Proceedings of the 2007 Winter Simulation Conference
S. G. Henderson, B. Biller, M.-H. Hsieh, J. Shortle, J. D. Tew, and R. R. Barton, eds.

SINGLE-STAGE MULTIPLE-COMPARISON PROCEDURE FOR QUANTILES AND OTHER PARAMETERS

Marvin K. Nakayama

Computer Science Department
New Jersey Institute of Technology
Newark, N.J. 07102, U.S.A.

ABSTRACT

We present a single-stage multiple-comparison procedure for comparing parameters of independent systems, where the parameters are not necessarily means or steady-state means. We assume that for each system, the parameter has an estimation process that satisfies a central limit theorem (CLT) and that we have a consistent variance-estimation process for the variance parameter appearing in the CLT. The procedure allows for unequal run lengths or sample sizes across systems, and also allows for unequal and unknown variance parameters across systems. The procedure is asymptotically valid as the run lengths or sample sizes of all system grow large. One setting the framework encompasses is comparing quantiles of independent populations. It also covers comparing means or other moments of independent populations, functions of means, and steady-state means of stochastic processes.

1 INTRODUCTION

Suppose that we have k systems, labeled $1,2,\ldots,k$, that are to be compared relative to a given parameter. Let θ_i denote the value of the parameter for system i. For example, suppose there are 10 possible designs for a fault-tolerant computing system, and we want to compare the alternatives in terms of their 0.9-quantile of the time to failure.

This paper presents a single-stage multiple-comparison procedure to compare $\theta_1,\theta_2,\ldots,\theta_k$, where we assume larger values of θ_i are better. Specifically, we consider multiple comparisons with the best (MCB, Hsu 1984), which produces simultaneous confidence intervals for $\theta_i - \max_{j\neq i}\theta_j$ for $i=1,2,\ldots,k$. Note that if $\theta_i - \max_{j\neq i}\theta_j > 0$, then system i is the best. In some situations when secondary considerations (e.g., ease of installation) are taken into account, one may opt to implement a non-optimal design, as long as it is "close enough" to the best. For example, we may choose a system i with $\theta_i - \max_{j\neq i}\theta_j < 0$ as long as $\theta_i - \max_{j\neq i}\theta_j > -\delta$ for some specified $\delta > 0$, so system i is not the best but is within δ of the best.

As the run length or sample size of each system grows to infinity, our MCB procedure is asymptotically valid under the following two assumptions. First we assume that for each system i, θ_i has an estimation process that satisfies a central limit theorem (CLT). The CLT has a variance parameter σ_i^2, and we require that we have a consistent estimator for σ_i^2. We allow for the run lengths or sample sizes of the different systems to differ, and we allow for the σ_i^2, $i=1,2,\ldots,k$, to be unequal and unknown.

Most previous work on multiple-comparison procedures (Hochberg and Tamhane 1987, Swisher, Jacobson, and Yucesan 2003, Kim and Nelson 2006) assumes that θ_i is the mean of a normally distributed population, and that independent and identically distributed (i.i.d.) sampling is used within each population. Exceptions include Nakayama (1997) and Damerdji and Nakayama (1999), which compare steady-state means using standardized time series (STS) methods (Schruben 1983, Glynn and Iglehart 1990) for "estimating" σ_i^2. STS methods do not yield consistent estimates of the σ_i^2 when the run length grows large but the number of batches remains fixed, so the results in the current paper do not include the multiple-comparison methods using STS.

Our framework in the current paper allows for comparing systems relative to quite general parameters, including means or other moments of independent (not necessarily normally distributed) populations, functions of means, and steady-state means of stochastic processes. One specific example we consider here is comparing quantiles of independent populations.

The rest of the paper is organized as follows. Section 2 describes the mathematical framework we adopt. We present our MCB procedure in Section 3, and Section 4 shows explicitly how the procedure applies to comparing quantiles of independent populations. Section 5 presents empirical results comparing quantiles, and we summarize our findings in Section 6.

1-4244-1306-0/07/$25.00

2 MATHEMATICAL FRAMEWORK

Suppose that we have k systems, labeled $1,2,\ldots,k$, where the ith system has parameter θ_i. The goal is to compare the k systems in terms of the θ_i. We assume that for each i, we have an estimation process $\widehat{\theta}_i = [\widehat{\theta}_i(t) : t > 0]$ for θ_i, where $\widehat{\theta}_i(t)$ is the estimator of θ_i based on running a simulation of system i for a run length of t or taking t i.i.d. samples from population i, depending on the context. Let $N(a,b)$ denote a normal random variable with mean a and variance b, and let $\Rightarrow$ denote weak convergence (see Chapter 5 of Billingsley 1995). We assume that each $\widehat{\theta}_i$ satisfies a central limit theorem:

Assumption 1 *The estimation processes $\widehat{\theta}_1,\ldots,\widehat{\theta}_k$ are independent, and there exists a finite positive constant η such that for each i,*

$$t^\eta \left[\widehat{\theta}_i(t) - \theta_i\right] \Rightarrow N(0,\sigma_i^2) \tag{1}$$

as $t \to \infty$, where $0 < \sigma_i < \infty$ is a constant.

In most applications, the parameter η in Assumption 1 takes on the canonical value of $1/2$, but we do not require this. Glynn and Whitt (1992) list examples of processes satisfying the CLT in (1). For example, the CLT holds (under various conditions) for sample means or other sample moments of i.i.d. samples, functions of sample means, quantile estimators, and time-average rewards of stochastic processes having a steady state. Specifically, for the last example, let $X_i = [X_i(t) : t \geq 0]$ be a stochastic process on state space S_i representing the evolution over time of system i, and let $f_i : S_i \to \Re$ be a real-valued "reward" function on S_i. Then we can define $\widehat{\theta}_i(t) = (1/t)\int_0^t f_i(X_i(s))\,ds$, which is the time-average reward of X_i over the interval $[0,t]$. Under a wide variety of assumptions (Glynn and Iglehart 1990), $\widehat{\theta}_i$ satisfies the CLT in (1) with $\eta = 1/2$, and θ_i is the steady-state mean reward of X_i.

We call σ_i^2 in (1) the *variance parameter* for system i. We assume that for each i, we have a variance-estimation process $V_i = [V_i(t) : t > 0]$ that is consistent, in the following sense:

Assumption 2 *For each system i, $V_i(t) \Rightarrow \sigma_i^2$ as $t \to \infty$.*

In our previous example of comparing steady-state mean rewards, suppose that each process X_i is regenerative (Crane and Iglehart 1975), and let $0 \leq A_{i,0} < A_{i,1} < A_{i,2} < \cdots$ be the sequence of regeneration epochs of system i. For $j = 1,2,\ldots$, let $\tau_{i,j} = A_{i,j} - A_{i,j-1}$ be the length of the jth cycle of system i. Also, define $Y_{i,j} = \int_{A_{i,j-1}}^{A_{i,j}} f_i(X_i(s))\,ds$ to be the cumulative reward over the jth cycle of system i. Assume that $E[\tau_{i,1}] < \infty$ and that there exists a finite constant θ_i such that $E[Y_{i,1} - \theta_i\tau_{i,1}] = 0$ and $E[(Y_{i,1} - \theta_i\tau_{i,1})^2] < \infty$. Then $\theta_i = E[Y_{i,1}]/E[\tau_{i,1}]$ and the CLT in (1) holds with $\eta = 1/2$ and $\sigma_i^2 = E[(Y_{i,1} - \theta_i\tau_{i,1})^2]/E[\tau_{i,1}]$. Define $N_i(t) = \sup\{j \geq 0 : A_{i,j} \leq t\}$, which is the number of regenerative cycles that process i completes by time t. Then the variance estimator

$$V_i(t) = \frac{1}{t}\sum_{j=1}^{N_i(t)} \left[Y_{i,j} - \widehat{\theta}_i(t)\tau_{i,j}\right]^2$$

satisfies Assumption 2; see Glynn and Iglehart (1993) for details.

An example of a setting in which Assumption 1 holds for a non-canonical value of η is for estimators based on the Kiefer-Wolfowitz (1952) stochastic approximation algorithm. Ruppert (1982) shows that under certain regularity conditions, such estimators satisfy (1) with $\eta = 1/3$. Ventner (1967), p. 189, provides directions for constructing $V_i(t)$ such that Assumption 2 holds.

3 MCB PROCEDURE

We now present a MCB procedure for comparing the parameters θ_i, $i = 1,2,\ldots,k$, of independent systems.

1. Specify the number of systems $2 \leq k < \infty$, the confidence level $1-\alpha$ with $0 < \alpha < 1$, and $\bar{T} = (T_1,\ldots,T_k)$, where T_i is the run length (or sample size) for system i. Define the constant $\gamma = z_{(1-\alpha)^{1/(k-1)}}$, where z_β satisfies $\Phi(z_\beta) = \beta$ for $0 < \beta < 1$, and Φ is the distribution function of a standard (mean 0 and variance 1) normal distribution.
2. Run independent simulations of the k systems, where the simulation of system i has a run length of T_i.
3. For each system i, compute $V_i(T_i)$, and define the joint MCB intervals $I_i(\bar{T})$, $i = 1,2,\ldots,k$, for $\theta_i - \max_{j\neq i}\theta_j$, $i = 1,2,\ldots,k$, respectively, with

$$I_i(\bar{T}) = \left[-\left(\min_{\substack{j\in\mathscr{A}(\bar{T}),\\ j\neq i}}\left(\widehat{\theta}_i(T_i) - \widehat{\theta}_j(T_j) - W_{i,j}(\bar{T})\right)\right)^-, \left(\min_{j\neq i}\left(\widehat{\theta}_i(T_i) - \widehat{\theta}_j(T_j) + W_{i,j}(\bar{T})\right)\right)^+\right], \tag{2}$$

where $(x)^+ = \max(x,0)$, $-(x)^- = \min(x,0)$,

$$W_{i,j}(\bar{T}) = \gamma\sqrt{\frac{V_i(T_i)}{T_i^{2\eta}} + \frac{V_j(T_j)}{T_j^{2\eta}}},$$

and

$$\mathscr{A}(\bar{T}) = \left\{\ell : \min_{j\neq\ell}\left(\widehat{\theta}_\ell(T_\ell) - \widehat{\theta}_j(T_j) + W_{\ell,j}(\bar{T})\right) \geq 0\right\}.$$

In (2), we define $\min_{j\in\emptyset} x_j = 0$.

The following establishes the asymptotic validity of the MCB intervals in (2).

Theorem 1 *Suppose $T_i = \zeta_i T$ for $i = 1,2,\ldots,k$, where $\zeta_1 > 0,\ldots,\zeta_k > 0$, are any constants. If Assumptions 1 and 2 hold, then*

$$\lim_{T\to\infty} P\left\{\theta_i - \max_{j\neq i}\theta_j \in I_i(\bar{T}),\ i = 1,2,\ldots,k\right\} \geq 1-\alpha.$$

4 QUANTILES

We now specifically discuss the setting of comparing quantiles of independent populations, which covers the case of comparing quantiles in terminating simulations. For example, given k alternative designs for a manufacturing system that shuts down each night, we might be interested in comparing the alternatives in terms of the 0.9-quantiles of the average flow times of the first 10 jobs each day.

For $0 < y < 1$ and any distribution function G, we define the yth quantile of G to be $G^{-1}(y) \equiv \inf\{x : G(x) \geq y\}$. Let F_i be the distribution function of population i. Fix p with $0 < p < 1$, and let $\theta_i = F_i^{-1}(p)$ be the pth quantile of system i. For each system i, let $X_{i,1}, X_{i,2}, \ldots, X_{i,t}$ be t i.i.d. samples from F_i, and define the empirical distribution function $F_{i,t}(x) = (1/t)\sum_{j=1}^{t} 1\{X_{i,j} \leq x\}$, where $1\{A\}$ is the indicator function of the event A. Also, define

$$\widehat{\theta}_i(t) = F_{i,t}^{-1}(p), \tag{3}$$

the estimator of θ_i based on t samples from F_i.

Assume that for each i, F_i is differentiable at θ_i, and $F_i'(\theta_i) > 0$, where prime denotes derivative. Then the CLT in Assumption 1 holds with $\eta = 1/2$ and

$$\sigma_i^2 = \frac{p(1-p)}{[F_i'(\theta_i)]^2}; \tag{4}$$

e.g., see p. 77 of Serfling (1980). To develop an estimator for σ_i^2, define constants $q_{i,t}$ such that

$$q_{i,t} = p + \sqrt{\frac{p(1-p)}{t}} + o\left(\frac{1}{t^{1/2}}\right) \tag{5}$$

as $t \to \infty$, where $f(t) = o(g(t))$ for functions f and g means $f(t)/g(t) \to 0$ as $t \to \infty$. Then the variance estimator

$$V_i(t) = t\left[F_{i,t}^{-1}(q_{i,t}) - F_{i,t}^{-1}(p)\right]^2 \tag{6}$$

satisfies Assumption 2; e.g., see p. 94 of Serfling (1980). Thus, we can apply the MCB procedure in Section 3 with $\widehat{\theta}_i(t)$ and $V_i(t)$ defined in (3) and (6), respectively. In the current setting of comparing quantiles, T_i in the MCB procedure of Section 3 denotes the number of i.i.d. samples taken from population i.

5 EMPIRICAL RESULTS

We ran some experiments with our MCB procedure to compare the pth quantiles of independent populations. Specifically, we compared $k = 4$ exponentially distributed populations, where the ith population has distribution function $F_i(x) = 1 - e^{-x/\mu_i}$ for $x \geq 0$, so its mean is μ_i. In all our experiments, we fixed $\mu_1 = \mu_2 = \mu_3 = 1$, and we varied μ_4 over the values 1.1, 1.2, 2 and 5. Thus, the pth quantile of the ith population is $\theta_i = -\mu_i \ln(1-p)$, which we assume is unknown, and for any fixed p, population 4 is the best since θ_4 is the largest among $\theta_1,\ldots,\theta_4$, which we also assume is unknown. We used the same sample size $T_1 = \cdots = T_4 = T$ for each population, and we set T to be 20, 80 or 320. We varied the quantile level p between 0.1 and 0.9. From (4) we can calculate $\sigma_i^2 = \mu_i^2 p/(1-p)$, which we assume is unknown, and we estimated σ_i^2 in our experiments using $V_i(T)$ in (6) with $q_{i,T} = p + \sqrt{p(1-p)/T}$ in (5). We ran 10^4 independent replications for each set of parameters, where we constructed MCB intervals having nominal confidence level $1-\alpha = 0.9$ in each replication.

Table 1 gives the coverage results from our simulations. In all but one case, for a fixed sample size T and quantile level p, the coverage increases as μ_4 increases, which also corresponds to $\theta_4 - \theta_i$, $i = 1,2,3$, increasing. Moreover, for a fixed configuration of means and quantile level p, coverage increases as the sample size T for each population increases (i.e., as we go across a row), with the coverage levels close to or at the nominal level of 90% for the largest sample size $T = 320$. This agrees with our asymptotic theory in Theorem 1. In addition, considering the coverages for all p and means for each sample size T, we see that there is a larger range of coverages for smaller T than for larger T. This may indicate that quantile estimators and their variance estimators require large sample sizes to be accurate. Finally, comparing the coverages for the different quantile levels p for the same configuration of means and for same T, we see that the coverages are typically lower for $p = 0.1$ and 0.9 than for $p = 0.3$, 0.5 or 0.7 when T is either 20 or 80, which is when coverages are significantly below the nominal level of 0.9. This may indicate that extreme quantiles (i.e., when p is close to 0 or 1) are harder to estimate than those with p close to 0.5.

For comparison Table 2 presents results from running experiments comparing the means of the 4 exponential distributions, so now each parameter $\theta_i = \mu_i$. In this case the coverages are all close to the nominal level of $1-\alpha = 0.9$, even for the smallest sample size $T = 20$. Hence, the asymptotics for MCB intervals comparing means seem to take effect for smaller sample sizes than they do when comparing quantiles.

Table 1: Coverage results (in percents) for MCB intervals comparing pth quantiles of $k = 4$ exponential populations, with T samples from each population.

p	Means	T 20	80	320
0.1	1, 1, 1, 1.1	72.4	75.7	86.3
0.1	1, 1, 1, 1.2	72.6	76.1	86.7
0.1	1, 1, 1, 2	73.9	78.8	88.3
0.1	1, 1, 1, 5	75.7	81.6	90.1
0.3	1, 1, 1, 1.1	84.9	86.8	89.3
0.3	1, 1, 1, 1.2	85.1	87.1	89.5
0.3	1, 1, 1, 2	86.1	88.0	90.7
0.3	1, 1, 1, 5	87.2	89.6	92.4
0.5	1, 1, 1, 1.1	83.6	85.3	86.7
0.5	1, 1, 1, 1.2	83.8	85.7	87.0
0.5	1, 1, 1, 2	84.7	86.6	88.7
0.5	1, 1, 1, 5	85.3	88.4	90.6
0.7	1, 1, 1, 1.1	86.7	89.1	90.2
0.7	1, 1, 1, 1.2	86.6	89.2	90.5
0.7	1, 1, 1, 2	86.9	89.8	91.4
0.7	1, 1, 1, 5	87.0	90.9	92.4
0.9	1, 1, 1, 1.1	74.9	78.2	88.6
0.9	1, 1, 1, 1.2	74.8	78.3	88.9
0.9	1, 1, 1, 2	73.8	79.7	89.6
0.9	1, 1, 1, 5	72.8	80.3	90.8

Table 2: Coverage results (in percents) for MCB intervals comparing the means of $k = 4$ exponential populations, with T samples from each population.

Means	T 20	80	320
1, 1, 1, 1.1	91.3	91.6	91.7
1, 1, 1, 1.2	90.9	91.4	91.7
1, 1, 1, 2	88.9	91.0	92.4
1, 1, 1, 5	88.6	92.0	93.9

6 CONCLUSIONS

We presented a single-stage MCB procedure for comparing parameters $\theta_1, \ldots, \theta_k$ of independent systems. The procedure is asymptotically valid when the estimator of each θ_i satisfies a CLT and there is a consistent estimator of the variance parameter appearing in the CLT. The procedure allows for unequal run lengths across systems, and unknown and unequal variances. This framework encompasses comparing means of populations, functions of means, quantiles, and steady-state means.

We consider two-stage MCB procedures in Nakayama (2006) for the same general framework as in the current paper. The two-stage procedures have random run lengths for each system, and to establish the asymptotic validity of these procedures, we require a slight strengthening of the CLT assumption so that the CLT also holds when applied at the random time.

REFERENCES

Billingsley, P. 1995. *Probability and measure*. Third ed. New York: John Wiley & Sons.

Crane, M., and D. L. Iglehart. 1975. Simulating stable stochastic systems, III: Regenerative processes and discrete-event simulations. *Operations Research* 23:33–45.

Damerdji, H., and M. K. Nakayama. 1999. Two-stage multiple-comparison procedures for steady-state simulations. *ACM Transactions on Modeling and Computer Simulations* 9:1–30.

Glynn, P. W., and D. L. Iglehart. 1990. Simulation output analysis using standardized time series. *Mathematics of Operations Research* 15:1–16.

Glynn, P. W., and D. L. Iglehart. 1993. Conditions for the applicability of the regenerative method. *Management Science* 39:1108–1111.

Glynn, P. W., and W. Whitt. 1992. The asymptotic validity of sequential stopping rules in stochastic simulations. *Annals of Applied Probability* 2:180–198.

Hochberg, Y., and A. C. Tamhane. 1987. *Multiple comparison procedures*. New York: Wiley.

Hsu, J. C. 1984. Constrained simultaneous confidence intervals for multiple comparisons with the best. *Annals of Statistics* 12:1136–1144.

Kiefer, J., and J. Wolfowitz. 1952. Stochastic estimation of the maximum of a regression function. *Annals of Mathematical Statistics* 23:462–466.

Kim, S.-H., and B. L. Nelson. 2006. Selecting the best system. In *Elsevier Handbooks in Operations Research and Management Science: Simulation*. Amsterdam: Elsevier.

Nakayama, M. K. 1997. Multiple-comparison procedures for steady-state simulations. *Annals of Statistics* 25:2433–2450.

Nakayama, M. K. 2006. Two-stage selection and multiple-comparison procedures for simulations with weakly consistent variance estimators. In preparation.

Ruppert, D. 1982. Almost sure approximations to the Robbins-Munro and Kiefer-Wolfowitz processes with dependent noise. *Annals of Probability* 10:178–187.

Schruben, L. W. 1983. Confidence interval estimation using standardized time series. *Operations Research* 31:1090–1108.

Serfling, R. J. 1980. *Approximation theorems of mathematical statistics*. New York: John Wiley & Sons.

Swisher, J. R., S. H. Jacobson, and E. Yucesan. 2003. Discrete-event simulation optimization using ranking, selection, and multiple comparison procedures: A survey. *ACM Transactions on Modeling and Computer Simulation* 13:134–154.

Ventner, J. H. 1967. An extension of the Robbins-Munro procedure. *Annals of Mathematical Statistics* 38:181–190.

AUTHOR BIOGRAPHY

MARVIN K. NAKAYAMA is an associate professor in the Department of Computer Science at the New Jersey Institute of Technology. He received a Ph.D. in operations research from Stanford University. He won second prize in the 1992 George E. Nicholson Student Paper Competition sponsored by INFORMS and is a recipient of a CAREER Award from the National Science Foundation. He is the stochastic models area editor for *ACM Transactions on Modeling and Computer Simulation* and the simulation are editor for *INFORMS Journal on Computing*. His research interests include applied probability, statistics, simulation and modeling. His e-mail address is ⟨marvin@njit.edu⟩, and his web page is ⟨web.njit.edu/~marvin⟩.