Sequential Procedure for Simultaneous Estimation of Several Percentiles

Kimmo E. E. Raatikainen
University of Helsinki, Department of Computer Science
Teollisuuskatu 23, SF-00510 Helsinki, FINLAND
e-mail: Kimmo.Raatikainen@Helsinki.FI
Telefax: + 358 0 7084441

Abstract

Percentiles are convenient indices to characterize the entire range of the values of simulation outputs. However, they have only seldom been used in simulation studies. One of the reason has been the lack of sequential estimation procedures, which are needed to obtain estimates of predefined accuracy. In this paper we introduce a sequential procedure for estimating several percentiles simultaneously.

The procedure uses the *extended P^2 algorithm* to estimate the percentiles. The variances of the percentile-estimates are estimated using a spectral method. Since the method applies a Bonferroni inequality, the covariances between the percentile-estimates are not needed. The procedure is shown to produce estimates having the predefined accuracy in eight queueing network models, representing multiprogrammed and time-shared computer systems.

Key words: *Percentile Estimation, Variance Estimation, Run Length Control*

1. Introduction

In most simulation studies the mean values of output data have been analyzed. However, in many situations the means provide an insufficent or even a misleading characterization of the output data. On the contrary, a suitably selected set of percentiles reflects all the essential distributional features of the phenomenon analyzed by simulation. Percentiles have only seldom been used in

*Manuscript of paper published in *Transactions of the Society for Computer Simulation 7,1* (March 1990): 21–44

simulation studies. We believe that the reason is the complexity of percentile estimation. The calculation of estimates has been regarded as a too cumbersome task in discrete event simulation. The accuracy estimation is not a simple task since the output sequence is, in general, autocorrelated. Therefore, sophisticated approaches are necessary to obtain an estimate of the accuracy. Only a few papers have been published about the estimation of percentiles in simulation context. Methods of estimating a single percentile and its variance have been introduced in [9, 15, 20, 21, 7]. All these methods are based on a fixed, predefined length of output sequence.

One of the basic problems in sequential estimation of percentiles is that the whole output sequence must be stored and (at least partially) sorted if the estimates are based on the order statistics. The *P^2 algorithm* [10] solves the problem if a single percentile is to be estimated. The algorithm estimates single percentiles without storing and sorting the observations. In [17] we introduced the *extended P^2 algorithm* that estimates several percentiles simultaneously. The extension is important since a single percentile estimate is sufficient only in a few specific situations.

The use of sequential estimation procedures where the length of a simulation run is not predefined is of practical importance. Estimates produced by a simulation usually have an accuracy requirement determined by the application. The analyst wants to obtain estimates, which meet his or her accuracy requirement. Running the simulation less than it is necessary to satisfy the accuracy requirement would not provide the information needed. On the other hand, longer runs would be a waste of computing time; see also the conclusions in [14].

In one-dimensional situations the accuracy requirement is usually given as the maximum relative half-width of the confidence interval. In [18] we proposed a sequential procedure for simultaneous estimation of several percentiles. There we generalized the relative half-width criterion to multidimensional situations using a relative Eucledian distance. In this paper we will introduce another sequential procedure based on another accuracy requirement.

When a one-dimensional estimate $\hat{\theta}$ satisfies the given maximum relative half-width criterion of the confidence interval (ε, α), we have a confidence of at least $1-\alpha$ that the true, unknown value θ lies in the interval $[\,(1-\varepsilon)\hat{\theta}\,,\,(1+\varepsilon)\hat{\theta}\,]$. In many situations the components of a multidimensional estimate, $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_m)'$, are of interest both as single independent values and as a vector. In such a situation it is natural to require that all the components of the vector satisfy simultaneously a relative half-width criterion.

In this study we apply the following accuracy requirement: *The estimate $\hat{\boldsymbol{\theta}}$ is accuracte enough, if we have a confidence of at least $1-\alpha$ that the components of the true unknown vector $\boldsymbol{\theta}$ are in the intervals* $[(1-\varepsilon_1)\hat{\theta}_1, (1+\varepsilon_1)\hat{\theta}_1], \ldots, [(1-\varepsilon_m)\hat{\theta}_m, (1+\varepsilon_m)\hat{\theta}_m]$ or, in other words

$$\Pr\{|\theta_1 - \hat{\theta}_1| \le \varepsilon_1|\hat{\theta}_1|, \ldots, |\theta_m - \hat{\theta}_m| \le \varepsilon_m|\hat{\theta}_m|\} \ge 1-\alpha\ .$$

Hence, the accuracy requirement is specified by $(\varepsilon_1, \ldots, \varepsilon_m, \alpha)$.

The current procedure has three important advantages over that given in [18], which was to best of our knowledge the first sequential procedure for simultaneous estimation of several percentiles. First, only variances of the percentile estimates, not the covariances between them, need to be estimated. This is due to the applied accuracy requirement allowing the use of a Bonferroni inequality. Secondly, the sequence length need not be double between successive tests of termination condition, since we use segmentation in the variance estimation. The third advantage is in the assumptions about the stochastic properties of the output sequence. The procedure given in [18] assumed that the percentile estimates have a multinormal limiting distribution. The current procedure assumes only that each marginal distribution is asymptotically normal. Hence the usual ϕ-mixing condition, not a stronger one, should be satisfied. The method of estimating the percentiles, and their variances and evaluating the accuracy measure are described in the following section.

The simulation experiments are reported in the last section. We analyzed the response time sequences from simulations of eight different queueing network models. The analyzed models are networks of 2–6 servers that represent multiprogrammed and time-shared computer systems. The experiments were performed in three phases. The first two of them were initial verifications of the *extended P^2 algorithm* and the method of estimating the variances. In the last phases we applied the proposed sequential procedure to estimate the 50^{th}, 75^{th}, and 90^{th} percentile of the response times from each of the eight models.

2. Description of the Method

We apply the *extended P^2 algorithm* to estimate several percentiles simultaneously. In this section we describe how the variances of percentile estimates can be estimated. In addition, we give an accuracy requirement for multidimensional estimates, such as a set of percentiles. Finally we show how a mechanism to control the run length of simulation can be established using the estimated variances and a Bonferroni inequality. Before these topics we give a short description of the *extended P^2 algorithm*. The detailed algorithm is given in [17].

Estimating Percentiles with the Extended P^2 Algorithm

A straightforward method of estimating percentiles is based on sorting (at least partially) the observations. Suppose that a sequence of N observations $\{X_j\}_1^N$ is available. Let $X_{(j)}$ denote the j^{th} observation in the ordered sequence; $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(N)}$. In other words, $X_{(j)}$ is the j^{th} order statistic. Based on the order statistics the estimate of the $100p^{\text{th}}$ percentile, x_p, is $X_{([1+(N-1)p])}$, where $[\cdot]$ denotes rounding to the nearest integer.

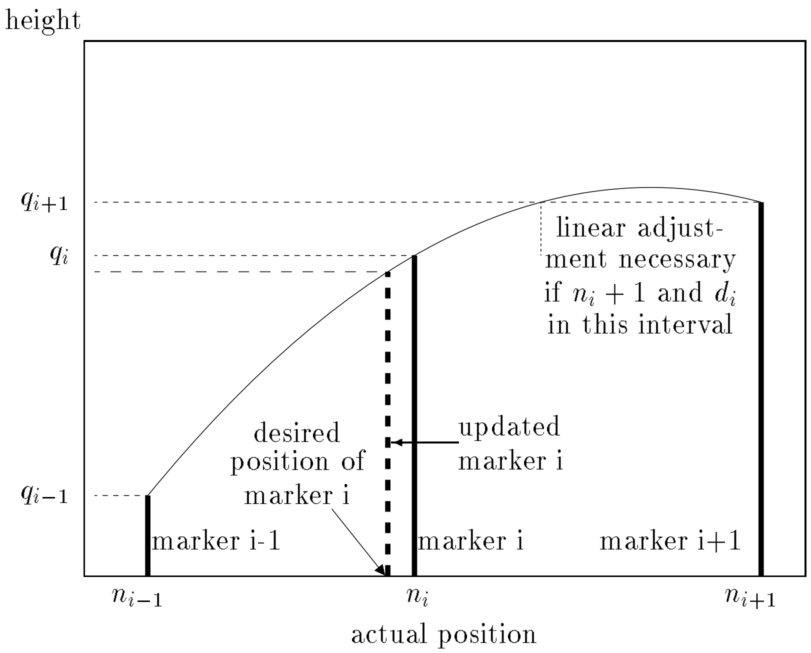


Figure 1: Example of Marker Update
Since $d_i \leq n_i - 1$, n_i is decreased by 1 and new q_i is evaluated.

Instead of storing and sorting the observations, the *extended P^2 algorithm* approximates the inverse of the empirical cumulative distribution function by a piecewise-parabola. The approximation is done by maintaining the 'marker information'. Each marker has a height, an actual position, and a desired position. Each marker is associated with the estimation of a specific percentile.

Suppose that a marker is used to estimate the $100p^{\text{th}}$ percentile of the sequence $\{X_j\}_1^N$. Now the desired position of the marker is $1 + (N - 1)p$. The algorithm tries to keep the actual position of the marker, n, close to $[1 + (N - 1)p]$. The height approximates the value of $X_{(n)}$ and is thus an estimate of x_p.

Suppose that percentiles $x_{p_1}, \ldots, x_{p_m}$ are to be estimated. The algorithm maintains m+2 principal markers and m+1 middle markers. The principal markers are used to estimate the minimum, the required percentiles, and the maximum. Each middle marker estimates the percentile at the point midway between two adjacent principal markers. The objective of the middle markers is to stabilize the updates of the marker information.

The markers are updated, if necessary, after each observation. After N observations, the actual and desired positions of the markers at the minumum and at the maximum are 1 and N, respectively. Their heights are the minimum and the maximum values in the sequence. The updating of the actual and desired positions of the other markers is simple. The actual position of each marker the height of which is greater than the observation is increased by one. If the actual position is now off to the left or to the right of its desired position by more than one position, then the height and the actual position are adjusted. The adjustment of the marker height is done using either a piecewise-parabolic or a piecewise-linear formula. The parabolic formula is preferable but sometimes the linear formula must be used in order to keep the marker heights in a nondecreasing order. Figure 1 demonstrates the updating.

The parabolic formula assumes that the curve passing through any three adjacent markers is of the form $q_i = an_i^2 + bn_i + c$, where q_i is the height and n_i is the actual position of the i^{th} marker. The movement of the actual position of a marker is always only one position. The parabolic adjustments are

$$\begin{aligned} q_i \leftarrow q_i &+ \frac{d}{n_{i+1} - n_{i-1}}\Big((n_i - n_{i-1} + d)\frac{q_{i+1} - q_i}{n_{i+1} - n_i} \\ &+(n_{i+1} - n_i - d)\frac{q_i - q_{i-1}}{n_i - n_{i-1}}\Big) , \\ n_i \leftarrow n_i &+ d , \end{aligned} \tag{1}$$

where $d = 1$, if the movement is to the right, and $d = -1$, if the movement is to the left.

The linear adjustments are

$$\begin{aligned} q_i &\leftarrow q_i + d\frac{q_{i+d} - q_i}{n_{i+d} - n_i} , \\ n_i &\leftarrow n_i + d . \end{aligned} \tag{2}$$

Although the *extended* P^2 *algorithm* proceeds only one observation at a time, we collect the observations in segments, each of length N_b. The first segment is used to initialize the algorithm. The initialization is based on the order statistics. The actual positions are initialized to $n_i = [d_i]$, where d_i is the desired position of the i^{th} marker, $i = 1, \ldots, 2m + 3$. Then it is checked that the actual positions are strictly increasing. If not, some adjustments must be made. The heights are then initialized to $q_i = x_{(n_i)}$.

Estimating the Variance of Percentile Estimates

Although our percentile estimates are not based on the order statistics, we use the asymptotic properties of the order statistics in estimating the variances of the percentile estimates. The empirical results given in [17] and this paper indicate that the behaviour of the percentile estimates based on the *extended* P^2 *algorithm* and the order statistics is usually very similar.

When the output sequence $\{X_j\}_1^N$ is stationary and satisfies the ϕ-mixing condition, the percentile estimate $\hat{x}_p$, based on the order statistic, has a normal limiting distribution (see Appendix I). Hence, as N becomes large, the variance of $\hat{x}_p$ can be approximated by

$$\mathrm{Var}(\hat{x}_p) \approx \hat{h}_p(0) \Big/ \left(N \hat{f}(\hat{x}_p)^2\right), \tag{3}$$

where $\hat{f}(\hat{x}_p)$ is the estimated (marginal) density of the X_j's at x_p and $\hat{h}_p(\omega)$ is the estimated spectral density of the binary process $\{I_j(x_p)\}$ at frequency ω:

$$I_j(x_p) = \begin{cases} 1 & \text{, if } X_j \leq x_p \text{ ,} \\ 0 & \text{, if } X_j > x_p \text{ .} \end{cases} \tag{4}$$

The *extended P^2 algorithm* approximates the inverse of the empirical cumulative distribution function by a piecewise-parabola, $\hat{F}^{-1}(y) = ay^2 + by + c$, in the neighbourhood of p. Hence the density function can be approximated by $\hat{f}(x) = (b + 2a\hat{F}(x))^{-1}$ in the neighbourhood of x_p.

The spectral density of the binary process $\{I_j(x_p)\}$ has a central role, when the variance of $\hat{x}_p$ is estimated. The generation of this process requires that the whole sequence $\{X_j\}$ is available. The problem of storing the whole sequence is avoided, when $\{I_j(x_p)\}$ is approximated by the process $\{\hat{I}_{j,k}(\hat{x}_p)\}$:

$$\hat{I}_{j,k}(\hat{x}_p) = \begin{cases} 1 & \text{, if } X_{j,k} \leq \hat{x}_{p,k} \\ 0 & \text{, otherwise} \end{cases} \quad , \; j = 1, \ldots, N_b \,, \; k = 1, \ldots, K \,, \tag{5}$$

where $X_{j,k}$ is the j^{th} item in the k^{th} segment (the $j + (k-1)N_b{}^{\text{th}}$ observation), and $\hat{x}_{p,k}$ is the estimate of x_p based on k segments (kN_b observations).

In each segment we evaluate the spectral density estimates, $\hat{h}_k(\omega_j)$, for $j = 1, \ldots, n < N_b/(4M + 2)$ by averaging $2M + 1$ adjacent periodogram values of the sequence $\{\hat{I}_{j,k}(\hat{x}_p)\}_{j=1}^{N_b}$. The overall spectral density estimates are taken as

$$\hat{\hat{h}}_p(\omega_j) = \frac{1}{K} \sum_{k=1}^{K} \hat{h}_k(\omega_j) \,, \tag{6}$$

where K is the number of segments. For details, see Appendix II.

A convenient method to estimate the spectral density at frequency 0 has been developed in [8]. A low order polynomial is fitted to the logarithms of the smoothed periodogram values. We apply a similar regression approach to construct an approximately unbiased estimate of $h_p(0)$ (see Appendix III.

Accuracy Requirement for Multidimensional Index

For a one-dimensional index θ, such as the mean or a single percentile point, the accuracy requirement is usually given as the upper bound of the relative half-width ε for the confidence interval of level $1-\alpha$. Using this requirement and asymptotic normality of the estimate $\hat{\theta}$ the simulation is continued until

$$\hat{s}/\hat{\theta} \le \varepsilon / t_d(1-\alpha/2)\,, \tag{7}$$

where $\hat{s}^2$ is the estimated variance of $\hat{\theta}$ with d degrees of freedom, and $t_d(1-\alpha/2)$ is the $100(1-\alpha/2)^{\text{th}}$ percentile point of Student's t-distribution with d degrees of freedom.

This requirement can be interpreted as $\Pr\{|\theta-\hat{\theta}| \le \varepsilon|\hat{\theta}|\} \ge 1-\alpha$. In other words, we have a confidence of at least $1-\alpha$ to trust that the true value θ is in the interval $[(1-\varepsilon)\hat{\theta},(1+\varepsilon)\hat{\theta}]$. A natural extension to multidimensional situations, where the index of interest is $\boldsymbol{\theta} = (\theta_1,\ldots,\theta_m)'$, is to require that we have a confidence of at least $1-\alpha$ to trust that each of the components is in the interval $[(1-\varepsilon_j)\hat{\theta}_j,(1+\varepsilon_j)\hat{\theta}_j]$, i.e.

$$\Pr\{|\theta_j-\hat{\theta}_j| \le \varepsilon_j|\hat{\theta}_j|\ ;\ j=1,\ldots,m\} \ge 1-\alpha\,. \tag{8}$$

Using a Bonferroni inequality (see e.g. [11, p. 41]) it can be concluded that the requirement (8) is satisfied, if

$$\Pr\{|\theta_j-\hat{\theta}_j| \le \varepsilon_j|\hat{\theta}_j|\} \ge 1-\alpha_j \quad \text{and} \quad \sum_{j=1}^{m}\alpha_j \le \alpha\,. \tag{9}$$

Suppose that the estimates are $\hat{\theta}_j$ and their estimated variances $\hat{s}_j^2$, each with d_j degrees of freedom. Using the confidence intervals based on normal approximations, $\Pr\{|\theta_j-\hat{\theta}_j| \le \hat{s}_j\, t_{d_j}(1-\alpha_j/2)\} = 1-\alpha_j$, the probabilities that $|\theta_j-\hat{\theta}_j| \le \varepsilon_j|\hat{\theta}_j|$ can be solved:

$$\alpha_j = 2F_{d_j}(-\varepsilon_j|\hat{\theta}_j|/\hat{s}_j)\,, \tag{10}$$

where $F_d(t)$ is the cumulative distribution function of the Student t-distribution with d degrees of freedom. Hence the criterion to terminate the simulation is $\sum\alpha_j \le \alpha$. If the criterion is not satisfied, additional segments of observations must be generated in order to obtain estimates of acceptable accuracy.

It is not necessary to evaluate the termination criterion after each segment. Since the $\hat{s}_j$'s are proportional to $1/\sqrt{KN_b}$, we can obtain a rough prediction for the number of segments, K', which is required to satisfy the termination criterion:

$$K' = K\sqrt{\sum\alpha_j/\alpha}\,, \tag{11}$$

where K is the current number of segments, and the α_j's are based on current estimates, $\hat{\theta}_j, \hat{s}_j^2, d_j$. Since K' is only a rough prediction, we actually perform the next termination test after K'' segments, where

$$K'' = \min\{2K, \max\{K+2, K'\}\}\,. \tag{12}$$

Implementation Details

The method described above has several internal and external parameters. The external parameters, the ε_j's and α are used to control the accuracy of the estimates. The problem of choosing suitable values is application dependent. However, our experience is that values greater than 0.15 will not provide stable estimates, since the variance estimates are based on asymptotic properties.

The internal parameters are N_b, M, and n, i.e. the length of the segments, the degree of 'local' averaging, and the number of frequencies in the regression. The first two of them affect the variance and the bias of the spectral estimates. Our experience is that $N_b = 512$ or 1024 is a reasonabe choice. The degree of 'local' averaging should be low, say $M = 0$ or $1/2$. Higher values of M will produce seriously biased estimates, if N_b is not increased. With $M = 1/2$ we have found that n should be 21–35. If $M = 0$, n should be somewhat greater, about 35–50. In the experiments, we used the following values: $N_b = 512$, $M = 1/2$, and $n = 31$.

3. Experimental Results

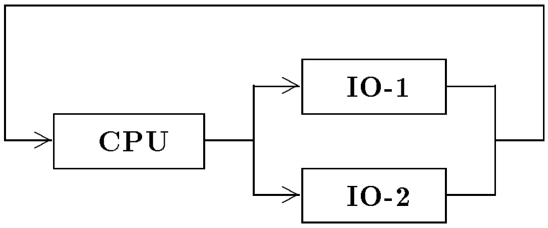
In the simulation experiments we analyzed three different questions. Our primary interest is the sequential estimation of several percentiles simultaneously. However, before we studied the run length control mechanism, we wanted to be sure that the *extended P^2 algorithm* and the variance estimation method do not introduce serious errors. Hence, we first analyzed the behaviour of the *extended P^2 algorithm*. The objective was to find out, whether the estimates based on the proposed algorithm have similar first- and second-order characteristics as the order statistics. In the second phase we studied the properties of the variance estimation method. Finally we analyzed the mechanism to control the length of simulation runs.

Analyzed Models

Since our primary interest is in modelling computer systems, we selected models, which represent them. In all the experiments we analyzed percentiles of the response times from eight different models. The models are used in many simulation studies, for example in [12, 13, 19, 8]. Figure 2 contains a summary of the models.

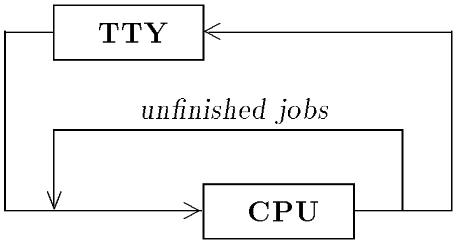
Models 1–4 are the so called Buzen models [5]. They represent a multiprogrammed computer system, where the number of active tasks is constant (N). All the service times are exponentially

Models 1–4



Model	1	2	3	4
N	4	8	8	8
S_{CPU}	1	1	1	1
S_{IO-1}	2	2	2.22	0.56
S_{IO-2}	2	2	20	5
p_{IO-1}	0.5	0.5	0.9	0.9
p_{IO-2}	0.5	0.5	0.1	0.1

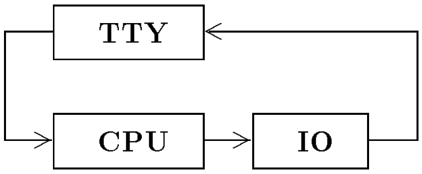
Model 5: Population 35



Node	Type	Serv. Rate
TTY	IS	0.04
CPU	RR*	1.25

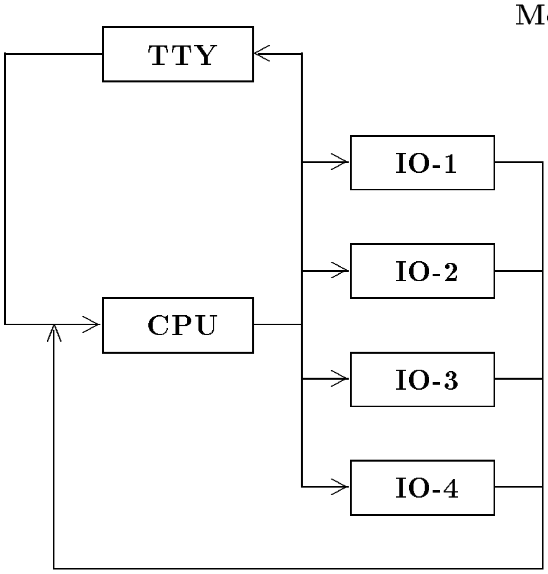
* quantum .1; overhead .015

Model 6: Population 40



Node	Type	Serv. Distr.
TTY	IS	Exp(1/30)
CPU	PS	$H_2(0.3, 0.8, 12.5)$
IO	IS	Exp(2)

Models 7 and 8



Model	7	8
Population	25	25
S_{TTY}	100	100
S_{CPU}	1	1
$S_{IO-1} = S_{IO-2}$	1.39	5.56
$S_{IO-3} = S_{IO-4}$	12.5	25
p_{TTY}	0.2	0.2
$p_{IO-1} = p_{IO-2}$	0.36	0.36
$p_{IO-3} = p_{IO-4}$	0.04	0.04

Figure 2: Analyzed Models

distributed with means S_{CPU}, S_{IO-1}, and S_{IO-2}. After completion of service at the CPU–node the task enters either of the IO–nodes. The IO–node is selected according to probabilities p_{IO-1} and p_{IO-2}. When a task leaves an IO–node, a new task simultaneously enters the CPU–node.

Models 5–8 represent time-sharing computer systems. The collection of user terminals is modelled as an infinite server (TTY–node), where the waiting time for the service is always zero. Its service time represents the 'thinking-time' between a task completion and the activation of the next one.

In Model 5 all the service times are exponentially distributed. The scheduling policy in the CPU–node is round robin: Each job receives service in quantums of fixed size. If the service is not completed in a quantum, the job is removed into the end of the waiting queue. This model is treated analytically in [1]. A detailed simulation study is given in [19].

Model 6 is a variation of the 'standard' queueing network, where service times have high variability, see e.g. [25]. In the CPU–node service times are hyperexponentially distributed. A hyperexponential distribution, $\mathrm{H}_2(\pi, \mu_1, \mu_2)$, is a special case of the general Coxian distribution. The service time is a mixture of two exponential distributions: $\mathrm{Exp}(\mu_1)$ with probability π, and $\mathrm{Exp}(\mu_2)$ with probability $1 - \pi$.

Models 7 and 8 are so called central server models analyzed in [8]. All the service times are exponentially distributed with means S_{CPU}, S_{TTY}, S_{IO-1}, S_{IO-2}, S_{IO-3}, and S_{IO-4}, respectively. After the completion of service at the CPU–node the task either enters one of the IO–nodes or leaves the system (enters the TTY–node). The node is chosen according to probabilities p_{IO-1}, p_{IO-2}, p_{IO-3}, p_{IO-4}, and p_{TTY}, respectively.

Properties of the Extended P^2 Algorithm

In the first phase of the experiments, we compared the first- and second-order properties of the *extended P^2 algorithm* to those of the order statistics. We generated 101 independent sequences of response times using each of the eight models. The sequence lengths were 65 536. In each sequence the 5$^{\text{th}}$, 10$^{\text{th}}$, 25$^{\text{th}}$, 50$^{\text{th}}$, 75$^{\text{th}}$, 90$^{\text{th}}$, and 95$^{\text{th}}$ percentile were estimated with the *extended P^2 algorithm* and the order statistics using the first 1 024, 2 048, 4 096, 8 192, 16 384, and 32 768 observations in the sequence and then using the whole sequence. The initial state was constructed using the steady-state load distribution. In additon, a warm-up period of random length (uniform in $[1\,025, 2\,048]$) was used to reduce the initialization bias.

The comparison of the first-order properties is based on paired differences. We constructed the 90% conficence intervals for the differences. These intervals indicates that usually the estimates obtained by the *extended P^2 algorithm* are close to those given by the order statistics. Only in ten cases (out of $8 \times 7 \times 7 = 392$) zero does not belong to the interval. These cases are the 90$^{\text{th}}$

percentile in Models 1, 2, 4, 5, 7, and 8; and the 95^{th} percentile in Models 1,5, 7, and 8. The run length was 65 536 in each of these cases. The differences are not of practical importance since the mid points of the confidence interval are less than 0.5% of the expected value of the corresponding percentile.

In addition to the confidence intervals, we considered the mean relative differences,

$$\sum_{i=1}^{n} |y_i - x_i| / \sum_{i=1}^{n} x_i \,, \tag{13}$$

where the percentile estimates based on P^2 *algorithm* are denoted by the y_i's and on the order statistics by the x_i.

Figure 3 visualizes the mean relative differences. The figure indicates that in Models 3 and 8 the relative difference in a single percentile estimate may be quite large, when the sequence is short. But the overall conclusion is that usually there are no difference of practical significance between the estimates based on the P^2 *algorithm* and the order statistics.

In Models 3 and 8, where the maximum relative distances are considerably greater than in the other models, the IO–nodes are the 'bottlenecks' of the system. This means that the response times are dominated by the time spent at the IO–nodes. Both the models have two types of IO–nodes: faster and slower ones. The ratios of the service rates are 1 : 9 in Model 3 and 1 : 4.5 in Model 8. Hence the response time distribution in Model 3 is essentially a mixture of two quite different distributions. In Model 8 the situation is more complicated, since a task can visit the IO–nodes several times. The response time distribution in this case is a mixture of several different distributions, some of which are quite different from each other.

Since the estimated percentiles were based on the same sequences, the standard test of equal covariance matrices can not be used. Instead, we tested the hypothesis that the covariance matrix of the P^2 estimates is equal to a given matrix that is the sample covariance matrix of the estimates based on the order statistics. The test is described in [2, p. 264–267].

In four cases (out of $8 \times 7 = 56$) the hypothesis of equal covariance matrices was rejected at the significance level 0.10. These cases were Model 3 run lengths being 4 096 and 16 384, and Model 4 run lengths being 1 024 and 2 048. These cases were examined in details. It turned out that in Model 4 the rejections were due to one or two outliers. In Model 3 the situation turned out to be quite complicated. We analyzed the variances and coefficients of correlations separately. We observed that the 90^{th} percentile caused the rejections. When the hypothesis of the covariance matrices being equal was tested without the 90^{th} percentile, the hypothesis was not rejected.

The 90^{th} percentile is problematic in Model 3. Since both the IO–nodes are the 'bottlenecks' of the system, the response time distribution is essentially a mixiture of two quite different distributions, and the 'mixing probability' happens to be 0.9. In six cases the *extended* P^2 *algorithm*

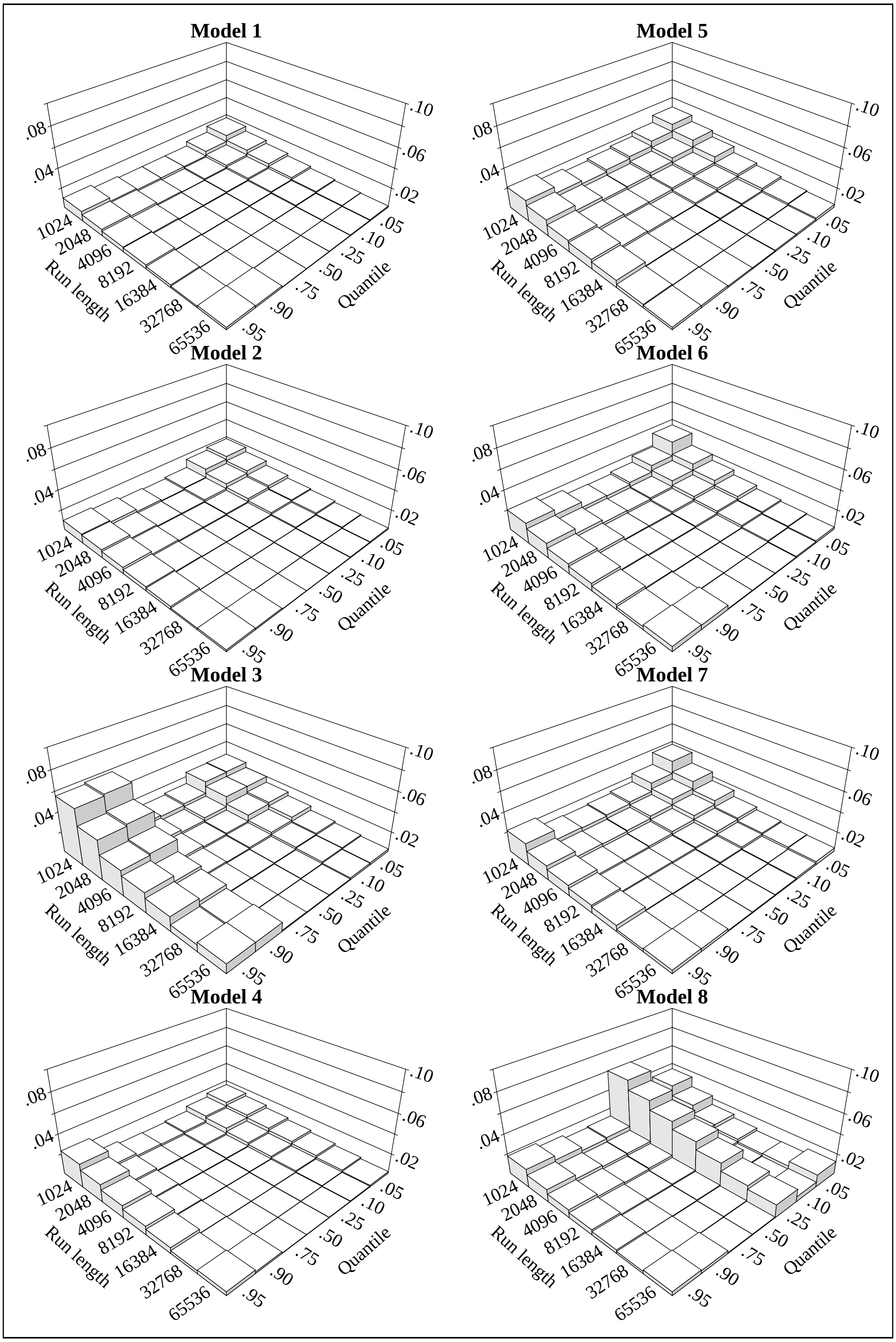


Figure 3: Mean Relative Differences Between Estimated Quantiles

produced essentially greater estimates of the 90$^{\text{th}}$ percentile than the order statistics. When these sequences were examined, we observed that the initialization of the *extended* P^2 *algorithm* was the key problem: The order statistic estimates of high percentiles were far away from their expectations. Since the updates of the P^2 estimates are stabilized by the 'middle-markers', the *algorithm* requires a long sequence to recover from the initial error due to order statistic estimates.

Despite the significant differences in quite a few sequences, our overall conclusion is that the second-order properties of the P^2 estimates are usually very close to those of the estimates based on order statistics. Hence, the variance estimation method can be based on the asymptotic variances of sample percentiles.

Properties of Variance Estimates

In the second phase of the experiments we analyzed the properties of the variance estimates. We were interested in the accuracy and stability of the estimated variances. We generated 101 new independent sequences of response times from the eight models. Using sequence lengths of 1 024, 2 048, 4 096, 8 192, 16 384, 32 768, and 65 536 we estimated the 5$^{\text{th}}$, 10$^{\text{th}}$, 25$^{\text{th}}$, 50$^{\text{th}}$, 75$^{\text{th}}$, 90$^{\text{th}}$, and 95$^{\text{th}}$ percentile together with their variances.

Since the true variances of the estimates are unknown, we compare the variance estimates based on the proposed method to the sample variances of the independent replications. The ratios of the means of estimated variances to the replication variances are given in Table 1. With the exception of Models 3 and 6 the proposed variance estimation method seems to produce reasonable estimates of variances. In Model 3 the variances of all the percentiles, but the 90$^{\text{th}}$ one, are seriously underestimated.

In Model 6 the variance estimates based on our method are considerably smaller than the replication variances, especially when the sequence becomes longer. We found that the replication variances, with the exception of the 95$^{\text{th}}$ percentile, are almost equal for all the sequence lengths. We also found that the parametric and nonparametric confidence intervals of 90% for the estimated percentiles overlap. The parametric confidence interval is $\bar{x} \pm 1.66s$, where $\bar{x}$ is the replication mean and s^2 is the replication variance. The nonparametric confidence interval is $(x_{(6)}, x_{(96)})$. This may indicate that Model 6 does not satisfy the ϕ-mixing condition.

The stability of the variance estimates is measured using their coefficient of variation (standard deviation divided by mean), which are given in Table 2. With the expection of the 90$^{\text{th}}$ percentile in Model 3 the coefficients of variations are not (statistically) significantly greater than those of the corresponding χ^2-distributed random variables.

Model	Percentile	Run Length 1024	2048	4096	8192	16384	32768	65536
1	5	1.1234	1.1113	1.1893	1.0403	1.2332	1.1947	1.0091
	10	1.1708	1.0214	.8883	.9582	1.0260	.8393	.9127
	25	1.2134	1.0465	.9914	.9392	.9744	.7571	.9611
	50	.9891	.9727	.8607	.8696	.9469	.8461	.9553
	75	.9755	1.2943	1.0295	1.1350	1.0821	.9653	.9611
	90	.9046	1.0024	1.0927	1.0210	1.0853	1.1897	1.0947
	95	1.1165	1.2368	1.2980	1.2746	1.2865	1.2643	1.2259
2	5	1.4378	1.0692	1.3508	1.3449	.9618	1.0999	1.2093
	10	1.2037	1.0283	1.3758	1.3449	1.1429	1.2585	1.1448
	25	1.1944	1.0268	1.3167	1.3388	1.3891	1.8769	1.6949
	50	1.2926	1.1123	1.2596	1.2776	1.5268	1.5655	1.5596
	75	1.3074	1.5124	1.3496	1.3989	1.7087	1.5463	1.9879
	90	1.1546	1.3790	1.5358	1.4232	1.4071	1.4268	1.5099
	95	1.4390	1.3772	1.4038	1.5010	1.4845	1.4621	1.3429
3	5	.2440	.3409	.3710	.3479	.4433	.3307	.3321
	10	.2128	.3186	.3134	.2673	.3133	.2600	.2592
	25	.2206	.2779	.2500	.1911	.1985	.1817	.1857
	50	.2642	.3124	.2796	.2098	.2003	.1901	.1962
	75	.4830	.5712	.5082	.4196	.4141	.4246	.4288
	90	.5675	.6741	1.4424	2.1979	2.9101	3.0432	2.6893
	95	.3706	.4086	.3362	.2434	.2873	.2902	.2425
4	5	.5897	.6181	.7654	.6848	.6534	.5581	.5984
	10	.6683	.7280	.6963	.7151	.7131	.5487	.6149
	25	.7868	.9021	.8628	.9260	.8522	.6280	.7718
	50	.8863	.9028	1.0367	.9756	.8857	.7060	.9294
	75	1.1469	1.0062	1.3407	1.2437	1.1552	1.0030	1.2629
	90	1.0186	.7683	.9350	1.0074	1.2206	1.2253	1.2654
	95	.8232	1.1676	1.2418	1.1321	1.3094	1.2370	1.3613
5	5	1.0005	.9288	.8271	.8210	.8593	.8182	1.0455
	10	.9803	.8067	.8110	.7087	.6996	.7198	.7839
	25	.8882	.8185	.7302	.6896	.7526	.7068	.6937
	50	.9355	.8865	.7418	.7195	.7865	.7126	.7484
	75	.7912	.7641	.6951	.6913	.7780	.6891	.6143
	90	.6652	.6134	.6638	.6575	.6251	.5467	.5453
	95	.7441	.6392	.7238	.7509	.4973	.5558	.5381
6	5	.7130	.6700	.3863	.2467	.1504	.0837	.0458
	10	.7108	.5926	.3783	.2615	.1421	.0777	.0430
	25	.7238	.4533	.3469	.2345	.1200	.0653	.0336
	50	.8988	.6632	.5003	.2934	.1519	.0895	.0474
	75	1.0454	1.1292	1.1776	.8995	.6683	.5314	.4269
	90	1.1628	1.0101	1.0870	.9141	.7288	.6182	.3990
	95	1.2778	1.2221	1.0849	1.0468	1.0360	1.0858	.9068
7	5	.9757	.9863	.8467	.8761	.9827	1.1406	.9229
	10	.8473	.9936	.8171	.8781	.8307	.9410	.8923
	25	.7899	.9853	1.0138	1.0858	.9997	.9718	.8641
	50	.8751	1.0812	.9577	.9023	.9340	.8641	.9682
	75	.7495	.9854	.9255	.8956	.9263	.9314	.9859
	90	.7731	.9627	.7527	.8954	.8356	.9303	.8547
	95	1.0030	1.1180	.7957	.9192	.9060	.8608	.8547
8	5	.9588	1.1183	.9903	1.3472	1.2592	1.1689	1.1777
	10	1.6802	1.5881	1.3739	1.3823	1.3327	1.7713	1.5069
	25	.5884	.7594	.6898	.5860	.5872	.5671	.7460
	50	.7576	.8052	1.0862	1.1494	1.1219	.9505	1.1212
	75	.8491	.6907	.9374	1.0406	.8658	.7835	.9142
	90	1.0050	.8195	.9651	1.0047	.8956	.8129	.9379
	95	1.0253	.7329	1.0480	.9634	.8771	.7963	1.0006

Table 1: Ratios of the Means of Estimated Variances to Replication Variances

Model	Percentile	Run Length 1024	2048	4096	8192	16384	32768	65536
1	5	.22523	.14964	.11150	.07593	.05746	.04037	.02805
	10	.20155	.15219	.11236	.07904	.05620	.04108	.02514
	25	.18503	.12985	.09535	.07134	.04928	.03055	.02054
	50	.17892	.11272	.07697	.05192	.03744	.02659	.01940
	75	.18579	.11732	.07450	.06534	.04426	.03031	.02200
	90	.24394	.17601	.12537	.08663	.06039	.04145	.02710
	95	.30470	.24716	.16509	.14991	.09970	.06971	.05239
2	5	.23291	.15157	.11079	.07971	.05883	.04453	.03009
	10	.25086	.17352	.11039	.07919	.05794	.04082	.02613
	25	.19556	.12625	.08355	.05988	.04554	.03334	.02290
	50	.16588	.12012	.07903	.05477	.03868	.03062	.02256
	75	.19347	.14437	.09571	.07163	.04815	.03660	.02545
	90	.30127	.22011	.16047	.11037	.07891	.05998	.04132
	95	.42794	.36744	.28097	.17938	.11130	.09884	.05811
3	5	.44590	.32261	.25081	.18580	.13988	.10110	.07663
	10	.36939	.26512	.19567	.15243	.11080	.08143	.05503
	25	.23388	.17549	.12621	.10184	.07090	.05032	.03618
	50	.20170	.13779	.09544	.06449	.04466	.03493	.02444
	75	.31958	.21014	.13574	.10302	.07200	.05150	.03843
	90	.62161	.53909	.47661	.37199	.26679	.19371	.15663
	95	.37988	.27400	.23166	.17276	.10661	.08123	.06754
4	5	.41448	.31398	.25871	.19865	.15665	.09406	.06662
	10	.30186	.23184	.17198	.12625	.09370	.06108	.04371
	25	.19473	.14104	.09873	.07777	.05168	.03672	.02839
	50	.17301	.12328	.09298	.06180	.04285	.03012	.02313
	75	.22034	.15742	.11461	.07799	.05732	.03757	.02422
	90	.26959	.20579	.17424	.11019	.09009	.06225	.04455
	95	.52749	.35657	.27560	.21668	.18910	.12710	.08129
5	5	.26735	.22292	.16103	.12138	.08642	.06356	.04773
	10	.25411	.17916	.11898	.09455	.06234	.04380	.03194
	25	.19746	.14713	.10065	.07035	.04730	.03571	.02840
	50	.16458	.11047	.07377	.05436	.03559	.02919	.02168
	75	.18256	.12447	.08646	.06339	.04727	.03745	.02304
	90	.25608	.20776	.14428	.09368	.08009	.05350	.03798
	95	.36925	.30281	.25616	.16728	.14894	.11710	.08681
6	5	.21136	.14127	.11979	.08974	.07958	.07030	.06519
	10	.22779	.16963	.11267	.08050	.07132	.06115	.05049
	25	.18506	.13648	.08721	.07126	.05465	.04839	.04179
	50	.16574	.11795	.07703	.05938	.04739	.03825	.03141
	75	.18481	.14118	.10185	.06986	.04858	.03878	.03170
	90	.24199	.18308	.13854	.10946	.09680	.08328	.08402
	95	.33158	.26601	.21398	.15079	.10826	.07613	.06431
7	5	.31203	.20342	.14370	.10219	.08192	.05570	.04030
	10	.22352	.18284	.14682	.09439	.06752	.04579	.03294
	25	.20505	.13197	.09261	.05975	.04651	.03372	.02451
	50	.19644	.12647	.08641	.05757	.04306	.02870	.02206
	75	.18720	.13168	.10286	.06967	.05108	.03171	.02594
	90	.28771	.20289	.14638	.09748	.06127	.04623	.03243
	95	.30203	.28387	.22734	.14264	.11089	.08100	.05883
8	5	.21411	.14780	.10582	.07967	.05488	.03955	.02826
	10	.23757	.16868	.12014	.09932	.07155	.04410	.02781
	25	.21864	.13847	.09357	.06981	.05253	.03763	.02440
	50	.18321	.11947	.07574	.05530	.03994	.02790	.01834
	75	.21187	.13526	.09082	.06607	.05364	.03787	.02251
	90	.23899	.17541	.11866	.08097	.06280	.04345	.03063
	95	.27802	.21413	.16290	.13363	.10245	.06382	.04723

Table 2: Coefficients of Variation to Estimated Variances

Model	1	2	3	4	5	6	7	8
Run Length								
Mean	1125	1642	13079	1744	8070	2879	6367	3447
St. deviat.	306	560	6513	586	695	1005	2035	952
Coverage	.94	.95	.87	.99	.90	.91	.96	.94
Fraction of outliers								
in all dimensions	0	0	0	0	.02	0	0	0
in two dimensions	.01	0	.02	0	.04	.01	0	0
in one dimension	.05	.05	.11	.01	.04	.08	.04	.06

Table 3: Summary of Sequential Estimation Runs

Sequential Estimation of Several Percentiles

In the last phase of the experiments we estimated the 50^{th}, 75^{th}, and 90^{th} percentile of the response time from the eight models using the proposed sequential estimation procedure. The accuracy requirement was $\varepsilon = 0.1$ for each percentile and $\alpha = 0.1$. A heuristic interpretation of the used requirement is that the relative errors of all the estimates are less than 10 % with a probability not less than 0.90. The minimum run length was 1024.

With each of the eight models we carried out 101 independent replications. Since the true values of percentiles are unknown, we used the means of the order statistic estimates (phase one of the experiments) as the 'exact' values.

A summary of these sequential simulation runs is given in Table 3. The table contains the means and standard deviations of the run lengths, coverages, and fractions of outliers. With the exception of Model 3, the run lengths are quite short and stable.

The coverage is the fraction of runs, where all the estimates are accurate enough, i.e. $|\theta_j - \hat{\theta}_j| \leq \varepsilon|\hat{\theta}_j|$ for all j, where the $\hat{\theta}_j$'s are the estimated percentiles and the θ_j's are their 'exact' values. Only in Model 3 the observed coverage is less than 0.9, i.e. $1-\alpha$. However, if we repeat a Bernoulli trial, where the probability of success is 0.9, 101 times, 88 is not a statistically exceptional result. In fact, the 95% confidence interval for the number of successes is $[85, 97]$.

The fractions of outliers are classified into three cathegories. They correspond to the number of dimensions, in which the estimates do not satisfy the given accuracy requirement. Only in Model 5 there are estimates, which are outliers in all of the three dimensions.

Figures 4a and 4b give a visualization of the obtained percentile estimates. The three-dimensional estimates are projected on the faces of a cube. On each of the three visible faces a two-dimensional (marginal) scatter-plot is pasted. (For details, see [23].) The plotareas are

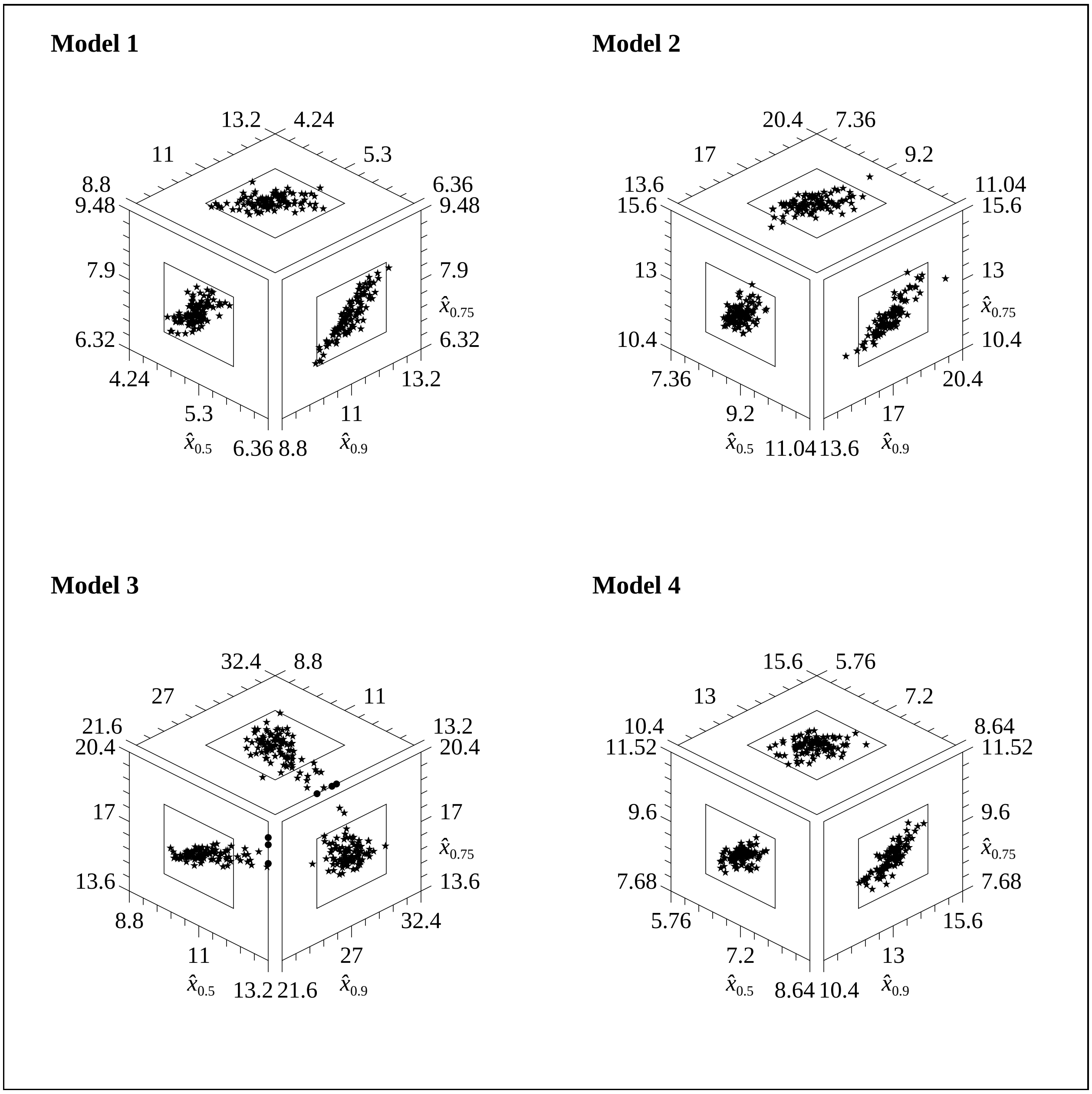


Figure 4a: Estimated Prercentiles

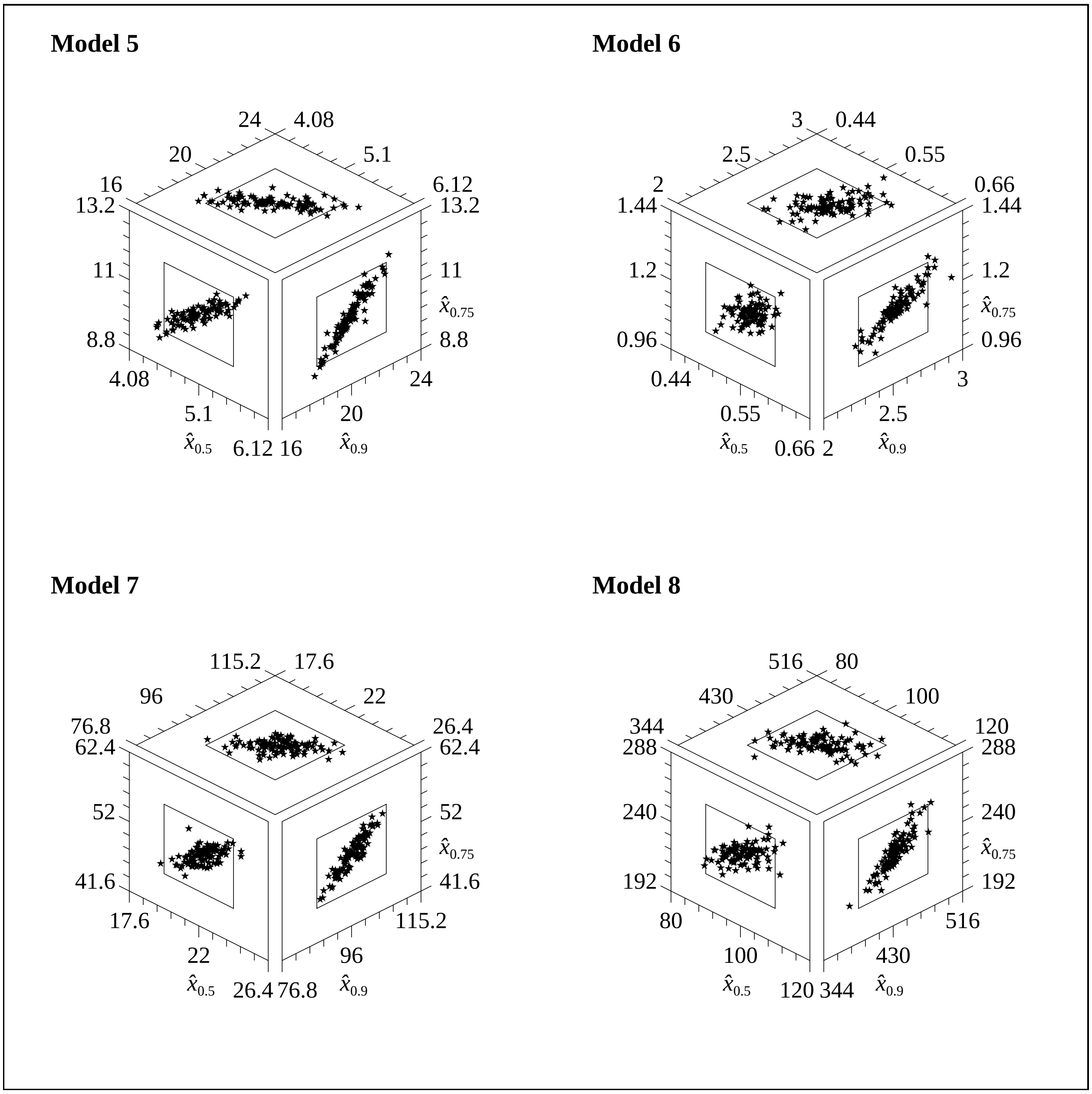


Figure 4b: Estimated Prercentiles

$\theta_j \pm 20\%$. Estimates, which are out of the plotarea, are shown as bullets (•) on the frames. The parallelograms on the faces correspond to $\theta_j \pm 10\%$. With the exception of Model 3, there are no serious departures from the 'exact' values of percentiles. In Model 3, the estimated medians exceed their expectation by more than 20% in three cases.

4. Conclusions

The overall conslusion from these experiments is that the proposed sequential estimation procedure provides the percentile estimates within the required accuracy. We have reported various difficulties in one of the eight models. We have not tried to use any other method to analyze that system. However, our intuitive belief is that the model would be difficult also to other methods, since its response time distribution is essentially a mixture of two quite different distributions.

References

[1] Adiri, I. and B. Avi-Itzhak. 1969. "A Time-Sharing Queue with a Finite Number of Customers". *Journal of the ACM* **16**, *no. 2* (Apr.): 315–323.

[2] Anderson, T. W. 1958. *Introduction to Multivariate Statistical Analysis*. Wiley, New York, N.Y.

[3] Bartlett, M. S. and D. G. Kendal. 1946. "The Statistical Analysis of Variance Heterogeneity and the Logarithmic Transformation". *Supplement to the Journal of the Royal Statistical Society* **8**, *no. 1*: 128–138.

[4] Blackman, R. B. and J. W. Tukey. 1959. *The Measurement of Power Spectrum from the Point of View of Communications Engineering*. Dover, New York, N.Y.

[5] Buzen, J. P. 1971. *Queueing Network Models of Multiprogramming*. Ph.D. Dissertation, Division of Engineering and Applied Physics, Harvard University, Cambridge, Mass.

[6] Cox, D. R. and D. V. Hinkley. 1974. *Theoretical Statistics*. Chapman and Hall, London, U.K.

[7] Heidelberger, P. and Lewis, P. A. W. 1984. "Quantile Estimation in Dependent Sequences". *Operations Research* **31**, *no. 1* (Jan.-Feb.): 185–209.

[8] Heidelberger, P. and P. D. Welch. 1981. "A Spectral Method for Confidence Interval Generation and Run Length Control in Simulations". *Communications of the ACM* **24**, *no. 4* (Apr.): 233–245.

[9] Iglehart, D. L. 1976. "Simulating Stable Stochastic Systems, VI: Quantile Estimation". *Journal of the ACM* **23**, *no. 2* (Apr.): 347–360.

[10] Jain, R. and I. Chlamtac. 1985. "The P^2 Algorithm for Dynamic Calculation of Quantiles and Histograms without Storing Observations". *Communications of the ACM* **28**, *no. 10* (Oct.): 1076–1085.

[11] Kleijnen, J. P. C. 1987. *Statistical Tools for Simulation Practitioners*. Marcel Dekker, New York, N.Y.

[12] Law, A. M. and J. S. Carson. 1979. "A Sequential Procedure for Determining the Length of a Steady-State Simulation". *Operations Research* **27**, *no. 5* (Sep.-Oct.): 1011–1025.

[13] Law, A. M. and W. D. Kelton. 1982. "Confidence Intervals for Steady-State Simulations, II: A Survey of Sequential Procedures". *Management Science* **28**, *no. 5* (May): 550–562.

[14] Law, A. M. and W. D. Kelton. 1984. "Confidence Intervals for Steady-State Simulations: I. A Survey of Fixed Sample Size Procedures". *Operations Research Science* **32**, *no. 6* (Nov.-Dec.): 1221–1239.

[15] Moore, L.W. 1980. *Quantile Estimation Methods in Regenerative Processes*. Ph.D. Dissertation, Dept. of Statistics, Univ. of North Carolina, Chapel Hill, N.C.

[16] Olshen, R. A. 1967. "Asymptotic Properties of the Periodogram of a Discrete Stationary Process". *Journal of Applied Probability* **4**, *no. 4* (Dec.): 508–528.

[17] Raatikainen, K. E. E. 1987a. "Simultaneous Estimation of Several Percentiles". *Simulation* **49**, *no. 4* (Oct.): 159–164.

[18] Raatikainen, K. E. E. 1987b. "Run Length Control for Simultaneous Estimation of Several Percentiles in Dependent Sequences". In *Proceedings of the Conference on Methodology and Validation, 1987* (Simulation Series, Vol 19, no. 1). Society for Computer Simulation, San Diego, Calif.: 54–59.

[19] Sargent, R. G. 1976. "Statistical Analysis of Simulation Output Data". In *Proceedings of the Symposium on the Simulation of Computer Systems* (Boulder, Colorado, Aug 10–12, 1976). Association for the Computing Machinery, New York, N.Y.: 39–50.

[20] Seila, A. F. 1982a. "A Batching Approach to Quantile Estimation in Regenerative Simulations". *Management Science* **28**, *no. 5* (May): 573–581.

[21] Seila, A. F. 1982b. "Estimation of Percentiles in Discrete Event Simulation". *Simulation* **39**, *no. 6* (Dec.): 193–200.

[22] Sen, P. K. 1972. On the Bahadur Representation of Sample Quantiles for Sequences of ϕ-mixing Random Variables". *Journal of Multivariate Analysis* **2**, *no. 1* (Mar.): 77–95.

[23] Tukey, P. A. and J. W. Tukey. 1981. "Preparation; Prechosen Sequences of Views". In *Interpreting Multivariate Data*. Ed. V. Barnett. Wiley, Chichester, U.K.: 189–213.

[24] Welch, P. D. 1961. "A Direct Digital Method of Power Spectrum Estimation". *IBM Journal of Research and Development* **5**, *no. 2* (Apr.): 141–156.

[25] Zahorjan, J., E. D. Lazowska, and R. L. Garner. 1983. "A Decomposition Approach to Modelling High Service Time Variability." *Performance Evaluation* **3**, *no. 1* (Feb.): 35–54.

I Asymptotic Properties of Order Statistic

The following result is given in [22].

Let $\{X_i\}_{i=-\infty}^{\infty}$ be a stationary sequence of random variables defined on a probability space $(\Omega, \mathcal{A}, P)$. Let $\mathcal{M}_{-\infty}^{j}$ and $\mathcal{M}_{j+n}^{\infty}$ be the σ-fields generated by $\{X_i\}_{i=-\infty}^{j}$ and $\{X_i\}_{i=j+n}^{\infty}$, respectively. Suppose that $E_1 \in \mathcal{M}_{-\infty}^{j}$, $E_2 \in \mathcal{M}_{j+n}^{\infty}$. If for all j $(-\infty < j < \infty)$ and $n \geq 1$

$$\begin{aligned}
&\left| \, P(E_2|E_1) - P(E_2) \, \right| \leq \phi(n) \, , \\
&1 \geq \phi(1) \geq \phi(2) \geq \ldots \, , \; \lim_{n\to\infty} \phi(n) = 0 \text{ , and} \\
&\sum_{n=1}^{\infty} [\phi(n)]^{1/2} < \infty \, ,
\end{aligned}$$

then the sequence $\{X_i\}$ satisfies the ϕ-mixing condition.

Let $F(x)$ be the (marginal) distribution function of X_i. Let $\hat{x}_p$ be the sample p-quantile ($100p^{\text{th}}$ percentiles); $\hat{x}_p = X_{(\lceil np \rceil)}$. Under the following conditions $\hat{x}_p$ $(0 < p < 1)$ has asymptotic normal distribution:

1. $F(x)$ is absolutely continuous in some neighbourhood of x_p,
2. the density function $f(x)$ is continuous , positive, and finite in some neighbourhood of x_p, and
3. $\{X_i\}$ satisfies the ϕ-mixing condition.

Futhermore, $\hat{x}_p$ is asymptotically unbiased, and $\lim\{n\mathrm{Var}(\hat{x}_p)\} = h_p(0)/f(x_p)^2$ as $n \to \infty$, where $h_p(w)$ is the spectral density of binary sequence $\{I_j(x_p)\}$ at frequency w:

$$I_j(x) = \begin{cases} 1 & \text{, if } X_j \leq x \, , \\ 0 & \text{, if } X_j > x \, . \end{cases}$$

II Estimating Spectral Densities

Most of the methods for estimating spectral densities $h(\omega)$ are based on the periodogram. The periodogram, $\{I(n/N)\}$, of sequence $\{z_j\}_{j=1}^N$ is defined by

$$I(n/N) = \frac{1}{N}\left|\sum_{j=0}^{N-1} z_{j+1}\exp(-i2\pi jn/N)\right|^2 .$$

The periodogram can be efficiently computed using the fast Fourier transform, especially when N is a power of two. Under very general conditions (see e.g. [16]) the periodogram ordinates have the following approximate properties:

$$\begin{aligned} \mathrm{E}[I(n/N)] &\approx h(n/N) & 0 < n < N/2\,, \\ \mathrm{Var}[I(n/N)] &\approx h(n/N)^2 & 0 < n < N/2\,, \\ \mathrm{Cov}[I(n/N, m/N)] &\approx 0 & 0 < n \neq m < N/2\,. \end{aligned}$$

Asymptotically periodogram ordinates are distributed as multiples of independent χ^2 random variables with two degrees of freedom. Hence the variances of periodogram ordinates do not decrease as N increases. There are two main approaches to reduce the variance of spectral density estimates. The first one is based on local, usually weighted averaging, and the second one on averaging over time segments.

In averaging over time segments, the sequence is divided into non-overlapping subsequences, each of length N_b. Let $z_{j,k}$, $j = 1, \ldots, N_b$
, , $k = 1, \ldots, K$ denote the $j + (k-1)N_b{}^{\mathrm{th}}$ observation in the original sequence. In each segment periodograms are evaluated:

$$I_k(n/N_b) = \frac{1}{N_b}\left|\sum_{j=0}^{N_b-1} z_{j+1,k}\exp(-i2\pi jn/N_b)\right|^2 .$$

The spectral density estimates are then

$$\hat{h}(n/N_b) = \frac{1}{K}\sum_{k=1}^{K} I_k(n/N_b)\ , \quad 0 < n < N_b/2\,.$$

In [24] it is shown that the variances are reduced by the factor K in practical situations. The spectral density estimates at different frequencies remain approximately uncorrelated. Hence if we average $2M+1$ adjacent spectral density estimates $\hat{h}((n-M)/N_b), \ldots, \hat{h}((n+M)/N_b)$, we obtain variance reduction by factor $K(2M+1)$. If only the non-overlapped averages are used, the estimates are still approximately uncorrelated.

Since the order of averagings does not chage the estimates, we can first evaluate the averaged spectral estimates in each segment

$$\hat{h}_k(\omega_j) = \frac{1}{2M+1} \sum_{m=-M}^{M} I_k(\omega_j + m/N_b)\,,$$
$$\omega_j = \frac{j(2M+1)-M}{N_b}\,,\ j = 1,\ldots,n < N_b/(4M+2)$$

and then the averages of segments

$$\hat{\hat{h}}(\omega_j) = \frac{1}{K}\sum_{k=1}^{K} \hat{h}_k(\omega_j)\,.$$

A common practice, established in [4], is to consider $\hat{\hat{h}}(\omega_j)/h(\omega_j)$ as a multiple of a χ^2 variable. The degrees of freedom is taken as $2\mathrm{E}^2\{\hat{\hat{h}}(\omega_j)\}/\mathrm{Var}\{\hat{\hat{h}}(\omega_j)\}$, which in our case is $2K(2M+1)$.

III Derivation of Approximately Unbiased Regression Estimates

We base our estimate of $h(0)$ on the regression model $\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}$, where

$$y_j = \log\{\hat{\hat{h}}(\omega_j)\} + \mu_1, \qquad \mathbf{X} = \begin{pmatrix} 1 & \omega_1 & {\omega_1}^2 \\ \vdots & \vdots & \vdots \\ 1 & \omega_n & {\omega_n}^2 \end{pmatrix},$$

and the ε's are independently and identically distributed: $E(\varepsilon) = 0$, and $E(\varepsilon^k) = \mu_k,\ k > 1$.

When $\hat{\hat{h}}(\omega_j)/h(\omega_j)$ is considered as a multiple of a χ^2 random variable with d degrees of freedom, the first four μ's are then (see [3]):

$$\begin{aligned} \mu_1 &\approx 1/d + 1/(3d^2)\,, \\ \mu_2 &\approx 2/(d-1)\,, \\ \mu_3 &\approx -4/(d-1)^2\,,\ \text{and} \\ \mu_4 &\approx 8/(d-1)^3 + 12/(d-1)^2\,. \end{aligned}$$

The least square estimate of $\boldsymbol{\alpha}$ is given by $\hat{\boldsymbol{\alpha}} = \mathbf{C}\mathbf{y} = \boldsymbol{\alpha} + \mathbf{C}\varepsilon$, where $\mathbf{C} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Hence $E\{(\hat{\alpha}_0 - \alpha_0)^k\} = \mu_k \sum {c_{1j}}^k = \mu'_k$.

Our parameter of interest is $\exp(\alpha_0)$, whose natural estimate is $\exp(\hat{\alpha}_0)$. In general $g(\hat{T})$ is a biased estimate of $g(T)$, but the bias can be reduced (see e.g. [6, p. 260]). Using Taylor expansion $\exp(\hat{\alpha}_0)$ can be written as

$$\exp(\hat{\alpha}_0) = \exp(\alpha_0)\sum_{k=0}^{\infty} \frac{(\hat{\alpha}_0 - \alpha_0)^k}{k!}\,.$$

Hence an approximately unbiased estimate of $h(0)$ is obtained as

$$\hat{h}(0) = \exp(\hat{\alpha}_0)/(1 + \mu'_2/2 + \mu'_3/6 + \mu'_4/24).$$

When a confidence interval is constructed using normal approximation, the degrees of freedom in the variance estimate is needed. A commonly used method is to consider $\hat{h}(0)/h(0)$ as a multiple of a χ^2 variate. The degrees of freedom is then taken as $2E\{\chi^2\}^2/\mathrm{Var}(\chi^2)$. Since $\hat{h}(0)/h(0) = (\sum(\hat{\alpha}_0 - \alpha_0)^k/k!)/(1 + \mu'_2/2 + \mu'_3/6 + \mu'_4/24)$, the degrees of freedom is about $3/(3\mu'_2 + 2\mu'_3 + \mu'_4)$.