COVERAGE ERROR FOR CONFIDENCE INTERVALS ARISING IN SIMULATION OUTPUT ANALYSIS†

Peter W. Glynn

Stanford University

Abstract

Coverage error asymptotics for confidence intervals arising in simulation are discussed. Asymptotic expansions, to order $O(n^{-1})$ (n is the sample size), are given for confidence intervals associated with sequences of independent and identically distributed random variables, as well as regenerative processes. Implications for simulation are emphasized.

1. INTRODUCTION

One of the major problems that arises in the statistical analysis of simulation output is the generation of confidence intervals for parameters of interest. However, a major practical obstacle remains: Coverage rates tend to be substantially lower than the confidence level indicated. This phenomenon manifests itself even in those cases in which the procedures have an asymptotically consistent large sample theory. For a discussion of this problem, we refer the reader to Section 8.5.1 of Law and Kelton (1982).

In this paper, we will study asymptotic expansions associated with the error of the coverage probability. We begin, in Section 2, with a discussion relating to the interpretation of a confidence interval. The asymptotics associated with confidence interval error turn out to be interpretation-dependent. In Section 3, we consider error expansions for confidence intervals for the mean of a sequence of independent and identically distributed (i.i.d.) random variables. This case is of interest when the simulator is concerned with analyzing the output of a terminating simulation. It is also intimately connected to the application of the technique of replication to the steady-state confidence interval problem. In this context, we are able to precisely identify the primary sources of error in the converage rate, and we discuss the relevance of the error expansion to the choice between using a t-variate or normal in the confidence interval procedure. In Section 4, we examine the regenerative confidence interval that arises when a simulator is interested in the steady-state mean of a regenerative stochastic process. We show that the coverage error has a form similar to that in the i.i.d. case. Our results indicate that the coverage rate difficulty inherent in the regenerative confidence interval is of the same order of magnitude as the coverage rate error faced in the

---

†This research was supported by NSF Grant MCS79-09139, ONR Contract N00014-76-C-0578 (NR 042-343) and a Natural Sciences and Engineering Research Council of Canada Postgraduate Scholarship.

Dr. Peter W. Glynn, Department of Industrial Engineering, University of Wisconsin-Madison, 1513 University Avenue, Madison, WI 53706

Proceedings of the 1982 Winter Simulation Conference
Highland * Chao * Madrigal, Editors

82CH1844-0/82/0000-0369 $00.75 © 1982 IEEE

classical i.i.d. case. Finally in Section 5, we state some conclusions, and briefly discuss two new small sample confidence interval methodologies that appear to have enhanced coverage properties.

2. THE INTERPRETATION OF A CONFIDENCE INTERVAL

Let $\{X_n;\ n \geq 0\}$ be a sequence of random variables representing the output of a simulation, and suppose that a confidence interval for the parameter θ is required. The goal of the simulator is to find a sequence of random variables

$$L_n = L_n(X_0, \ldots, X_n)$$

$$R_n = R_n(X_0, \ldots, X_n)$$

such that the random interval $[L_n,R_n]$ corresponds to a "confidence" set for θ. In precise terms, $[L_n,R_n]$ is a $100(1-\alpha)\%$ confidence interval for θ if

$$P\{\theta \in [L_n,R_n]\} = 1-\alpha \quad . \qquad (2.1)$$

However, (2.1) does not fully specify the choice of L_n and R_n, as the following example shows.

(2.2) **Example.** Let $\{X_n;\ n \geq 0\}$ be normal random variables with unknown mean θ and known variance σ^2, denoted $N(\theta,\sigma^2)$. Let $\Phi(x) = P\{N(0,1) \leq x\}$, and put $z(p) = \Phi^{-1}(p)$ for $0 < p < 1$. Then, it is easily verified that

$$[\bar{X}_n - z(p+1-\alpha)\sigma/n^{1/2}, \quad \bar{X}_n - z(p)\sigma/n^{1/2}] \ ,$$

where $\bar{X}_n = \sum_{i=1}^{n} X_i/n$, is a $100(1-\alpha)\%$ confidence interval for θ, provided $0 < p < \alpha$.

In the above example, most simulators would agree that the "correct" choice of p is $\alpha/2$. There are two reasons for that choice. First, the length of the interval is minimal for $p = \alpha/2$. Secondly, the interval $[L_n,R_n]$ obtained through that choice has the property that

$$P\{\theta \geq L_n\} = P\{\theta \leq R_n\} = 1-\alpha/2 \ , \qquad (2.3)$$

$$P\{\theta < L_n\} = P\{\theta > R_n\} = \alpha/2 \ .$$

We shall call an interval $[L_n,R_n]$ satisfying (2.3) a $100(1-\alpha)\%$ balanced confidence interval for θ. Observe that (2.3) implies (2.1) so that any balanced confidence interval is a confidence interval in the sense of (2.1).

In many simulation problems, it appears that the simulation practitioner would have a preference for a balanced interval. For example, suppose that a simulation of a queueing system produces a 90% confidence interval $[L_n,R_n]$ for the mean customer waiting time θ. The simulator can conclude only that

$$0 \leq P\{\theta < R_n\},\ P\{\theta > L_n\} \leq 0.1$$

whereas, for a balanced interval, the conclusion

$$P\{\theta > R_n\} = P\{\theta < L_n\} = 0.05$$

is possible. Clearly, a balanced interval gives more information to the simulator.

It should be pointed out that much of the statistical theory of confidence intervals in a parametric setting relates to balanced intervals. For a discussion of the desirability of balanced intervals from a Bayesian viewpoint, see Efron (1981).

3. COVERAGE RATES FOR I.I.D. CONFIDENCE INTERVALS

Let $\{X_n;\ n \geq 0\}$ be a sequence of i.i.d. random variables with $0 < \sigma^2 = \text{var}(X_n) < \infty$. Let us first examine the results in the simplest possible setting: The goal is to produce a confidence interval for $\mu = EX_n$, given that σ^2 is known.

The standard confidence interval procedure, in this case, starts from a Central Limit Theorem (CLT) for $n^{1/2}\ (\bar{X}_n-\mu)/\sigma$. Simple algebraic manipulation shows that $[L_n(p),\ R_n(p)]$ is an approximate $100(1-\alpha)\%$ confidence interval for μ, where

$$L_n(p) = \bar{X}_n - z(p+1-\alpha)\sigma/n^{1/2}$$

$$R_n(p) = \bar{X}_n - z(p)\sigma/n^{1/2}$$

for $0 < p < \alpha$.

Before proceeding, let us observe that the coverage rate error in the interval $[L\ (p),\ R_n(p)]$, denoted $\varepsilon_n(p)$, is given by $\left|P\{\mu_n^n \in [L_n(p),\ R_n(p)]\} - (1-\alpha)\right|$. As for the balanced situation, it is clear that the interval $[L_n(p),\ R_n(p)]$ is asymptotically balanced only if $p = \alpha/2$. Thus, we shall henceforth restrict our discussion of balanced error to this case, and designate $(\varepsilon_n^\ell, \varepsilon_n^r)$ as our error descriptor, where

$$\varepsilon_n^\ell = \left|P\{\mu < L_n(\alpha/2)\} - \alpha/2\right|$$

$$\varepsilon_n^r = \left|P\{\mu > R_n(\alpha/2)\} - \alpha/2\right|$$

Our error estimates will be written in terms of the "big O", "little o" notation (i.e., $g(n) = O(f(n))$ if $g(n)/f(n)$ is bounded; $g(n) = o(f(n))$ if $g(n)/f(n)$ goes to zero).

(3.1) **Theorem.** (i) If $EX_n^4 < \infty$, then the error terms $\varepsilon_n(p)$, ε_n^ℓ, ε_n^r are all $O(n^{-1/2})$.

(ii) If, in addition, X_n has a distribution with a Lebesgue density component, then

a) $$\varepsilon_n(p) = \left|Sk(X_0)\cdot(g(z(p)) - g(z(p+1-\alpha)))\right|/6n^{1/2} + O(\tfrac{1}{n}),$$

b) $$\varepsilon_n^\ell = \left|Sk(X_0)\cdot g(z(\alpha/2))\right|/6n^{1/2} + O(\tfrac{1}{n}) = \varepsilon_n^r,$$

where

$$g(x) = (1-x^2)d/dx(\Phi(x)) ,$$

and

$$Sk(X_0) = E(X_0-\mu)^3/\sigma(X_0)^3.$$

Proof. Part (i) of the theorem is a direct consequence of the Berry-Esseen theorem (see Feller (1971), p. 542). The proof of the second half of the result follows from simple algebraic manipulation of the expansions cited in Theorems 1 and 3 of Section 16.4 of Feller (1971).

The first part of the theorem tells us that under an appropriate moment assumption, the error $\varepsilon_n(p)$, ε_n^ℓ, and ε_n^r decrease at least as fast as $n^{-1/2}$, for all p. The second half of the result shows that when X_n has a density component, then $\varepsilon_n(\alpha/2)$ is $O(1/n)$, whereas for $p \neq \alpha/2$, $\varepsilon_n(p)$ is $O(n^{-1/2})$. On the other hand, the balanced error for $[L_n(\alpha/2), R_n(\alpha/2)]$ is always $O(n^{-1/2})$.

Thus, using $p = \alpha/2$ in the interval $[L_n(p), R_n(p)]$ produces a confidence interval with the "best" coverage rate (i.e., $O(1/n)$ as opposed to $O(n^{-1/2})$), as well as the asymptotically minimal length. However, $[L_n(\alpha/2), R_n(\alpha/2)]$ achieves this result in a very curious way. Part (ii.b) of Theorem 3.1 shows that $P\{\mu \leq L_n(\alpha/2)\}$ will differ from $\alpha/2$ by $O(n^{-1/2})$. Hence, if $P\{\mu \leq L_n(\alpha/2)\}$ is greater (say) than $\alpha/2$ by $O(n^{-1/2})$, then $P\{\mu \geq R_n(\alpha/2)\}$ must be less than $\alpha/2$ by precisely the same $O(n^{-1/2})$ error (in order that $\varepsilon_n(\alpha/2)$ be $O(1/n)$). This suggests, in this case, that a "better" confidence interval would be achieved by shifting the interval to the left slightly. It is also worth observing that the coverage rate accuracy of $p = \alpha/2$ is highly unstable in the sense that $\varepsilon_n(\alpha/2+\eta)/\varepsilon_n(\alpha/2) \to \infty$ for any non-zero η.

Some caution should be exercised in trying to extend the second conclusion of the theorem to the case of discrete random variables. The result depends on an Edgeworth expansion which breaks down in the discrete case (see [3], p. 539, for a related comment).

Our above analysis required that σ^2 be known. Of course, in general the simulator does not know this parameter, and hence it must be estimated from the simulation output sequence $\{X_n\}$. The key result in forming confidence intervals in this setting is the CLT for t_n, where $t_n = n^{1/2}(\bar{X}_n-\mu)/s_n$ and

$$s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n} X_i^2 - (\bar{X}_n)^2 .$$

The intervals $[L_n(p), R_n(p)]$ are approximate $100(1-\alpha)\%$ confidence intervals for $0 < p < \alpha$, provided

$$L_n(p) = \bar{X}_n - z(p+1-\alpha)s_n/n^{1/2},$$

$$R_n(p) = \bar{X}_n - z(p)s_n/n^{1/2}.$$

(3.2) **Theorem.** (i) If $EX_n^8 < \infty$, then the error terms $\varepsilon_n(p)$, ε_n^{ℓ}, ε_n^{r} are all $O(n^{-1/2})$.

(ii) If, in addition, X_n has a distribution with a Lebesgue density component, then

a) $$\varepsilon_n(p) = \Big|Sk(X_0)\cdot(h(z(p)) - h(z(p+1-\alpha)))\Big|/n^{1/2} + O(\tfrac{1}{n}) ,$$

b) $$\varepsilon_n^{\ell} = \Big|Sk(X_0)\cdot h(z(\alpha/2))/n^{1/2} + O(\tfrac{1}{n}) = \varepsilon_n^{r} ,$$

where $h(x) = [(1+2x)d/dx(\Phi(x))]/6$.

The proof of this theorem is similar to that of Theorem 3.1. The key step is to prove a Berry-Essen type result and obtain an Edgeworth expansion for t_n; this can be found in Glynn (1982).

Notice that the conclusions of this theorem are qualitatively very similar to those of Theorem 3.1, where σ_2 was known. For example, $\varepsilon_n(\alpha/2)$ is again $O(1/n)$, whereas ε_n^{ℓ}, ε_n^{r} are both $O(n^{-1/2})$. However, the coefficient in $n^{-1/2}$ has changed. The proof of the Edgeworth expansion mentioned above shows that the error of order $n^{-1/2}$ occurs precisely as a consequence of the skewness $Sk(X_0)$) of $\{X_n; n \geq 0\}$, and the correlation between s_n^2 and $\bar{X}_n$. This is in contrast to the situation where σ^2 is known, in which case all the $n^{-1/2}$ error emanates from skewness alone. Recall that the coefficient of skewness $Sk(X_0)$ measures the asymmetry of a distribution.

Before concluding this section, we turn to the question of using t-variates rather than a normal distribution to generate confidence intervals. Let T_k have a Student's t-distribution with k degrees of freedom. Put $\Gamma_k(x) = P\{T_k \leq x\}$, and set $z_k(p) = \Gamma_k^{-1}(p)$. Many simulators (e.g., [6], p. 288) advise the use of the interval $[\bar{L}_n(p), \bar{R}_n(p)]$, where

$$L_n(p) = \bar{X}_n - z_{n-1}(p+1-\alpha)s_n/n^{1/2},$$

$$\bar{R}_n(p) = \bar{X}_n - z_{n-1}(p)s_n/n^{1/2},$$

rather than the previously defined interval for μ. Let $\bar{\varepsilon}_n(p)$, $\bar{\varepsilon}_n^{\ell}$, $\bar{\varepsilon}_n^{r}$ be the coverage rate error and balancing errors associated with the t-variate confidence interval.

(3.3) **Theorem.** For $0 < p < \alpha$,

(i) $$\bar{\varepsilon}_n(p) = \varepsilon_n(p) + o(n^{-1/2})$$

(ii) $$\bar{\varepsilon}_n^{\ell}(p) = \varepsilon_n^{\ell}(p) + o(n^{-1/2})$$

$$\bar{\varepsilon}_n^{r}(p) = \varepsilon_n^{r}(p) + o(n^{-1/2})$$

The proof of this result can be found in [4]. The content of this theorem is the that t-variate modification is of "small order" in the sense that the leading term in the error (of order $n^{-1/2}$) is precisely the same as that obtained via normal theory. However, it should be mentioned that the coverage error $\bar{\varepsilon}_n(\alpha/2)$ may differ from $\varepsilon_n(\alpha/2)$ in its leading term. In any case, it seems clear that the desirability of using a t-variate, as opposed to a normal, deserves further study.

4. COVERAGE RATES FOR REGENERATIVE CONFIDENCE INTERVALS

Let $\{X_t; t \geq 0\}$ be a regenerative stochastic process. Then, there exist random times $T_1, T_2, \ldots$ such that for any (measurable) real-valued function f, $\{(Y_k, \tau_k); k \geq 1\}$ is i.i.d., where

$$Y_k = \int_{T_k}^{T_{k+1}} f(X_s)\, ds$$

$$\tau_k = T_{k+1} - T_k .$$

The goal of the simulator is to find a confidence interval for the steady-state mean of $\{f(X_t); t \geq 0\}$. It can be shown that this is equivalent to obtaining an interval for $r = EY_1/E\tau_1$ (see Crane and Iglehart (1975)). The regenerative method for output analysis depends on a CLT, under the assumption $0 < \sigma^2(Y_k - r\tau_k) < \infty$, $E\tau_k < \infty$, for the statistic $n^{1/2}(r_n - r)/v_n$, where

$$r_n = \sum_{k=1}^{n} Y_k / \sum_{k=1}^{n} \tau_k$$

$$v_n^2 = \frac{n}{(T_n - T_1)^2} \left(\sum_{k=1}^{n} Y_k^2 - 2r_n \sum_{k=1}^{n} Y_k \tau_k + r_n^2 \sum_{k=1}^{n} \tau_k^2 \right)$$

As for the i.i.d. case of Section 3, the normal approximation yields $[L_n(p), R_n(p)]$ as an approximate $100(1-\alpha)\%$ confidence interval for r, where

$$L_n(p) = r_n - z(p+1-\alpha) v_n / n^{1/2},$$

$$R_n(p) = r_n - z(p) v_n / n^{1/2}.$$

(4.1) **Theorem.** i) If $EY_n^8 < \infty$ and $E\tau_n^8 < \infty$, then the error terms $\epsilon_n(p)$, ϵ_n^{ℓ}, ϵ_n^r are all $O(n^{-1/2})$.

(ii) If, in addition, the distribution of (Y_n, τ_n) has a component which is a Lebesgue density in the plane, then

a) $\epsilon_n(p) = \left| k(z(p)) - k(z(p+1-\alpha)) \right| / n^{1/2} + O(1/n).$

b) $\epsilon_n^{\ell} = \left| k(z(\alpha/2)) \right| + O(1/n) = \epsilon_n^r$

where

$$k(x) = (\alpha + \beta x^2)\, d/dx(\Phi(x)) ,$$

$$\alpha = EZ_1^3 / 6\sigma^3(Z_1)$$

$$\beta = EZ_1^3 / 3\sigma^3(Z_1) - E\tau_1 Z_1 / \sigma(Z_1) E\tau_1$$

$$Z_k = Y_k - r\tau_k .$$

This theorem is a consequence of a Berry-Esseen result and Edgeworth expansion for regenerative processes, which appears in [4].

The important point here is the similarity to the i.i.d. case. The order of error is precisely the same as that which we encountered in Section 3. Hence, it could be argued, on the basis of error comparison, that the steady-state simulation problem is no more difficult than the terminating simulation problem, provided that the regenerative method is used.

As for the form of the error coefficient in $n^{-1/2}$, we observe that $Sk(Z_1)$ plays an important role, as in the i.i.d. case. Additional terms in $E\tau_1 Z_1 / E\tau_1 \cdot \sigma(Z_1)$ also appear, however: These are contributed by the bias of r_n. Thus, the prime sources of error in the regenerative case are asymmetry effects, correlation between point and variance estimates, and bias problems in the point estimate. This has an important implication for research efforts directed at producing "correct" coverage rates for confidence intervals. It shows that reducing one source of error, such as bias in the point estimate, should not be expected to necessarily improve coverage.

5. CONCLUDING REMARKS

In this paper, we have provided an overview of the coverage error problem for confidence interval generation in simulation. We have shown that the qualitative character of the error appears reasonably insensitive to the "fine" structure of the simulation output sequence. The regenerative case indicates that point estimate bias, correlation between point and variance estimators, and asymmetry all play equally important roles in determining coverage rates in the steady-state simulation problem. Our results also show the need for more research on the question of whether t-variates or normals should be used to generate intervals. Furthermore, we have shown that the parameter p plays a critical role in determining the amount of error. However, we caution the reader that although $p = \alpha/2$ appears optimal in the sense of error, the result is highly unstable. Finally, we have introduced the concept of a balanced confidence interval and shown that its error asymptotics are somewhat different from the standard error criterion.

In [4], we study two procedures that appear to have promising coverage characteristics. The first technique is a regenerative bootstrap (see [2] for the bootstrap in the i.i.d. case), and the second method involves an application of a so-called "normalization" transformation. The latter procedure is based on an idea of Johnson (1978). The asymptotic error expansion for these intervals indicates an improvement over currently used intervals. These improvements have also manifested themselves empirically.

ACKNOWLEDGEMENTS

The author would like to thank Professors Bradley Efron and Donald L. Iglehart for stimulating conversations on the topics covered in this paper.

BIBLIOGRAPHY

1. Crane, M. and D.L. Iglehart (1975). Simulating Stable Stochastic Systems: III. Regenerative Processes and Discrete-Event Simulations. Operations Research 23, 33-45.

2. Efron, B. (1981). Nonparametric Standard Errors and Confidence Intervals. Canad. J. Statist. 9, 139-172.

3. Feller, W. (1971). An Introduction to Probability Theory and its Applications, Volume II. Wiley, New York.

4. Glynn, P.W. (1982). Simulation Output Analysis for General State Space Markov Chains. Ph.D. dissertation, Department of Operations Research, Stanford University.

5. Johnson, N. (1978). Modified t Tests and Confidence Intervals for Asymmetrical Populations. J. Amer. Statist. Assoc. 73, 536-544.

6. Law, A.M. and W.D. Kelton (1982). Simulation Modeling and Analysis. McGraw-Hill, New York.