

USING PARALLEL REPLICATIONS FOR SEQUENTIAL ESTIMATION OF MULTIPLE STEADY STATE QUANTILES

Mirko Eickhoff
Computer Science
University of Canterbury
Christchurch, New Zealand
m.eickhoff@cosc.canterbury.ac.nz

Don McNickle
Management
University of Canterbury
Christchurch, New Zealand
don.mcnicke@canterbury.ac.nz

Krzysztof Pawlikowski
Computer Science
University of Canterbury
Christchurch, New Zealand
krys.pawlikowski@canterbury.ac.nz

ABSTRACT

Simulation output data analysis in performance evaluation studies of complex stochastic systems such as the Internet is typically limited to mean values, even though it provides very limited information about the analysed system's performance. Quantile analysis is not as common, even though it can provide much deeper insights into the system of interest. A set of quantiles can be used to approximate a cumulative distribution function, providing full information about a given performance characteristic of the simulated system. In this paper, we will present two new methods for estimating steady state quantiles and distribution functions. The quantiles are estimated using simulation output data from concurrently executed independent replications. They are calculated sequentially and on-line to guarantee that their final statistical errors do not exceed a permitted threshold.

Categories and Subject Descriptors

I.6.6 [Simulation and Modeling]: Simulation Output Analysis; G.3 [Probability and Statistics]: Distribution Functions; Stochastic Processes

General Terms

Performance, Algorithms, Experimentation

Keywords

Discrete-event simulation, steady state, quantiles, sequential estimation

1. INTRODUCTION

Results of mean value analysis of simulation output data can answer such questions about a simulated system's behaviour as: "what is the mean delay experienced by a file sent across a telecommunication network?" or "how many customers, on average, are waiting in a queue?". Results of

quantile analysis can tell us, for example "what is the largest delay experienced by 90% of files sent across a telecommunication network?" or "what is the probability of more than k customers waiting in a queue?". Such results equip the decision-maker with much better knowledge about analysed systems than simple mean values.

The complexity of quantile estimation is higher than that of mean value estimation, but it can provide a deeper insight into the system of interest. This is true especially when several quantiles are estimated. A set of several quantiles can be used to approximate cumulative distribution functions (CDFs). The estimation of probability distributions is the ultimate goal of quantitative simulation, conducted for performance evaluation of such dynamic stochastic systems as the Internet, logistic systems, etc.

Here, we will discuss two new sequential methods for estimation of multiple quantiles of steady state distributions. The methods are based on multiple independent replications. The performance of the proposed methods is assessed analytically and evaluated by experiments.

In the next section we will review some basic results on estimation of quantiles and in Section 2.2 we show the advantages of selecting a set of quantiles with disjoint confidence intervals. This is followed by Section 3, which contains the discussion of two methods of quantile estimation. In Section 4 the methods are evaluated in experiments in which selected output processes are simulated. Conclusions are given in Section 5. All discussions and results which are presented in this article are an excerpt of [7].

2. QUANTILES

A short survey about quantile estimation in simulation output analysis can be found in [8]. In this section we would like to briefly discuss the basics of quantile estimation. Our discussion of order statistics and quantiles in Section 2.1 is mainly based on [5], [1] and [3].

2.1 Order Statistics and Quantiles

Let x_1, x_2, \dots, x_p be a set of observations of independent and identically distributed random variables X_1, X_2, \dots, X_p with common CDF $F_X(x)$. Furthermore, let $\{y_i\}_{i=1}^p$ be the ordered sequence of $\{x_i\}_{i=1}^p$, i.e. $y_1 \leq y_2 \leq \dots \leq y_p$. Then Y_i is called the i th order statistic and y_i is its realisation. Because $y_i \leq y_{i+1}$, order statistic are dependent and not

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Valuetools '07, October 23-25, 2007, Nantes, France
Copyright 2007 ICST 978-963-9799-00-4.

identically distributed. The CDF of Y_i is given by

$$\begin{aligned} F_{Y_i}(x) &= \Pr[Y_i \leq x] \\ &= \sum_{j=i}^p \binom{p}{j} (F_X(x))^j (1 - F_X(x))^{p-j}, \end{aligned} \quad (1)$$

see e.g. [20]. This equation allows the construction of distribution free confidence intervals for quantiles.

Let x_q define a value in the range of X , so that $F_X(x_q) = q$. Therefore,

$$x_q = F_X^{-1}(q) = \inf\{x | F_X(x) \geq q\}$$

is the population quantile of order q , if $F_X(x)$ is continuous. We will focus on continuous distributions only. If $F_X(x)$ is non-continuous this definition is ambiguous. We define the interval $[Y_l, Y_u]$ to be a distribution-free confidence interval for a population quantile, where

$$\begin{aligned} \Pr[Y_l \leq x_q \leq Y_u] &= \Pr[Y_l \leq x_q] - \Pr[Y_u < x_q] \\ &\geq \sum_{j=l}^{u-1} \binom{p}{j} q^j (1-q)^{p-j}. \end{aligned} \quad (2)$$

This equation can be derived from (1) with $F_X(x_q) = q$. Therefore, it is independent of the general form of $F_X(x)$. This property will be used extensively, because it enables us to establish a confidence interval for an unknown distribution.

The sample quantile \hat{x}_q estimates the population quantile x_q , when a certain value of q is specified. A common estimate is e.g. $\hat{x}_q = y_{\lfloor pq+1 \rfloor}$. However, many other estimators are known. For example the weighted sum of two neighbouring order statistics is another common estimator. In literature regarding simulation output data this is discussed e.g. in [22].

We will extend beyond estimation of x_q for a specified value of q by estimating the whole CDF of a given measure on the basis of several quantiles, assuming we are free to decide which values of q are appropriate, and choosing them on the basis of the size p of a given sample. A sorted random sample provides natural order statistics. Therefore, we are looking for the population quantile x_q that is represented by the expected value of the i th order statistic. $q = F_X(x_q)$ has to be estimated and x_q is given by y_i . We can see the dependence of q on the form of $F_X(x)$, thus, the bias of a general estimator of q depends on the sample size p . In [4] properties of the approximation

$$\mathbb{E}[Y_i] \approx F_X^{-1}\left(\frac{i}{p+1}\right) \quad (3)$$

are discussed. The error decreases with increasing sample size p and depends on derivatives of $F_X(x)$ as well as on the location of the quantile. Equation (3) suggests estimating q by $\hat{q}_i = i/(p+1)$. This estimate is approximately unbiased for any form of $F_X(x)$ and it is the best choice for the uniform case ([5]). If the form of $F_X(x)$ is given, specialised estimators are known. [5] shows that for the exponential case $\hat{q}_i^{(e)} = i/(p+0.5)$ and for the normal case $\hat{q}_i^{(n)} = (i-0.5)/p$ have better asymptotic properties than (3). However, a general solution for the unknown case is given by (3) and we can assume that $F_X(y_i) \approx \hat{q}_i$, if p is sufficiently large.

2.2 Disjoint Confidence Intervals

Equation (2) allows us to construct a confidence interval for a population quantile x_q based on two order statistics Y_l and Y_u . $\Pr[Y_l \leq x_q \leq Y_u]$ can be calculated for arbitrary ranks $1 \leq l \leq u \leq p$. It is not necessary, but could be desirable, that x_q splits the confidence interval $[Y_l, Y_u]$ in two parts, so that half of the probability mass is in each part.

DEFINITION 1. Let Y_c be an approximately unbiased estimator of x_q , i.e. $\mathbb{E}[Y_c] \approx F_X^{-1}(q) = x_q$. The confidence interval $\Pr[Y_l \leq x_q \leq Y_u] \geq 1 - \alpha$ is balanced if

$$\begin{aligned} \Pr[Y_l \leq x_q \leq Y_c] &\geq \frac{1 - \alpha}{2} \quad \text{and} \\ \Pr[Y_c \leq x_q \leq Y_u] &\geq \frac{1 - \alpha}{2} \end{aligned} \quad (4)$$

hold.

This definition comes from the concept of mid-p confidence intervals, see e.g. [2]. Other common approaches are to construct a confidence interval that has minimum width or that has $x_q - y_l = y_u - x_q$. However, we construct the confidence interval on basis of (4) because in the balanced case u and l can be calculated independently of each other.

Let the two confidence intervals $\Pr[Y_{l_1} \leq x_{q_1} \leq Y_{u_1}] \geq 1 - \alpha$ and $\Pr[Y_{l_2} \leq x_{q_2} \leq Y_{u_2}] \geq 1 - \alpha$ be balanced. Estimates of x_{q_1} and x_{q_2} are dependent on each other because they are taken from the same sample. However, by choosing disjoint confidence intervals, i.e. $u_1 \leq l_2$, we can assure that at most $\frac{\alpha}{2}$ of the probability mass of both distributions, given by (2), overlap. If α is sufficiently small, e.g. $\alpha \leq 0.1$, high correlation between estimates of x_{q_1} and x_{q_2} can be avoided. To split the ordered sequence $\{y_i\}_{i=1}^p$ into a maximum number of disjoint balanced confidence intervals an algorithmic approach is needed and described in [8]. An example of a selected set of quantiles with $p = 999$ and $\alpha = 0.05$ is shown in Figure 1. Equation (2) is applied for the selected q -quantiles and $\Pr[Y_i \leq x_q \leq Y_{i+1}]$ is depicted for all i . The overlapping parts of the graphs show the remaining dependence among neighbouring quantiles.

3. NEW METHODS OF SEQUENTIAL ANALYSIS

We assume that simulation output data are collected during steady state, i.e. data are collected after observation index l_F , where

$$\forall (t \geq l_F, \Delta \geq 0, x) : F_{X_t}(x) \approx F_{X_{t+\Delta}}(x). \quad (5)$$

The truncation point l_F specifies the boundary between the initial transient phase of simulation and the onset of analysis of the probability distribution in steady state. A method for determining this point is presented in [9]. Note that this truncation point is different from the truncation point which would be used in the mean value analysis. This new method of quantile analysis is based on concurrent independent replications of simulation. In such a scenario, replications of the same simulation are executed in parallel, using either different pseudo random number generators, or non-overlapping sequences of pseudo-random numbers from the same generator.

Using p independent replications, we obtain the observations $x_{j,i}$, where $1 \leq j \leq p$ is the replication index and

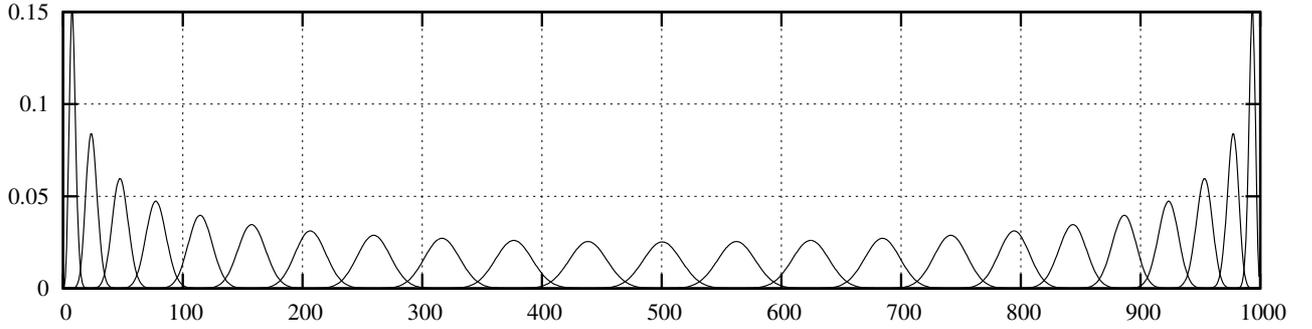


Figure 1: $\Pr[Y_i \leq x_q \leq Y_{i+1}]$ (ordinate), see (2), versus rank i (abscissa) for selected values of q .

$l_F \leq i < \infty$ is the observation index. The independence of all replications implies that the observations $\{x_{j,i}\}_{j=1}^p$ are independent of each other. This is valid for all observation indexes i . Therefore, for a fixed observation index the statistical methods, which are valid for independent and identically distributed random samples, are directly applicable to $\{x_{j,i}\}_{j=1}^p$.

The common definition of the population quantile has to be extended by the observation index i :

$$x_{q,i} = F_{X_i}^{-1}(q) = \inf\{x | F_{X_i}(x) \geq q\}.$$

Let $\{y_{j,i}\}_{j=1}^p$ be the ordered values of $\{x_{j,i}\}_{j=1}^p$ and let $\{y_{j,i}\}_{i=l}^{\infty}$ be a realisation of the stochastic process $\{Y_{j,i}\}_{i=l}^{\infty}$. In this sense $Y_{j,i}$ represents the j th order statistic at observation index i .

3.1 Mean of Order Statistics

Our aim is to estimate several quantiles of $F_{X_{\infty}}(x)$. Here, a natural approach is to estimate those quantiles x_{q_j} which are represented by the order statistics $Y_{j,i}$. For an unknown distribution, q_j can be estimated by (3). The difference in distributions of X_i , where $i \geq l_F$, is negligible. Therefore, the mean of the j th order statistics

$$\hat{X}_{\hat{q}_j} = \frac{1}{n - l_F + 1} \sum_{i=l_F}^n Y_{j,i} \quad (6)$$

is a point estimator of $F_{X_{\infty}}^{-1}(q_j)$.

THEOREM 1. *The mean of the j th order statistics $\hat{X}_{\hat{q}_j}$ is an approximately unbiased estimator of $F_{X_{\infty}}^{-1}(q_j)$ for large p and $i \geq l_F$.*

PROOF. The expected value of (6) is

$$\mathbb{E}[\hat{X}_{\hat{q}_j}] = \frac{1}{n - l_F + 1} \sum_{i=l_F}^n \mathbb{E}[Y_{j,i}]. \quad (7)$$

$\mathbb{E}[Y_{j,i}] = F_{X_i}^{-1}(q_j)$ holds for large values of p , see (3) and [4] or [5]. Furthermore, all $X_{l_F}, X_{l_F+1}, \dots$ are assumed to be identically distributed, i.e. $\forall(i \geq l_F) : F_{X_i}(x) = F_{X_{\infty}}(x)$.

Equation (7) simplifies to

$$\begin{aligned} \mathbb{E}[\hat{X}_{\hat{q}_j}] &= \frac{1}{n - l_F + 1} \sum_{i=l_F}^n F_{X_i}^{-1}(q_j) \quad (8) \\ &= \frac{1}{n - l_F + 1} \sum_{i=l_F}^n F_{X_{\infty}}^{-1}(q_j) \\ &= F_{X_{\infty}}^{-1}(q_j). \end{aligned}$$

The estimator $\hat{X}_{\hat{q}_j}$ is asymptotically unbiased, i.e. $\mathbb{E}[\hat{X}_{\hat{q}_j}] - F_{X_{\infty}}^{-1}(q_j) = 0$, because (8) holds for large p and $i \geq l_F$. \square

Every simulation is a statistical experiment. Point estimators never return exact values, even if they are unbiased. Confidence intervals are essential to provide convincing results. To establish a confidence interval for (6) its variance $\text{Var}[\hat{X}_{\hat{q}_j}]$ is needed. Note, that all $Y_{j,l_F}, Y_{j,l_F+1}, \dots$ are autocorrelated and the variance cannot be estimated directly. The form of (6) is, however, identical to mean value estimators of single simulation runs. Its special feature is that each component represents a quantile. Therefore, known techniques for variance estimation of mean value estimators can be applied. Spectral analysis (see e.g. [13]) and batch means (see e.g. [10]) are commonly used in mean value analysis.

Both, spectral analysis and batch means, have already been suggested in [12] for variance estimation in quantile analysis. The maximum transformation is used to obtain extreme quantiles of the output process. Here, we replace the maximum transformation with (6) and extend the method to multiple independent replications. Further details on how batch means and spectral analysis methods are applied can be found in [6].

Sequential analysis is the only way to obtain estimates with controlled error. Equation (6) is suitable for a sequential approach because n can be extended. Extensions of batch means and spectral analysis to a sequential approach are discussed in [15]. Let Δ_{q_j} be the halfwidth of the confidence interval of point estimate $\hat{x}_{\hat{q}_j}$, calculated on basis of $\text{Var}[\hat{X}_{\hat{q}_j}]$ and the Student t-distribution. The stopping criterion $\Delta_{q_j}/D \leq \epsilon_{max}$ can be used to stop a sequential approach. D is a value which is used to standardise the halfwidth of the confidence interval and ϵ_{max} is the maximum acceptable relative error, where $0 < \epsilon_{max} \leq 0.1$. In mean value analysis D is usually the point estimate itself. However, here we have to take into account that it is quite

likely that $F_{X_i}^{-1}(q_j) \approx 0$ holds for one of our quantile estimates. Furthermore, it is desirable to standardise all quantile estimates by the same value D . Therefore, we choose $D = \hat{x}_{\hat{q}_p} - \hat{x}_{\hat{q}_1}$, which is the estimated range of $F_{X_\infty}(x)$. This guarantees small relative errors for all estimated quantiles.

3.2 Pooling Spaced Data

Approximate independence can be achieved by establishing a pool of observations, which are spaced far apart from each other. Let s be an adequate space size. Then X_{l_F} , X_{l_F+s} , X_{l_F+2s} , \dots can be regarded as nearly independent. When using p replications the pool of observations is given by

$$P = \{\{x_{j,l_F+is}\}_{j=1}^p\}_{i=0}^\infty.$$

The size of this pool is unbounded and contains approximately independent and identically distributed data if l_F and s are large. Then, standard quantile estimators are directly applicable to estimate $F_{X_\infty}^{-1}(q)$.

The determination of an adequate value of s is similar to the determination of a batch size for batch means, as pooling is just a special kind of batching: The batch statistic is the first value of a batch. For this task correlation tests are needed. Many different correlation tests are known, like run tests (see e.g. [19]) or permutation tests (see e.g. [21]). The von Neumann ratio test is probably the most recommended correlation test today (see e.g. [11]). However, we wish to estimate a set of quantiles and, thus, have to find an overall space size s that is valid for all sequences $\{x_{j,l_F+is}\}_{i=0}^\infty$, where $1 \leq j \leq p$. Thus, we select a correlation test based on permutations and the median confidence interval (see e.g. [18]), which is described next.

Let $\hat{r}^{(p)}(P_1)$ be Pearson's correlation coefficient of the original lag-1 paired spaced sequence

$$\{(x_{j,l_F+is}; x_{j,l_F+(i+1)s})\}_{i=0}^{n-1},$$

and $\hat{r}^{(p)}(P_k)$ be Pearson's correlation coefficient for the lag-1 paired data of the k th permutation of $\{x_{j,l_F+is}\}_{i=0}^n$ with $2 \leq k \leq (n!)$. In [17] the first four moments of Pearson's correlation coefficient are derived. Here, the first and the third moment are of special interest: $E[\hat{r}^{(p)}] = 0$ holds even for small samples and $\text{Skew}[\hat{r}^{(p)}] = 0$ holds approximately. $\text{Skew}[\hat{r}^{(p)}]$ defines the degree of asymmetry of the distribution of $\hat{r}^{(p)}$. Therefore, $F_{\hat{r}^{(p)}}(0) = 0.5$ is approximately true. The null hypothesis of our test is that $\{x_{j,l_F+is}\}_{i=0}^n$ is an independent sequence.

$$\Pr \left[|\hat{r}^{(p)}(P_k)| < |\hat{r}^{(p)}(P_1)| \right] = \frac{1}{2}$$

holds under the null hypothesis and for a randomly chosen permutation P_k . For K randomly chosen permutations P_{k_1}, \dots, P_{k_K} we can derive

$$\Pr \left[\forall l (1 \leq l \leq K) : |\hat{r}^{(p)}(P_{k_l})| < |\hat{r}^{(p)}(P_1)| \right] = \frac{1}{2^K}.$$

On the basis of this equation the following confidence interval can be established:

$$\Pr \left[-\Delta \leq \hat{r}^{(p)}(P_1) \leq \Delta \right] = 1 - \frac{1}{2^K}$$

with halfwidth

$$\Delta = \max_{1 \leq i \leq K} \left(|\hat{r}^{(p)}(P_{k_i})| \right).$$

If $\hat{r}^{(p)}(P_1)$ is not within the confidence interval, the null hypothesis must be rejected at significance level $1 - \frac{1}{2^K}$. The advantage of using this confidence interval is that the assumption of zero skewness is milder than the assumption of a normal distribution. For only $K = 6$ permutations the significance level is already > 0.95 and K can be regarded as a value of choice. For our purpose of estimating the overall space size s for p independent replications this correlation test is performed on $\{x_{j,l_F+is}\}_{i=0}^n$ for each j .

A sequential approach needs to be able to extend the sample size. By adding an additional sequence $\{x_{j,l_F+s(n+1)}\}_{j=1}^p$ of previously unprocessed observations at index $n+1$ the sample size can be extended by p observations. For quantile estimation based on order statistics the sample has to be sorted. The most efficient way of sorting in this case is to merge two already sorted samples. Let assume that P of size pn is already sorted. The new sample $\{x_{j,l_F+s(n+1)}\}_{j=1}^p$ can be sorted in $O(p \log(p))$. Merging of P and the new observations can be done in $O(pn+p)$. So the total runtime of adding new observations to a sorted pool of data is $O(p \log(p) + p(n+1))$. Because usually $n \gg p$ holds, we can simplify the runtime to $O(pn)$, which is efficient.

Because our aim is to estimate $F_{X_\infty}(x)$ on base of several quantiles we select quantiles as described in [8]. This method selects quantiles so that their confidence intervals do not overlap. This implies that their probability mass is distributed to different order statistics, see Figure 1. Therefore, high correlation between quantile estimates is avoided. How many quantiles are selected, depends on the sample size pn . Here, a stopping criterion could be defined by simply setting a minimum number of quantiles which should be selected. The bigger the sample size the more quantiles can be selected. The simulation experiment can be stopped if the sample is large enough to estimate the minimum number of quantiles with disjoint confidence intervals.

On the other hand, a stopping criterion can depend on the size of the confidence interval. Let the confidence interval $\Pr[Y_l \leq x_q \leq Y_u] \geq 1 - \alpha$ be a confidence interval of the quantile x_q , where Y_i is the i th order statistic of P . Similar to the stopping criterion of the previous section, we can demand

$$\frac{Y_u - Y_l}{2(Y_{pn} - Y_1)} \leq \epsilon_{max}.$$

$Y_u - Y_l$ is divided by 2 to produce the halfwidth. It is standardized by the range $Y_{pn} - Y_1$ to avoid a division by a value close to zero and to standardise all quantiles by the same value.

3.3 Discussion

In the quantile estimation method of Section 3.1 the sample size p is given by the number of replications. Our focus is on performing all p replications in parallel. Then, the maximum number of replications is usually restricted by the number of workstations in a LAN, or the number of processors in a grid computing system. However, replications could be performed serially involving the overhead of storing the output data in adequate data structures. If p is not large enough to fulfil the conditions of Theorem 1, we can expect biased quantile estimates. However, the method of Section 3.1 does not allow to apply a mechanism which would control correlation between quantile estimates in one set.

This problem is solved by the quantile estimation method

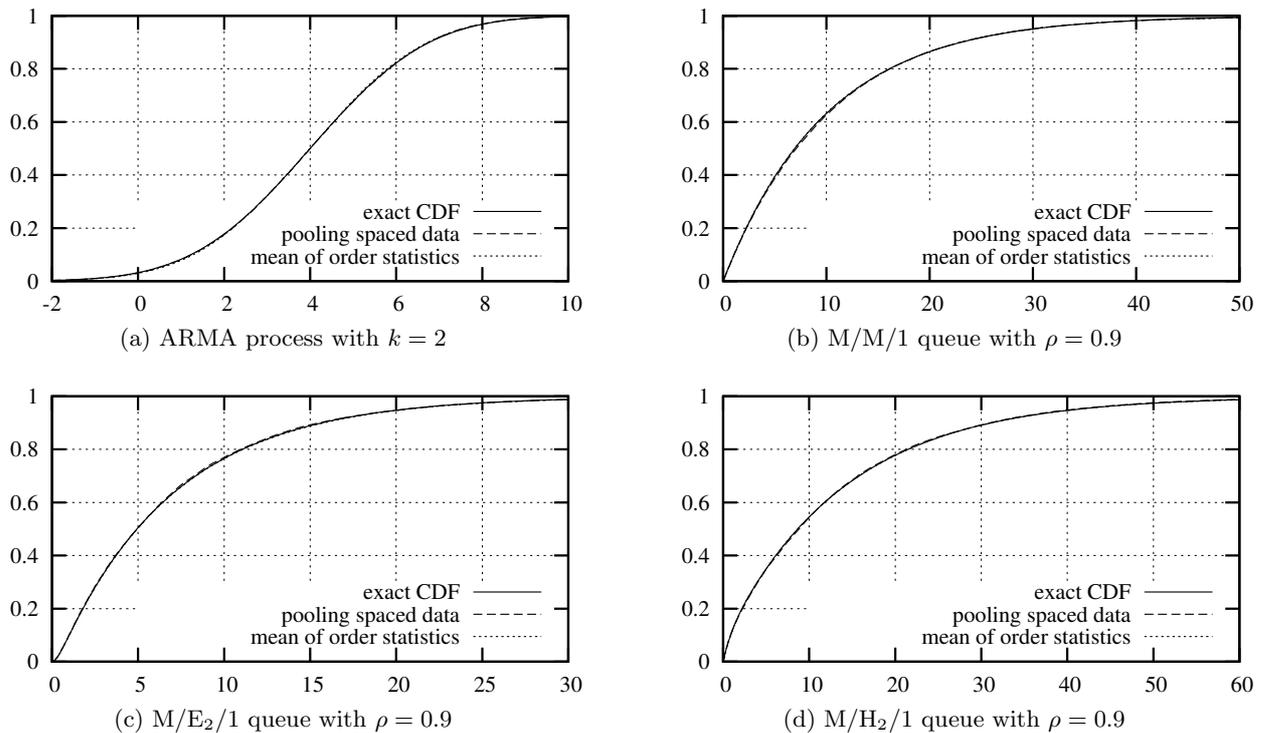


Figure 2: Exact and estimated CDFs.

of Section 3.2. Here, the sample P is increased until all quantiles have sufficiently small statistical error. The correlation between the quantile estimates is controlled by choosing disjoint confidence intervals. However, the drawback is that the memory requirement of this method is not constant.

4. EXAMPLES

In this section we will test the quantile estimation methods of Section 3.1 and 3.2 on various examples with known steady state distribution function. We will compare the estimated CDF with the expected CDF by comparing their shapes. Furthermore, we will perform coverage analysis for all estimated quantiles, which is done according to [16].

Example 1 is a geometrical ARMA process defined by

$$X_t = 1 + \epsilon_t + \sum_{i=1}^k \frac{1}{2^i} (X_{t-i} + \epsilon_{t-i}),$$

where ϵ_t is a Gaussian white noise process, thus, $F_{X_\infty}(x)$ is normally distributed. $E[X_\infty] = 4$ and $\text{Var}[X_\infty] = 117/25$ for $k = 2$ and $F_{X_\infty}(x) = N(x; 4, 117/25)$ follows. We chose $X_t = 0$ for $t \leq 0$, but this does not influence the steady state behaviour.

Example 2 is the response time X_t of an M/M/1 queue with $\mu = 1$ and $\lambda = 0.9$ so that $\rho = 0.9$. Here, we expect $F_{X_\infty}(x) = 1 - e^{-x\mu(1-\rho)}$ (see e.g. [14]) and the coefficient of variation is 1.

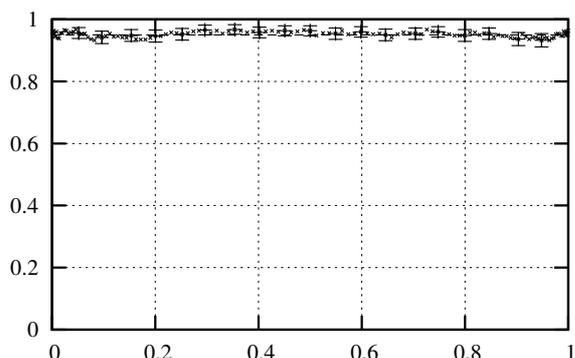
Example 3 is the response time X_t of an M/E₂/1 queue with $\mu = 1/0.45$ and $\lambda = 1$ so that $\rho = 0.9$. $F_{X_\infty}(x)$ can be calculated by inverting the Laplace-Stieltjes transform of the response time distribution using Maple. We avoid giving the exact form of $F_{X_\infty}(x)$ because of its complexity. Here,

the coefficient of variation of the service time is $1/\sqrt{2}$.

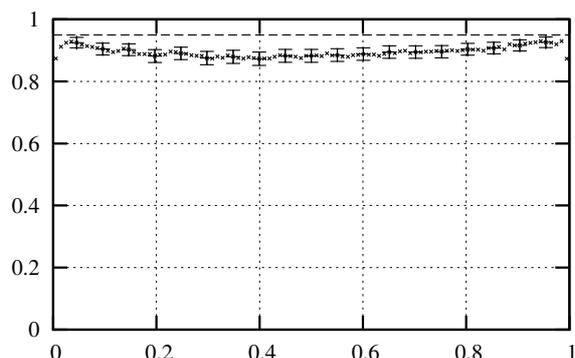
Example 4 is the response time X_t of an M/H₂/1 queue, where the service time is given by a 2 dimensional hyper-exponential distribution. To give a squared coefficient of variation equal to 2 with $\rho = 0.9$ we set $\lambda = 1$, $\mu_1 \approx 0.4696$, $\mu_2 \approx 1.7526$ and the probability of choosing μ_1 is ≈ 0.2113 . These parameters and $F_{X_\infty}(x)$ were again calculated with Maple.

We started all queueing models with an empty queue and no customer in service. For every simulation experiment we performed $p = 99$ replications in parallel. The run length of each simulation experiment is not fixed because we applied sequential analysis of the output data, as described in Section 3. In Figure 2 we can see the exact CDF compared with the estimated CDFs from *mean of order statistics* (Section 3.1) and *pooling spaced data* (Section 3.2) for all examples. The expected graph is barely distinguishable from estimated graphs. A Q-Q plot of those graphs confirms this assumption. This shows that the estimates from both methods are very close to the exact CDF.

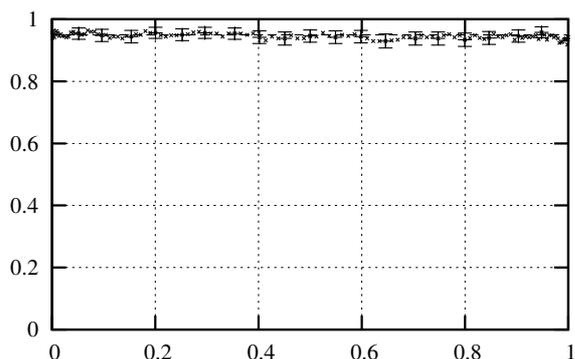
However, we present also results in a practical simulation framework and perform sequential coverage analysis for every estimated quantile separately. The results are presented in Figure 3. The abscissa shows q of the quantile x_q and the ordinate shows the calculated coverage. In all examples and for all quantiles the coverage of 0.95 is expected. We can see that the performance of the quantile estimation by pooling spaced data is very good because the coverage for all quantiles in all examples is around 0.95. This shows that estimates are approximately unbiased and the estimated confidence intervals have an appropriate size. For all experiments with pooling spaced data we applied \hat{q}_i , as defined in Sec-



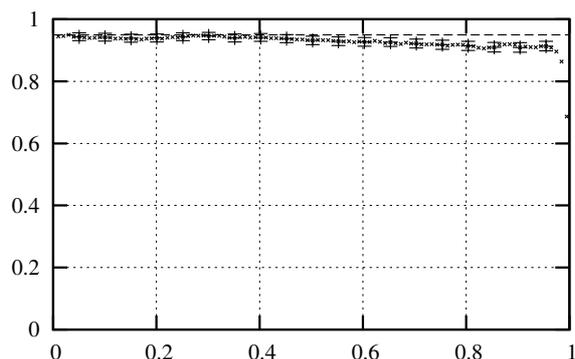
(a) PSD using \hat{q}_i : ARMA process with $k = 2$



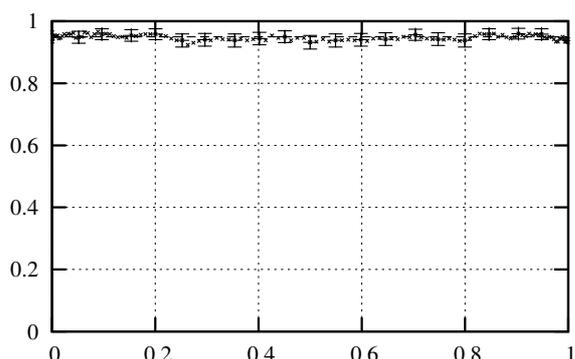
(b) MOS using $\hat{q}_i^{(n)}$: ARMA process with $k = 2$



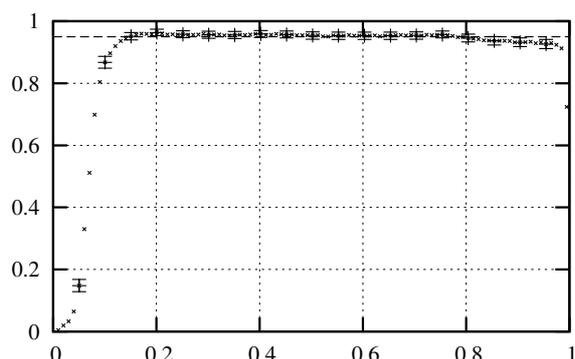
(c) PSD using \hat{q}_i : M/M/1 queue with $\rho = 0.9$



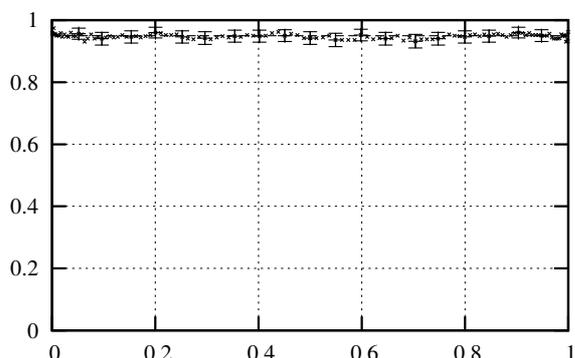
(d) MOS using $\hat{q}_i^{(e)}$: M/M/1 queue with $\rho = 0.9$



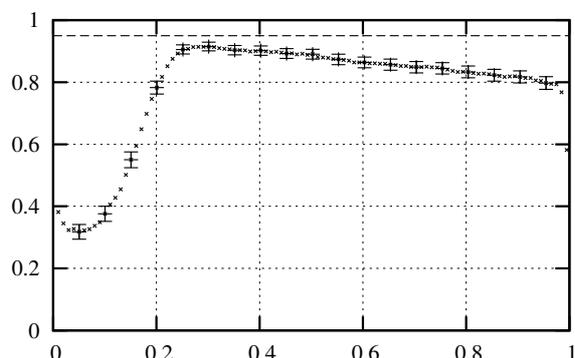
(e) PSD using \hat{q}_i : M/E₂/1 queue with $\rho = 0.9$



(f) MOS using $\hat{q}_i^{(e)}$: M/E₂/1 queue with $\rho = 0.9$

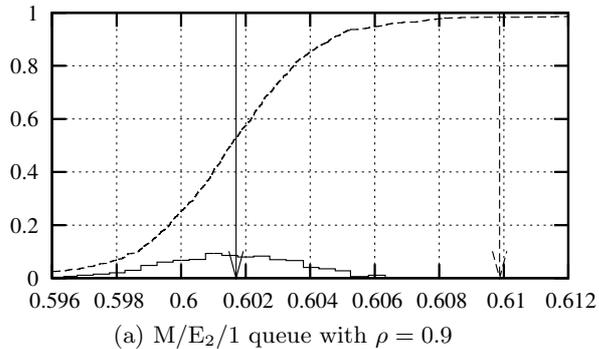


(g) PSD using \hat{q}_i : M/H₂/1 queue with $\rho = 0.9$

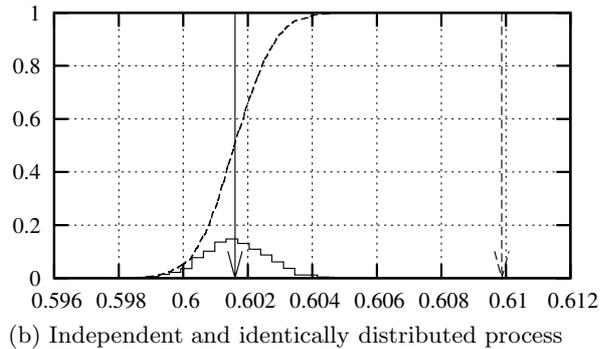


(h) MOS using $\hat{q}_i^{(e)}$: M/H₂/1 queue with $\rho = 0.9$

Figure 3: Coverage (ordinate) of q-quantiles (abscissa) estimated by pooling spaced data (PSD) and by mean of order statistics (MOS).



(a) M/E₂/1 queue with $\rho = 0.9$



(b) Independent and identically distributed process

Figure 4: Empirical CDF of the mean of the 5th order statistic, where $p = 99$.

tion 2.1, because the accuracy of final estimates is fine and, thus, there is no need for the specialised cases $\hat{q}_i^{(n)}$ or $\hat{q}_i^{(e)}$.

The coverage of the quantiles estimated by the mean of order statistics is almost as expected for Example 1 and 2, as one can see in Figures 3(b) and 3(d). The coverage is significantly smaller than 0.95 for extreme quantiles. However, here we used our knowledge of the form of the distribution function and applied the more specialised estimates $\hat{q}_i^{(e)}$ and $\hat{q}_i^{(n)}$, as defined in Section 2.1. In Figures 3(f) and 3(h) one can see that the coverage of the mean of order statistics for Example 3 and 4 is not as expected. Even non extreme quantiles show a coverage significantly smaller than 0.95. For this examples none of the estimators \hat{q}_i , $\hat{q}_i^{(e)}$ and $\hat{q}_i^{(n)}$ are optimal, the best results we obtained by applying $\hat{q}_i^{(e)}$. The bad coverage is caused by a constant value of the sample size p . Here, all order statistics provide a slightly biased estimate. The calculation of the mean does not eliminate this bias.

To prove this we draw the empirical distribution of the mean of the 5th order statistics for Example 3 in Figure 4(a). The bold arrow marks the position of the overall mean and the dashed arrow marks the expected value. The distribution is not centred around the expected value, thus, the estimator is biased. However, one can also see that the difference between the expectation and the mean $F_X^{-1}(x) - \hat{F}_X^{-1}(x) \approx 0.0082$ is small. This explains why the distributions in Figure 2(c) are almost identical and it shows that a constant sample size p causes the bias.

To show that the estimation of the variance by batch means or spectral analysis is not the problem we did another series of experiments. In this case we used an independent and identically distributed output process. Here, the data is drawn directly from the steady state distribution of Example 3. The empirical distribution of the mean of the 5th order statistics is depicted in Figure 4(b). The distribution of this process is more focused (smaller variance) compared to Figure 4(a) and the bias still remains approximately equal 0.0082. This shows that batch means or spectral analysis does not influence the result. The constant sample size p is the only source of bias.

Statistical accuracy of sequential estimation methods, in terms of low bias, is our primary performance measure of interest. Efficiency of sequential estimation methods, in terms of minimum number of observations to fulfil the stopping criterion, is a secondary performance measure. Even though we

believe that there is no significant difference in the efficiency of pooling spaced data and the mean of order statistics we did no detailed investigation on this topic. This is because pooling spaced data is clearly preferable due to higher accuracy.

5. CONCLUSIONS

We have suggested two different methods for steady state estimation of quantiles in simulation output analysis of multiple replications. The first method is to calculate the mean of order statistics of a constant sample size p , whereas the second method uses a pool P of independent observations.

The examples show that the pool of observations provides valid quantile estimates. The statistical properties of this method are as expected. This is true for all examined examples and all estimated quantiles. Therefore, this method is robust. It is suitable for automated simulation analysis because no previous knowledge is assumed.

The mean of order statistics only provides good quantile estimates if the general form of the distribution is known. In other cases the estimates are biased because the sample size p is constant. A constant sample size p violates the assumption of Theorem 1.

We can see that the use of multiple replications enables new methods of quantile estimation. Because of the independence of the replications standard quantile estimators can be applied. In Figure 2 we can see that the steady state distribution can be estimated. The estimated distributions are almost indistinguishable from the expected distributions. A small statistical error is guaranteed by a sequential approach of quantile analysis.

6. REFERENCES

- [1] B. C. Arnold and N. Balakrishnan. *Lecture Notes in Statistics: Relations, Bounds and Approximations for Order Statistics*. Springer, 1989.
- [2] G. Berry and P. Armitage. Mid-p confidence intervals: A brief review. *The Statistician*, 44(4):417–423, 1995.
- [3] W. Conover. *Practical Nonparametric Statistics*. John Wiley & Sons, Inc., New York, 1999.
- [4] F. N. David and N. L. Johnson. Statistical treatment of censored data. *Biometrika*, 41(1):228–240, June 1954.
- [5] H. A. David. *Order Statistics*. John Wiley & Sons, Inc., 1970.

- [6] M. Eickhoff. Steady state quantile estimation. In *Proceedings of the 13th GI/ITG Conference on Measurement, Modeling and Evaluation of Computer and Communication Systems (MMB'06)*, pages 155–171, 2006.
- [7] M. Eickhoff. *Sequential Analysis of Quantiles and Probability Distributions by Replicated Simulations*. Ph.D. thesis, Computer Science and Software Engineering, University of Canterbury, 2007.
- [8] M. Eickhoff, D. McNickle, and K. Pawlikowski. Analysis of the time evolution of quantiles in simulation. *International Journal of Simulation*, 7(6):44–55, 2006.
- [9] M. Eickhoff, D. McNickle, and K. Pawlikowski. Detecting the Duration of Initial Transient in Steady State Simulation of Arbitrary Performance Measures. In *Proceedings of the 2nd International Conference on Performance Evaluation Methodologies and Tools (Valuetools'07)*, 2007.
- [10] G. S. Fishman. Grouping observations in digital simulation. *Management Science*, 24(5):510–521, January 1978.
- [11] G. S. Fishman and L. S. Yarbber. An implementation of the batch means method. *INFORMS Journal on Computing*, 9(3):296–310, Summer 1997.
- [12] P. Heidelberger and P. Lewis. Quantile estimation in dependent sequences. *Operations Research*, 32(1):185–209, February 1984.
- [13] P. Heidelberger and P. D. Welch. A spectral method for confidence interval generation and run length control in simulations. *Communications of the ACM*, 24(4):233–245, April 1981.
- [14] R. Jain. *The Art of Computer Systems Performance Analysis*. John Wiley & Sons, Inc., 1991.
- [15] K. Pawlikowski. Steady-state simulation of queueing processes: a survey of problems and solutions. *ACM Computing Surveys*, 22:123–170, June 1990.
- [16] K. Pawlikowski, D. McNickle, and G. Ewing. Coverage of confidence intervals in sequential steady-state simulation. *Journal of Simulation Practise and Theory*, 6(3):255–267, 1998.
- [17] E. J. G. Pitman. Significance tests which may be applied to samples from any populations. ii. the correlation coefficient test. *Supplement to the Journal of the Royal Statistical Society*, 4(2):225–232, 1937.
- [18] J. C. Strelen. The accuracy of a new confidence interval method. In *Proceedings of the 2004 Winter Simulation Conference*, pages 654–662, 2004.
- [19] F. S. Swed and C. Eisenhart. Tables for testing randomness of grouping in a sequence of alternatives. *The Annals of Mathematical Statistics*, 14(1):66–87, March 1943.
- [20] W. R. Thompson. On confidence ranges for the median and other expectation distributions for populations of unknown distribution form. *The Annals of Mathematical Statistics*, 7(3):122–128, September 1936.
- [21] A. Wald and J. Wolfowitz. Statistical tests based on permutations of the observations. *The Annals of Mathematical Statistics*, 15(4):358–372, December 1944.
- [22] D. C. Wood and B. Schmeiser. Overlapping batch quantiles. In *Proceedings of the 1995 Winter Simulation Conference*, pages 303–307, 1995.