1 - \mu)^{12} < \infty$ and $\{X_i\}$ is ϕ-mixing with $\phi_i = \text{O}(i^{-9})$, then expression (6) takes the form (Goldsman and Meketon 1986)

$$\text{E}V_t = \sigma^2_\infty + \gamma(l+1)/t + \text{o}(1/b), \tag{7}$$

where

$$\gamma := -2\sum_{i=1}^{\infty} i\,\text{cov}(X_1, X_{1+i}).$$

Also (Chien et al. 1996),

$$\text{var } V_t = \frac{2\sigma_\infty^4(l+1)}{(l-1)^2} + \mathrm{O}(1/lb^{1/4}) + \mathrm{O}(1/l^2). \qquad (8)$$

Note that $\sigma_\infty^2 = \sum_{i=\infty}^{\infty} \text{cov}(X_1, X_{1+i})$.

Expression (6) leads to the representation

$$V_t - \sigma_\infty^2 = \underbrace{t\sigma_t^2 - \sigma_\infty^2}_{\text{error due to finite } t} \; \underbrace{-\, t\sigma_t^2 \left(\frac{1 - b\sigma_b^2/t\sigma_t^2}{1-b/t}\right)}_{\text{error due to ignoring correlation}} + \underbrace{\epsilon_t}_{\text{error due to random sampling between batches}}, \quad (9)$$

where ϵ_t has mean zero and variance (8). Hereafter, we collectively refer to the errors due to finite t and to ignoring correlation as systematic error. From expression (7) we see that the systematic error behaves as $\mathrm{O}((l+1)/t) = \mathrm{O}(1/b)$ whereas, from expression (8), $\sqrt{\text{var } \epsilon_t}$ behaves as $\mathrm{O}(1/l^{1/2})$, revealing the tradeoff between the two types of error that a choice of l and b induces.

In what follows, we adopt a strategy that leads to $l \propto t^{1/2}$ and $b \propto t^{1/2}$, once t becomes sufficiently large. Two considerations motivate this choice. First, since the approximating $100\times(1-\delta)$ percent confidence interval (5) provides the means of assessing how well $\bar{X}_t$ approximates μ, we would like its true coverage rate to be as close to $1-\delta$ as possible. Secondly, the square root option facilitates the efficient computation of a sequence of estimates $V_{t_1}, V_{t_2}, \ldots,$ based on subsequences $\{X_1, \ldots, X_{t_1}\}$, $\{X_1, \ldots X_{t_2}\}, \ldots,$ respectively, for $t_1 < t_2 < \ldots \leq t$, that provide a means for assessing the extent to which systematic error remains in V_t, our final estimate of σ_∞^2. The next subsection elaborates on this issue.

3.1 Interim Review

In what follows we take t_1 as given and $t_{j+1} := 2t_j$ for $j = 1, 2, \ldots$. For a sequence of i.i.d. random variables $X_1, X_2, \ldots,$ this choice implies corr$(\bar{X}_{t_j} \bar{X}_{t_{i+j}}) = \sqrt{t_j/t_{i+j}}$ for $i \geq 0$ and $j \geq 1$. More generally, if corr$(X_1, X_{1+j}) = \alpha^{|j|}$, for some $-1 < \alpha < 1$, then $\lim_{j\to\infty}$ corr$(\bar{X}_{t_j}, \bar{X}_{t_{i+j}}) = \lim_{j\to\infty} \sqrt{t_j/t_{i+j}}$. We choose $t_{j+1}/t_j := 2$ which implies corr$(\bar{X}_{t_j}, \bar{X}_{t_{i+j}}) = 2^{-i/2}$ in the independent and asymptotic (as $j \to \infty$) cases. Any growth factor less than 2 would induce higher correlation. Most importantly, this choice makes possible the O(t) computing time and the O$(\log_2 t)$ space bounds for LABATCH.2.

3.2 FNB and SQRT Rules

The description here closely follows Fishman (1996) and Fishman and Yarberry (1997). For $j = 1, 2, \ldots,$ let $l_j := l(t_j)$ and $b_j := b(t_j)$. We illustrate the benefits of interim review for two assignment rules, FNB and SQRT, that form the basis for the LBATCH and ABATCH rules that LABATCH.2 incorporates. Given (l_1, b_1), the FNB rule fixes $l_j = l_1$ for all j and doubles the batch size $b_{j+1} = 2b_j$ on successive reviews.

Given (l_1, b_1), the SQRT rule sets

$$l_2 = \tilde{l}_1 := \lfloor \sqrt{2} l_1 + .5 \rfloor \qquad (10)$$

$$b_2 = \tilde{b}_1 := \begin{cases} 3 & \text{if } b_1 = 1 \\ \lfloor \sqrt{2} b_1 + .5 \rfloor & \text{if } b_1 > 1 \end{cases} \qquad (11)$$

$$\begin{aligned} l_{j+1} &= 2l_{j-1} \\ b_{j+1} &= 2b_{j-1} \quad j = 2, 3, \ldots \ . \end{aligned}$$

These assignments induce $l_{j+1}/l_j \doteq \sqrt{2}$ and $b_{j+1}/b_j \doteq \sqrt{2}$. By choosing (l_1, b_1) from $\mathcal{B}$ in Table 1, we ensure that $2l_1b_1 = \tilde{l}_1\tilde{b}_1$ so that $t_j = l_jb_j = 2^{j-1}l_1b_1$ and, therefore, $t_{j+1}/t_j = 2$, as desired. This constraint proves valuable in Sec. 3.3, which describes batch-size rules that combine the FNB and the SQRT rules.

3.3 LBATCH and ABATCH Rules

Let H denote the hypothesis: On review j, the l_j batches, $Y_{1b_j}, \ldots, Y_{l_jb_j}$, are mutually independent. The LBATCH and ABATCH rules both use the outcome of a test of H to switch between the FNB and SQRT rules on succesive reviews. The net effect is to retain the desirable properties of each rule while reducing the influence of their limitations. The principal features of these hybrid rules are:

LBATCH Rule

- Start with the FNB rule on review 1.
- For $j \geq 1$, if H is rejected on review j, use the FNB rule on review $j+1$.
- Once H is accepted on review j, use the SQRT rule on reviews, $j+1, j+2, \ldots$.

Table 1: $\mathcal{B} := \{(l_1, b_1) : 1 \leq b_1 \leq l_1 \leq 100,\ 2l_1b_1 = \tilde{l}_1\tilde{b}_1\}^\dagger$

b_1	l_1	b_1	l_1
$\underline{1}$	$\underline{3}$	38	54
2	3,6	$\underline{39}$	$\underline{55}$
3	4	$\underline{41}$	58,$\underline{87}$
4	6	$\underline{43}$	$\underline{61}$
$\underline{5}$	$\underline{7}$,14,$\underline{21}$,28	44	62, 93
$\underline{7}$	10,$\underline{15}$,20,$\underline{25}$,30,$\underline{35}$	$\underline{46}$	$\underline{65}$
$\underline{8}$	$\underline{11}$	48	51,68,85
$\underline{9}$	$\underline{13}$	$\underline{49}$	$\underline{69}$
10	14, 21,28	$\underline{50}$	$\underline{71}$
$\underline{12}$	$\underline{17}$,34,$\underline{51}$,68,$\underline{85}$	$\underline{51}$	$\underline{60}$,72,$\underline{84}$,96
14	15,20,25,30,35	$\underline{53}$	$\underline{75}$
$\underline{15}$	$\underline{21}$,28	54	57
$\underline{17}$	24,$\underline{36}$,48,$\underline{60}$,72,$\underline{84}$,96	55	78
$\underline{19}$	$\underline{27}$,54	$\underline{56}$	$\underline{79}$
20	21, 28	58	82
$\underline{21}$	$\underline{25}$, 30, $\underline{35}$	$\underline{60}$	68,$\underline{85}$
$\underline{22}$	$\underline{31}$,62,$\underline{93}$	61	86
24	34,51,68,85	62	66
25	28	$\underline{63}$	$\underline{89}$
$\underline{26}$	$\underline{37}$	65	92
$\underline{27}$	38,$\underline{57}$	$\underline{66}$	$\underline{93}$
28	30,35	$\underline{67}$	$\underline{95}$
$\underline{29}$	$\underline{41}$,82	68	72,84,96
$\underline{31}$	44,$\underline{66}$,88	$\underline{70}$	$\underline{99}$
32	45	72	85
$\underline{33}$	$\underline{47}$	82	87
34	36,48,60,72,84,96	$\underline{84}$	$\underline{85}$
$\underline{36}$	$\underline{51}$,68,$\underline{85}$	$\underline{85}$	$\underline{96}$
37	52	88	93

[†] $\tilde{l}_1$ is defined in expression (10) and $\tilde{b}_1$ in expression (11).

and: **ABATCH Rule**

- Start with the FNB rule on review 1.
- For $j \geq 1$, if H is rejected on review j, use the FNB rule on review $j+1$.
- For $j \geq 1$, if H is accepted on review j, use the SQRT rule on review $j+1$.

By initially fixing l, the LBATCH rule allows batch size, b, to increase at the maximal rate when H is rejected, thus dissipating systematic error in V_{t_j} as fast as possible. Once H is accepted, the rule switches to the SQRT rule to dissipate the error in coverage rate as rapidly as possible. By testing H on every review, the ABATCH rule takes into consideration the possibility of Type II errors on successive replications. The LBATCH rule ignores this source of error whereas the ABATCH rule allows the procedure to correct itself.

As a consequence of this testing, batch size and number of batches on review j (> 1) are random. To acknowledge this property, we denote them as B_j and L_j respectively. Let K_j denote the number of test rejections of H on reviews $1, \ldots, j$. Then the LBATCH and ABATCH rules induce:

$$L_j = \begin{cases} 2^{(j-K_j)/2} l_1 & \text{if } j - K_j \text{ is even} \\ 2^{(j-K_j-1)/2} \tilde{l}_1 & \text{if } j - K_j \text{ is odd} \end{cases}$$

and

$$B_j = \begin{cases} 2^{(j+K_j-2)/2} b_1 & \text{if } j - K_j \text{ is even} \\ 2^{(j+K_j-1)/2} \tilde{b}_1 & \text{if } j - K_j \text{ is odd,} \end{cases}$$

so that $t_j = L_j B_j = 2^{j-1} l_1 b_1$ for $j = 1, 2, \ldots$.

To account for the randomness of $\{(L_j, B_j),\ j \geq 1\}$ with regard to limiting behavior, we rely on Yarberry (1993) and Fishman and Yarberry (1997) who give conditions for which $B_j W_{L_j B_j} \to \sigma^2_\infty$ as $j \to \infty$ w.p.1 and

$$\frac{\bar{X}_{t_j} - \mu}{\sqrt{W_{L_j B_j}/L_j}} \xrightarrow{\text{d}} \mathcal{N}\ (0, 1) \quad \text{as } j \to \infty. \tag{12}$$

□

3.4 Choosing t, l_1, and b_1

For given $(l_1, b_1) \in \mathcal{B}$, choosing t so that $J(t) := \log(t/l_1 b_1)/\log 2$ is an integer results in $t'(t) = t$. As a consequence, the LBATCH and ABATCH rules used all t observations to estimate $W_{L_{J(t)} B_{J(t)}}$ and $\bar{X}_{t'(t)} = \bar{X}_t$. Since choosing l_1, b_1, and t subject to the constraint may be too burdensome for some users, LABATCH.2 merely requires that t be specified and then chooses l_1 and b_1 to maximize $t'(t)$.

Let $\mathcal{B}(t)$ denote the subset of $\mathcal{B}$ that maximizes $t'(t)$. LABATCH.2 chooses (l_1, b_1) to be the element of $\mathcal{B}(t)$ that maximizes $J(t)$, the number of interim reviews. For $l_1 \leq 30$ and $t \geq 500$, this leads to $t'(t)/t \geq .898$.

We now reconcile

$$A_t := \frac{\bar{X}_t - \mu}{\sqrt{B_{J(t)} W_{L_{J(t)} B_{J(t)}}/t}} \tag{13}$$

whose properties establish the basis for the LABATCH.2 confidence interval (5), with $l(t) = L_{J(t)}, b(t) = B_{J(t)}$,

$W_{l(t)b(t)} = V_{t'(t)}/B_{J(t)}$, and

$$G_t := \frac{\bar{X}_{t'(t)} - \mu}{\sqrt{B_{J(t)}W_{L_{J(t)}B_{J(t)}}/t'(t)}} \qquad (14)$$

whose properties are those of the variate in (12) for $j = J(t)$. Recall that $t'(t) = t_{J(t)} = L_{J(t)}B_{J(t)}$.

Whereas A_t uses all t observations to estimate μ, G_t uses only the first $t'(t)$ observations. However, both use the $B_{J(t)}W_{L_{J(t)}B_{J(t)}}$ as an estimate of σ^2_∞. Since $(\bar{X}_{t'(t)} - \mu)/\sqrt{\sigma^2_\infty/t'(t)}$ and $(\bar{X}_t - \mu)/\sqrt{\sigma^2_\infty/t}$ are both asymptotically $\mathcal{N}$(0,1), by Assumption 2, and since $B_{L_{J(t)}}W_{LJ(t)B_{J(t)}}$ is strongly consistent, it a modification of a theorem 6 in Yarberry (1993) can yield asymptotic normality for A_t as well. With regard to statistical efficiency, A_t reduces interval width by a factor of $1 - \sqrt{t'(t)/t} \leq 1 - \sqrt{.889} = .0571$ when $(l_1 b_1) \in \mathcal{B}(t)$ for $t \geq 500$.

To test H, LABATCH.2 uses the von Neumann ratio (von Neumann 1941, Young 1941, Fishman 1978, 1996). The rightmost column in each Interim Review Tableau in Figure 1 lists the corresponding p-values, the probability, under H, of seeing a test statistic larger than the one observed. If p-value $\leq$ BETA, H is rejected; otherwise is accepted.

4 ADEQUACY OF THE WARM-UP INTERVAL

LABATCH.2 also provides an assessment of the extent of warm-up bias in $\bar{X}_t$. We again use the M/M/1 simulation. However, this time the run began with an arrival to an empty and idle system and data collection began with $X_1 :=$ waiting time of the first arrival. For this scenario, $X_1 = 0$ w.p.1 and $X_1, X_2, \ldots, X_t$ is stationary, only asymptotically (as $t \to \infty$). This is a more biased environment than one would expect to encounter when data collection begins after a user-specified warm-up interval.

Figure 4 displays the 99 percent confidence intervals for mean waiting time in queue, taken from the LABATCH.2 interim review tableau for Series 1 for this run. It suggests little bias after review 11, which corresponds to $2^{11} \times L_1 \times B_1 = 2048 \times 7 \times 5 = 71,680$ observations.

Suppose the user had specified $t = 4480$ for the sample path length. Then LABATCH.2 would have generated the same sample averages and confidence intervals for $L_1 = 7$ and $B_1 = 5$, but only for the first seven interim reviews. Moreover, a display of these results would have aroused considerably more concern about the dissipation of bias in $\bar{X}_t$. Interestingly, had we taken $t = 4480$ in our first simulation, which began in an equilibrium state, we might equally be suspicious of $\bar{X}_t$ based on the path it displays in Figure 2 for $j = 1, \ldots, 7$. However, this observation in no way mitigates the value of the assessment when we know for a fact that $\bar{X}_t$ may contain systematic bias as a result of starting in an arbitrarily chosen state and possibly choosing an inadequate warm-up interval.

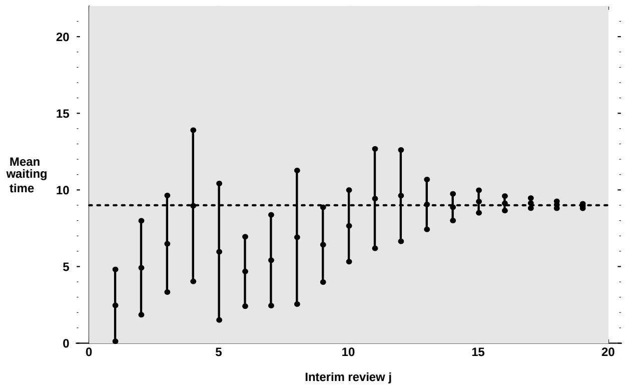


Figure 4: LABATCH.2 sample means and 99 percent confidence intervals for Series 1 ABATCH rule; simulation starts in empty and idle state

This example in no way mitigates the traditionally sound advice of truncating a warm-up interval in the sample data to reduce the influence of initial conditions.

5 FEATURES

Several features of LABATCH.2 allow wide latitude for using it in practice. For example, one can easily strip the header and trailer entries from the column display in the interim review tableau and then transfer the remaining tableau to a spreadsheet environment, thus facilitates graphmaking. Experience has shown that little effort is needed to effect the desired graphs, provided software such as Mathematica Ⓣ or EXCEL Ⓣ is available.

5.1 Two Modalities

As already mentioned, LABATCH.2 provides two ways of accessing data for statistical analysis. One requires the user to insert a call statement into the data-generating program which executes the call each time it produces a new data vector. Calling and executing BATCH_MEANS each of T times (Table 1) that a data vector with S_NUM entries is generated results in O(S_NUM×T) computing time and O(S_NUM× $\log_2$T) space being used to generate the LABATCH.2 output. Both complexities arise from choosing rules that cause either $B_{j+1} = 2B_j$ or $B_{j+1} \doteq \sqrt{2}B_j$ on successive reviews $j = 1, 2, \ldots$. The space bound is particularly appealing when $t =$T is very large. Yarberry (1993) and Alexopoulos et al. (1997) describe the basis for these complexities.

The other option allows LABATCH.2 to read its data from a file, giving the software a considerably broader range of application than merely for in-line generated sample records. We illustrate how this option works in the context of the M/M/1 example, but stress the applicability of the approach to stored sample data.

If IN_UNIT=30 and OUT_UNIT=15 (Table 1), then a main program needs to call BATCH_MEANS just once to cause LABATCH.2 to read its sample data from a file called c.30 fort.30 SIMU30 and to write its output to c.15 fort.15. SIMU15 While the O(S_NUM×T) time and O(S_NUM× $\log_2$T) space bounds remain for LABATCH.2, they do not tell the whole story. In particular, the input file requires space proportional to S_NUM×T, which for S_NUM as small as 1 can be substantial for sufficiently large T.

Programs written in C, FORTRAN, or SIMSCRIPT II.5 can implement the second option without qualification. Any program that provides access to and execution of a C, FORTRAN, or SIMSCRIPT II.5 subroutine can take advantage of the first option. In a simulation environment, a user-written program in a language that provides standard linkages for incorporating a subroutine generally will consume less time calling BATCH_MEANS and analyzing data than a program generated at the icon-level in a point-and-click environment.

5.2 Interacting with the Analysis

If SCREEN=0 in expression (2), then LABATCH.2 performs as described in Sec. 2. However, if SCREEN=1, then after executing interim review j, LABATCH.2 displays on the screen $\bar{X}_{t_j}$, the sample coefficient variation of $\bar{X}_{t_j}$, sometimes called the relative standard error, and $\sqrt{B_j W(L_j, B_j)}$, for Series i in column $i+1$ for $i = 1, \ldots, \min(6, \mathrm{S\_NUM})$, where screen size dictates the upper limit on i. LABATCH.2 then asks the user if he/she wishes to continue.

Figure 5 shows an abbreviated screen display for the steady-state M/M/1 simulation, where column 2 shows the sequence of estimates for Series 1 and column 3 does likewise for Series 2. If the user concludes that $\sqrt{B_j W(L_j, B_j)}$ has stabilized in each column, so that the systematic errors have become negligible, and that the C.V.($\bar{X}_{t_j}$) are sufficiently small, then he/she may terminate the simulation by typing "n". This action causes LABATCH.2 to compute confidence intervals for the S_NUM and to write both the final and interim review tableaus (Figure 1) to file OUT_UNIT. If this action occurs immediately after the on-screen display, for review j, then the final tableau, as well as the interim review tableaus, use $L_j B_j$ observations. If the screen display suggests that systematic error remains, then, provided that $L_j B_j < t$ typing "y" causes LABATCH.2 to collect additional data, to perform the next review, to display $\sqrt{B_{j+1} W(L_{j+1}, B_{j+1})}$ on screen, and to ask the user whether or not he/she wishes to continue. If $L_j B_j = t'(t)$ and the user types "y", then the simulation goes to completion and LABATCH.2 uses all t=T observations for the sample averages and the first $t'(t)$ observations for the final $B_{J(t)} W(L_{J(t)}, B_{J(t)})$'s.

```
X_BAR, C.V.(X_BAR), and Sqrt[B*W(L,B)] for
         Series 1 Through  2
*******************************************
       No. of
Review   Obs.               1        2

 1      35 X_BAR        0.1660D+02 0.1000D+01
          C.V.(X_BAR)   0.4557D-01 0.0000D+00
         Sqrt[B*W(L,B)] 0.4475D+01 0.0000D+00
continue[y/n]? y
  .         .             .        .
  .         .             .        .
  .         .             .        .

 9    8960 X_BAR        0.6747D+01 0.8770D+00
          C.V.(X_BAR)   0.1366D+00 0.1553D-01
         Sqrt[B*W(L,B)] 0.8721D+02 0.1289D+01
continue[y/n]? y
10    17920 X_BAR       0.7817D+01 0.8888D+00
          C.V.(X_BAR)   0.1080D+00 0.1106D-01
         Sqrt[B*W(L,B)] 0.1130D+03 0.1316D+01
continue[y/n]? y
11    35840 X_BAR       0.9513D+01 0.9001D+00
          C.V.(X_BAR)   0.1255D+00 0.8205D-02
         Sqrt[B*W(L,B)] 0.2260D+03 0.1398D+01
continue[y/n]? y
12    71680 X_BAR       0.9668D+01 0.8973D+00
          C.V.(X_BAR)   0.1136D+00 0.6581D-02
         Sqrt[B*W(L,B)] 0.2941D+03 0.1581D+01
continue[y/n]?
```

Figure 5: Abbreviated LABATCH.2 screen display when SCREEN=1

As illustration of how this procedure works in practice, consider review 11 in Figure 5, where C.V.(X_BAR)=.1255 for mean waiting time (Series 1) and C.V.(X_BAR)=.008205 for probability of waiting (Series 2) may encourage one to conclude that the sample averages (X_BAR) are statistically reliable. By constrast, the variability in Sqrt[B*W(L,B)], the estimates of σ_∞, for say, reviews 9 through 11 encourages continuation of the experiment; hence "y" in response to the query, produces the output on review 12.

REFERENCES

Alexopoulos, C., G.S. Fishman, and A.T. Seila. 1997. Computational experience with the batch means method, *Proceedings of the 1997 Winter Simulation Conference*, S. Andradóttir, K.J. Healy, D.H. Withers, and B.L. Nelson eds., IEEE, Piscataway, NJ.

Blomqvist, N. 1967. The covariance function of the M/G/1 queueing system, Skandinavisk Aktuarietidskrift, **50**, 157-174.

Chien, C.H. 1989. Small-sample theory for steady state confidence intervals, Technical Report 37, Department of Operations Research, Stanford University.

Damerdji, H. 1994. Strong consistency of the variance estimator in steady-state simulation output analysis, *Math. Oper. Res.*, **19**, 494-512.

Fishman, G.S. 1978. *Principles of Discrete Event Simulation*, Wiley, New York.

Fishman, G.S. 1996. *Monte Carlo: Concepts, Algorithms, and Applications*, Springer-Verlag, New York.

Fishman, G.S. 1997. LABATCH.2 for Analyzing Sample Path Data, Technical Report No. 97/04, Operations Research Department, University of North Carolina at Chapel Hill.

Fishman, G.S. and L.S. Yarberry. 1997. An implementation of the batch means method, *INFORMS J. on Comput.*, **9**, 296-310.

Goldsman, D. and M.S. Meketon. 1986. A comparison of several variance estimators for stationary increment stochastic process, Technical Report, School of Industrial and Systems Engineering, Georgia Institute of Technology.

Philipp, W. and W. Stout. 1975. Almost sure invariance principles for partial sums of weakly dependent random variables. *Mem. Am. Math. Soc.*, **161**.

Song, W.T. and B.W. Schmeiser. 1995. Optimal mean-squared-error batch sizes, *Man. Sci.*, **41**, 110-123.

von Neumann, J. 1941. Distribution of the ratio of the mean square successive difference and the variance, *Ann. Math. Statist.*, **12**, 367-395.

Yarberry, L.S. 1993. Incorporating a dynamic batch size selection mechanism in a fixed - sample-size batch means procedure, unpublished Ph.D. dissertation, Dept. of Operations Research, University of North Carolina, Chapel Hill.

Young, L.C. 1941. On randomness in ordered sequences, *Ann. Math. Statist.*, **12**, 293-300.

AUTHOR BIOGRAPHY

GEORGE S. FISHMAN is professor of operations research at the University of North Carolina at Chapel Hill. His areas of research include Monte Carlo methods and discrete-event simulation with current focus on Gibbs sampling, fast manufacturing simulation, and fluid flow discrete-event simulation models. He has contributed to the literature on the design of simulation experiments, sample path data analysis, pseudorandom number generation and non-uniform variate generation. His most recent book, “Monte Carlo: Concepts, Algorithms, and Applications", published by Springer-Verlag in 1996, was awarded the Lanchester Prize by INFORMS as its outstanding publication in 1996 in operations research and management science in the English language. It was also awarded the prize of the INFORMS College on Simulation for its 1997 outstanding publication in simulation.