DEPICTION OF TRANSIENT PERFORMANCE MEASURES USING QUANTILE ESTIMATION

Mirko Eickhoff,
Donald C. McNickle,
Krzysztof Pawlikowski

University of Canterbury
Christchurch, New Zealand
E-Mail: m.eickhoff@cosc.canterbury.ac.nz

KEYWORDS

Quantile Estimation, Transient Analysis, Multiple Replications in Parallel, Selection of Several Quantiles

ABSTRACT

For simulation output the estimation of several quantiles usually provides a deeper insight than mean value analysis. So far, quantile estimation has usually been applied to show the long run behaviour of a system. In this paper we describe a method to depict several quantiles over simulation time to show the transient behaviour. This method is based on independent replications and its capability is demonstrated by examples with different kinds of transient behaviour.

INTRODUCTION

The purpose of steady-state simulation is to study the long-run behavior of a system. Using estimators for mean value analysis, the results of the simulation can answer questions about the average system state like: How many customers are there on average in the queue? On the other hand, quantiles are known to be more robust against outliers than mean values. Quantile estimation can also answer questions like: What is the probability of more than k customers in the queue? Questions of this kind are often of more interest to the decision-maker. The complexity of quantile estimation is higher than the complexity of mean value estimation, but the estimation of quantiles can give 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 estimate the steady-state distribution function. The estimation of the steady-state distribution is the ultimate goal in steady-state simulation. For details on quantile estimation see e.g. [Heidelberger and Lewis, 1984], [Jain and Chlamtac, 1985], [Raatikainen, 1987] and [Chen and Kelton, 1999].

An extension of the estimation of several quantiles is to estimate these quantiles over model time. This provides a deep insight into the transient behaviour of the system of interest. In steady-state simulation this is useful to verify if a steady-state phase exists, i.e. that the probability distribution function of the analyzed performance measure converges to a steady-state distribution function. Furthermore it can be verified, e.g, whether the transient behaviour is monotone or if there are some unexpected issues which demand further investigation. However, the truncation point for the estimation of steady-state performance measures should not be determined by a visual inspection of these quantiles over time (compare [Bause and Eickhoff, 2003]).

In applications finite-horizon simulation is frequently used to examine a given situation with a certain initial state. In contrast to steady-state simulation the transient behavior of the system is the central point of analysis, even though, the usual approach is to estimate only mean values. Again, the estimation of several quantiles over time provides a deeper insight and extends the results of commonly used approaches. For a comparison of finite-horizon simulation and steady-state simulation see e.g. [Law and Kelton, 2000] and [Alexopoulos and Kim, 2002].

The main problem in quantile estimation for steady-state performance measures is that the output data $X_1, X_2, \ldots$ of a single simulation run is typically not stationary and is autocorrelated (see e.g. [Lee et al., 1999]). Therefore, the amount of required output data can be immense, which causes a problem when storing and sorting the output data. Using p independent replications of the simulation is a well known approach to obtain independent sequences of output data. If these replications are synchronized (see [Bause and Eickhoff, 2002]) an independent and identically distributed (iid) random sample $\{x_{j,i}\}_{j=1}^{p}$ of p observations of X_i is available at each observation index i. This property helps to overcome the main problem of quantile estimation in a single simulation run and enables the use of traditional quantile estimators for iid random samples.

Let $F_i(x) = \Pr\{X_i \leq x\}$ denote the cumulative probability distribution function of X_i. The q-quantile at observation index i is defined by the equation $q = F_i(x_q)$ and, therefore,

$$x_q = \inf\{x : F_i(x) \geq q\} = F_i^{-1}(q)$$

is the location of the q-quantile in the case of a continuous distribution $F_i(x)$. Let $\{y_{j,i}\}_{j=1}^{p}$ be the order statistic of $\{x_{j,i}\}_{j=1}^{p}$. A valid estimator for the location of the q-quantile at observation index i is given by

$$\hat{x}_q = y_{\lceil pq \rceil, i}$$

The half width of a confidence interval of $\hat{x}_q$ can be described in two ways:

$$\hat{x}_q \in x_q \pm \epsilon'_q \quad \text{and} \quad \hat{x}_q \in x_{q \pm \epsilon_q}$$

Proceedings 19th European Conference on Modelling and Simulation
Yuri Merkuryev, Richard Zobel, Eugène Kerckhoffs © ECMS, 2005
ISBN 1-84233-112-4 (Set) / ISBN 1-84233-113-2 (CD)

ϵ'_q describes an interval in the range of the measure and ϵ_q describes an interval in the range of the probability (see [Chen and Kelton, 1999]). Note, the interval $q \pm \epsilon_q$ should not exceed the bounds 0 and 1. In nonparametric statistics ϵ'_q can be calculated from

$$\begin{aligned} \Pr\{y_{l,i} \leq x_q < y_{u,i}\} &= 1 - \alpha_{l,u} \quad (1) \\ &= \sum_{j=l}^{u-1} \binom{p}{j} q^j (1-q)^{p-j} \end{aligned}$$

by decreasing l and increasing u until the chosen confidence level $(1 - \alpha) \leq (1 - \alpha_{l,u})$ is reached (see [Conover, 1999] and [Heidelberger and Lewis, 1984]). In [Chen and Kelton, 1999] is shown that ϵ_q can be chosen from the inequality

$$p \geq \frac{z^2_{1-\alpha/2} q(1-q)}{\epsilon_q^2} \quad (2)$$

where $z_{1-\alpha/2}$ is the $1 - \alpha/2$ quantile of the standard normal distribution. Both, Equation (1) and Inequality (2) do not depend on the output data itself. Therefore, both formulas can be used to estimate the half width before the simulation experiment starts. However, both formulas mainly depend on the number of replications p, because the confidence level $1-\alpha$ can be considered as a constant parameter. Therefore, p is the most important parameter in the methods described in subsequent sections.

To show the transient behavior of the system of interest a plot of several quantiles over time is needed. The quantiles should be chosen with non overlapping confidence intervals. Therefore, a method is needed which determines adequate quantiles based on parameter p, because the half width of the confidence interval of $\hat{x}_q$ depends on the number of replications p. In the following section two alternative methods are proposed and discussed. The better method is used to examine examples with a variety of different transient behaviors. In the last section some conclusions are given.

SELECTION OF QUANTILES

As already pointed out, the calculation of the confidence interval based on Equation (1) and Inequality (2) does not depend on the output data itself, but on the number of replications p, the confidence level $1 - \alpha$ and q itself. Because the confidence level can be considered as a given parameter the main question is: How to choose several q-quantiles as a function of p? The basic idea of the algorithms described in this section is to choose the 0.5-quantile as the starting point and to choose all other quantiles in a way that their confidence intervals do not overlap. A larger number of replications involves smaller confidence intervals and this enables the selection of more quantiles with non overlapping confidence intervals.

Our first method is based on Equation (1). In the beginning the first quantile 0.5 is given and its confidence interval is calculated by extending l and u until the wanted confidence level $1 - \alpha$ is reached. l and u describe the indexes in $\{y_{j,i}\}_{j=1}^{p}$ of the bounding values of the confidence interval. The selection of the next two quantiles which have a non overlapping and non disjoint confidence interval is not straight forward, because Equation (1) has no closed form. Therefore, we perform two binary searches in the directions above and below 0.5. The binary search in the direction below 0.5 stops if a quantile is found with a upper bound u' being equal to l. Analogously, the binary search in the upper direction stops if a quantile is found with a lower bound l' equal to u. The result of these binary searches are the next displayed quantiles. The binary searches are repeated, until it is not possible to find another quantile with a confidence interval in the unprocessed area between the last l and 1 (resp. u and p). This calculation can be performed before the simulation experiment starts and, therefore, the run time of this method does not really matter. For convenience a linear search, leading to a worse run time, could be performed instead of the binary search.

Equ. (1)		Inequ. (2)	
$p = 100$	$p = 1000$	$p = 100$	$p = 1000$
q (l;u)	q (l;u)	q ($q \pm \epsilon_q$)	q ($q \pm \epsilon_q$)
			.003 (0;.006)
	.010 (5; 16)		.012 (.006;.018)
	.024 (16; 32)		.026 (.018;.034)
	.042 (32; 53)		.045 (.034;.056)
.08 (3;13)	.066 (53; 79)	.09 (.04;.13)	.069 (.056;.082)
	.094 (79;110)		.098 (.082;.113)
	.127 (110;145)		.131 (.113;.148)
.19 (13;26)	.164 (145;184)	.20 (.13;.26)	.167 (.148;.187)
	.205 (184;226)		.208 (.187;.230)
	.250 (227;273)		.252 (.230;.274)
.34 (26;42)	.297 (273;321)	.34 (.26;.42)	.298 (.274;.322)
	.346 (321;371)		.346 (.322;.371)
	.396 (371;422)		.397 (.371;.422)
	.448 (422;474)		.448 (.422;.474)
.5 (42;59)	.5 (474;527)	.5 (.42;.58)	.5 (.474;.526)
	.553 (527;579)		.552 (.526;.578)
	.604 (579;630)		.603 (.578;.629)
	.655 (630;680)		.653 (.629;.678)
.67 (59;75)	.704 (680;728)	.66 (.58;.74)	.702 (.678;.726)
	.751 (729;774)		.748 (.726;.771)
	.795 (774;817)		.792 (.771;.813)
.81 (75;88)	.836 (817;856)	.80 (.74;.87)	.833 (.813;.852)
	.873 (856;891)		.869 (.852;.887)
	.906 (891;922)		.902 (.887;.918)
.93 (88;97)	.935 (922;948)	.91 (.87;.96)	.931 (.918;.944)
	.958 (948;969)		.955 (.944;.966)
	.977 (969;985)		.974 (.966;.982)
	.990 (985;996)		.988 (.982;.994)
			.997 (.994; 1)

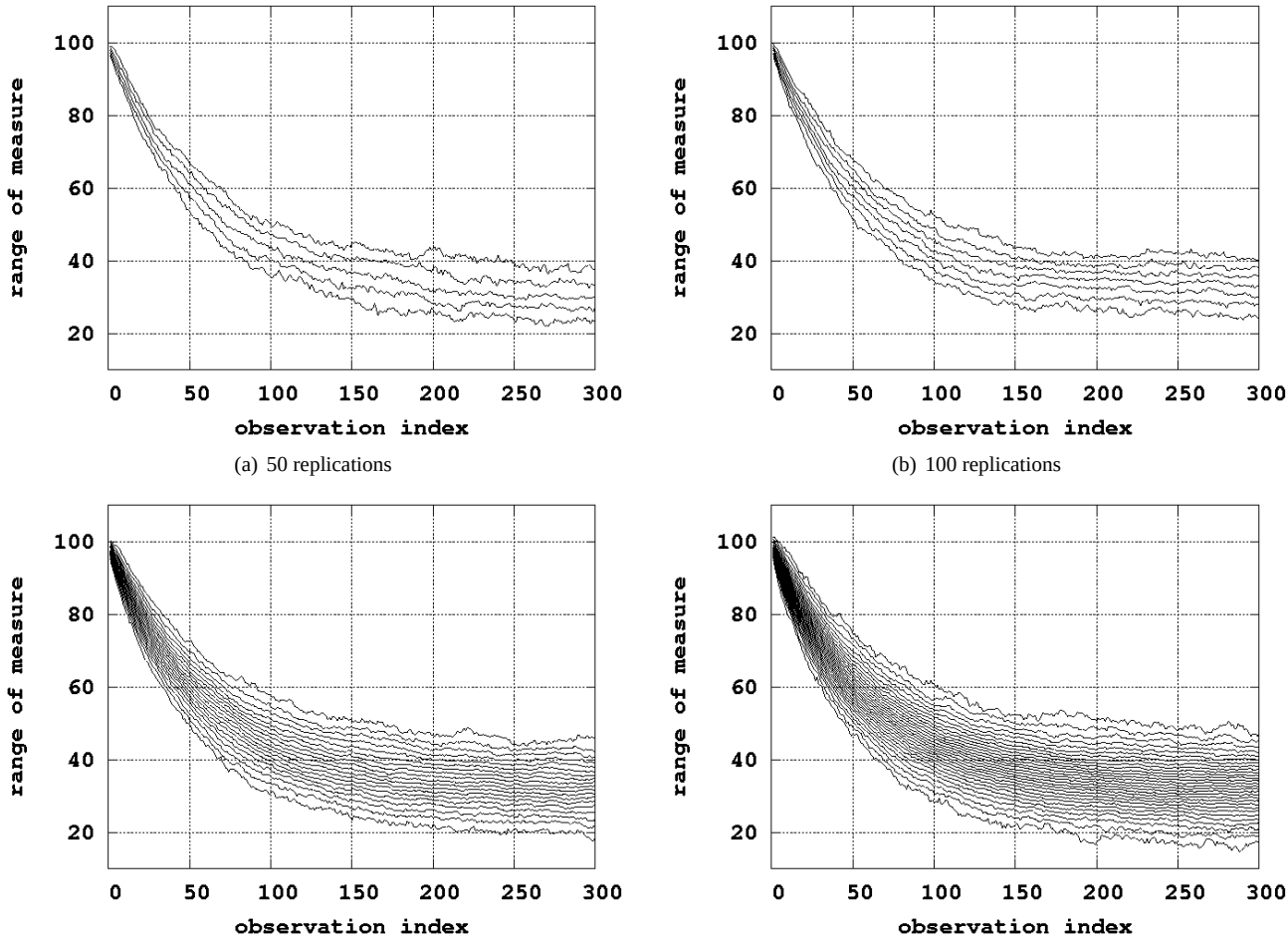
Table 1: This table shows the selected quantiles with their confidence intervals chosen by the method based on Equation (1) and by the method based on Inequality (2) with $1 - \alpha = 0.9$ and $p = 100$ resp. $p = 1000$.

The first two columns of Table 1 show the rounded result of this method for $p = 100$ and $p = 1000$ independent replications with a confidence level of $1 - \alpha = 0.9$. The method selects 7 quantiles for $p = 100$ and 27 quantiles for $p = 1000$. The values in brackets show the l and the u index of the quantile as defined in Equation (1).

The second investigated method is based on Inequality (2). Again, the starting point is the 0.5-quantile and the method searches for more quantiles in the directions below and above 0.5. In this case a binary search is not needed, because the next quantile can be calculated directly with the help of Inequality (2) and the following condition:

$$q_k - \epsilon_{q_k} = q_{k+1} + \epsilon_{q_{k+1}} \quad (3)$$

This condition is valid for the direction below 0.5, a condition for the direction above 0.5 can be formulated analogously. q_k is given and ϵ_{q_k} can be calculated by Inequality



(a) 50 replications

(b) 100 replications

(c) 500 replications

(d) 1000 replications

Figure 1: Several quantiles over time: ARMA process.

(2). Therefore, the substitution $a_k = q_k - \epsilon_{q_k}$ is calculable right away. Equation (3) can be transformed to:

$$a_k = q_{k+1} + z_{1-\alpha/2}\sqrt{\frac{q_{k+1}(1-q_{k+1})}{p}}$$

Eliminating the square root leads to

$$0 = q_{k+1}^2 b + q_{k+1} c_k + d_k$$

with $b = \frac{1}{z_{1-\alpha/2}^2} + \frac{1}{p}$, $c_k = -\frac{2a_k}{z_{1-\alpha/2}^2} - \frac{1}{p}$ and $d_k = \frac{a_k^2}{z_{1-\alpha/2}^2}$. Finally, q_{k+1} can be calculated by

$$q_{k+1} = \frac{-c_k - \sqrt{c_k^2 - 4bd_k}}{2b}. \qquad (4)$$

Equation (4) is valid for quantiles below 0.5. An equation for quantiles above 0.5 can be derived analogously. The search should be continued until the bounds of the interval $[0, 1]$ are exceeded.

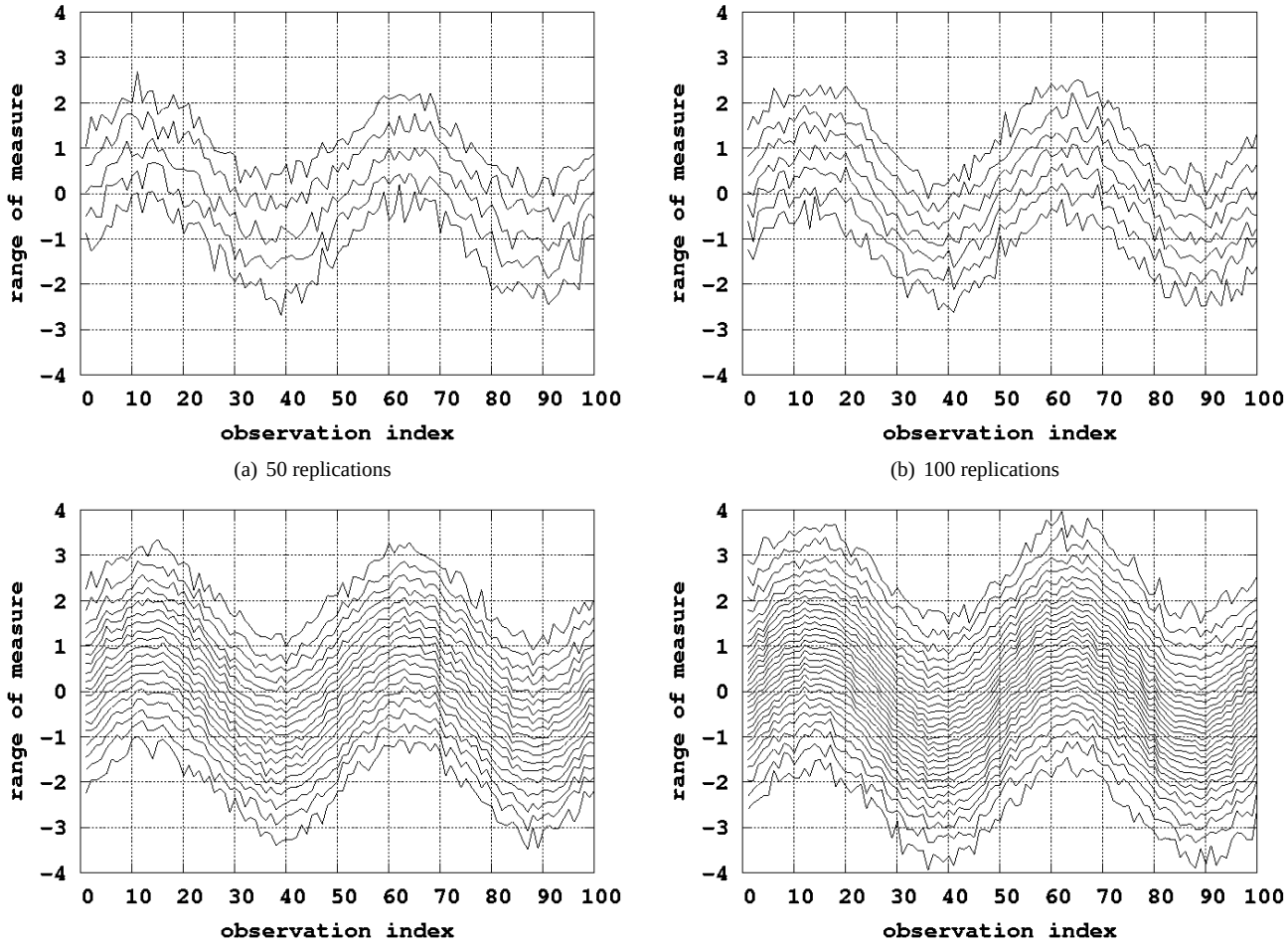
The rounded results of the second method are shown in the last two columns of Table 1. For $p = 100$ the second method selects 7 quantiles and for $p = 1000$ this method selects 29 quantiles. The values in brackets show the confidence interval of the belonging quantile in the range of the probability.

The results of the first and the second method are comparably accurate. However, the binary search of the first method seems to be circumstantial compared to the direct calculation by Equation (4) in the second method. Furthermore, the calculation of $\binom{p}{j}$ in Equation (1) involves the handling of very small and very big values. This might lead to problems in computer calculations and rounding errors. Therefore, we recommend the second of the described methods. All depictions in subsequent sections use quantiles selected by the second method.

EXAMPLES

In the previous section we described how to select several quantiles. In this section we use these selected quantiles to depict stochastic processes with well known statistic properties. This is useful to validate the results as in [Bause and Eickhoff, 2003]. The implementation of the stochastic processes is based on the random number generator described in [L'Ecuyer et al., 2002]. Because this generator allows the choice of many substreams, it is suitable for many independent replications. As already mentioned in the introduction, the independent replications of the stochastic processes are used to collect a set of independent data for each X_i with $1 \leq i \leq \infty$. To realize these stochastic processes the random numbers are transformed as follows.

ARMA Process: The behaviour of the first stochastic process is comparable with the behaviour of a storage in an inventory system. It is an ARMA$(5, 5)$ process which is



(a) 50 replications

(b) 100 replications

(c) 500 replications

(d) 1000 replications

Figure 2: Several quantiles over time: periodic process.

defined by

$$X_i = 1 + \epsilon_i + \sum_{k=1}^{5} \frac{1}{2^k}(X_{i-k} + \epsilon_{i-k}), k \geq 0$$

with the starting condition $X_{-5} = X_{-4} = X_{-3} = X_{-2} = X_{-1} = 100$. $\{\epsilon_i\}_{i=1}^{\infty}$ is an independent Gaussian white noise process ([Hamilton, 1994]). We selected four sets of quantiles for $p = \{50, 100, 500, 1000\}$ independent replications. The results are shown in Figure 1.

Periodic Process: The second observed process has a periodic behaviour and is defined by

$$X_i = a \cdot \sin(\omega i) + \epsilon_i$$

The cycle length of the sine oscillation is given by $T = \frac{2\pi}{\omega}$ with the amplitude a. And again $\{\epsilon_i\}_{i=1}^{\infty}$ is an independent Gaussian white noise process. This process is depicted in Figure 2 for $p = \{50, 100, 500, 1000\}$ independent replications.

Exponential Process: The behaviour of the third stochastic process is comparable with the behaviour of a buffer in a queueing system. It is defined by

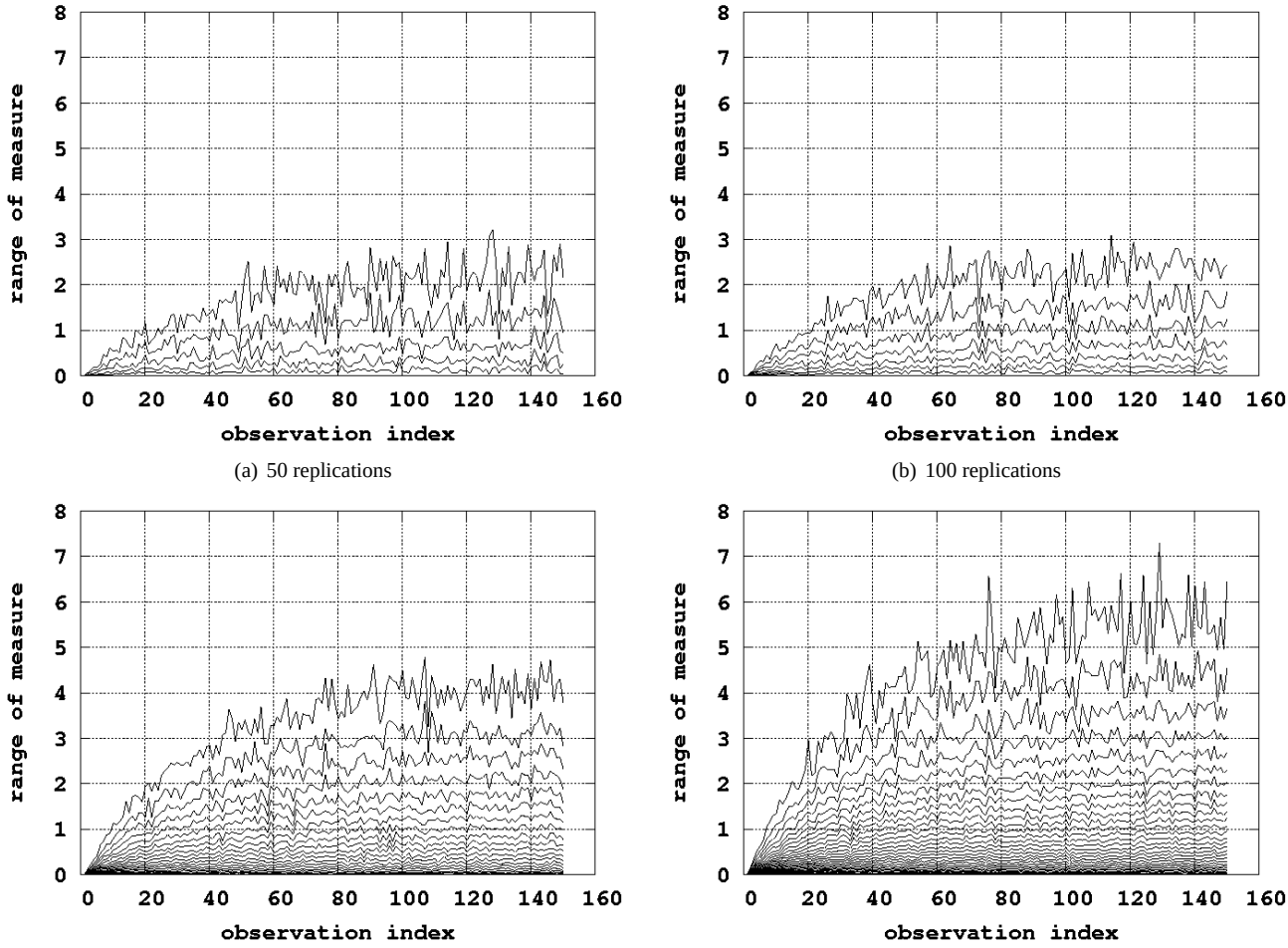
$$X_i = \epsilon_i' \cdot b(1 - e^{(i\frac{ln(0.05)}{l})}).$$

The process $\{\epsilon_i'\}_{i=1}^{\infty}$ is similar to the independent Gaussian white noise process, but its distribution is exponential (see [Law and Kelton, 2000]) with $\beta = 1$. The parameter b stretches the distribution. The part in brackets of the formula causes that the process is slowly converging towards its marginal distribution. This is depicted in Figure 3 for $p = \{50, 100, 500, 1000\}$.

Of course, a smaller value of p leads to a smaller set of selected quantiles. Additionally, the quantiles of the smaller set seem to fluctuate more. Further more, the quantiles of areas with lower probability fluctuate more than the ones of high probability. In Figure 1 and Figure 2 this can be observed when comparing the bounds 0 and 1 with the center (around 0.5) of the distribution. Because the distribution in Figure 3 is not symmetrical, the quantiles at bound 1 fluctuate more than the ones at bound 0.

These examples show, that this approach of depicting quantiles is suitable for both symmetrical and asymmetrical distributions, as well as converging and non converging processes. However, we recommend to use at least 50 independent replications. This makes sure that the selected set of quantiles includes at least 5 different quantiles.

Random Walk: The last examined process behaves like a random walk between 0 and 100. The random walks are not stopped at this thresholds, but all values higher (resp. lower) than these bounds are reduced to these values (compare [Bause and Beilner, 1999]). The peculiarity of this process is the constant value of the mean (e.g. the 0.5-quantile), whereas all other quantiles are not constant and tend to the thresholds 0 and 100 (see Figure 4(a) and 4(b)). Therefore, the cumulative distribution function is in the be-



(a) 50 replications

(b) 100 replications

(c) 500 replications

(d) 1000 replications

Figure 3: Several quantiles over time: exponential process.

ginning very steep around the 0.5-quantile (Figure 4(c)), but after a long simulation time it is very flat (Figure 4(d)). Analysis of only mean values would show a constant behaviour, even though this example is transient and the cumulative distribution is slowly converging to its marginal distribution.

CONCLUSIONS

We described two methods of selecting quantiles. Both methods delivered similar results. However, for further investigation we decided to use the method based on Inequality (2) because its complexity is lower.

This approach to depict several quantiles over time appears suitable for a variety of different transient behaviours. We recommend to use at least 50 independent replications to make sure, that the selected set of quantiles is reasonably large. In finite-horizon simulation the replications do not need to be processed in parallel. Therefore, a large number of replications, e.g. $p = 1000$, is feasible.

Our last investigated example shows, that the analysis of several quantiles provides a deeper insight into the analyzed process than mean value analysis. Drawing conclusions entirely based on mean value analysis is not recommendable for complex models.

REFERENCES

[Alexopoulos and Kim, 2002] Alexopoulos, C. and Kim, S. (2002). Output data analysis for simulations. *Proceedings of the 2002 Winter Simulation Conference*, pages 85–96.

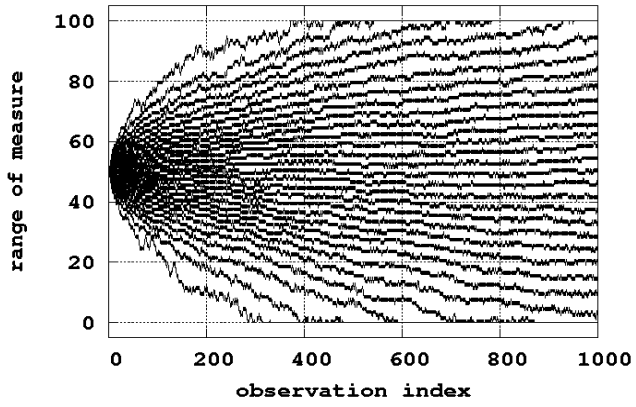
[Bause and Beilner, 1999] Bause, F. and Beilner, H. (1999). Intrinsic problems in simulation of logistic networks. *Proc. of the 11th European Simulation Symposium and Exhibition (ESS99)*, pages 193–198.

[Bause and Eickhoff, 2002] Bause, F. and Eickhoff, M. (2002). Initial transient period detection using parallel replications. *Proc. of the 14th European Simulation Symposium*, pages 85–92.

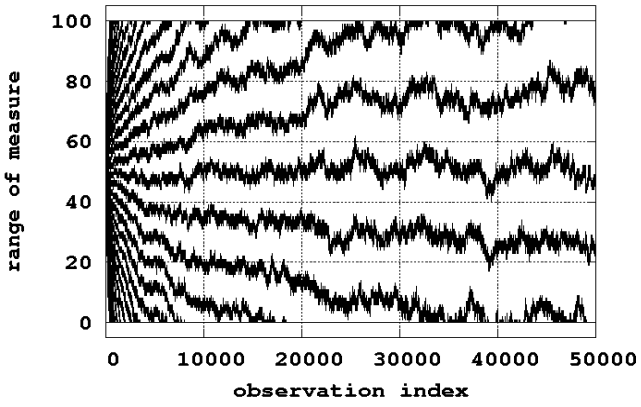
[Bause and Eickhoff, 2003] Bause, F. and Eickhoff, M. (2003). Truncation point estimation using multiple replications in parallel. *Proc. of the Winter Simulation Conference*, pages 414–421.

[Chen and Kelton, 1999] Chen, E. J. and Kelton, W. D. (1999). Simulation-based estimation of quantiles. *Proceedings of the 1999 Winter Simulation Conference*, pages 428–434.

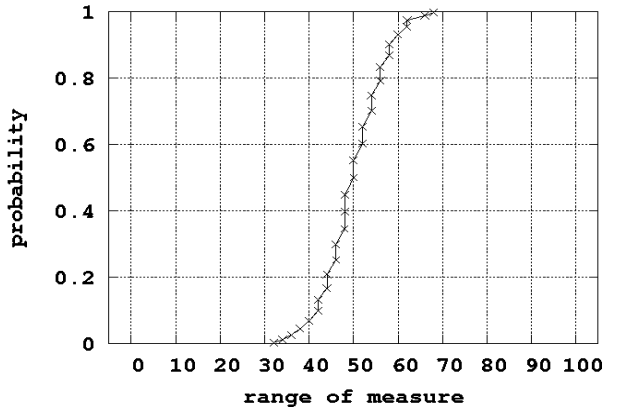
[Conover, 1999] Conover, W. (1999). *Practical Nonparametric Statistics*. John Wiley & Sons, Inc., New York.



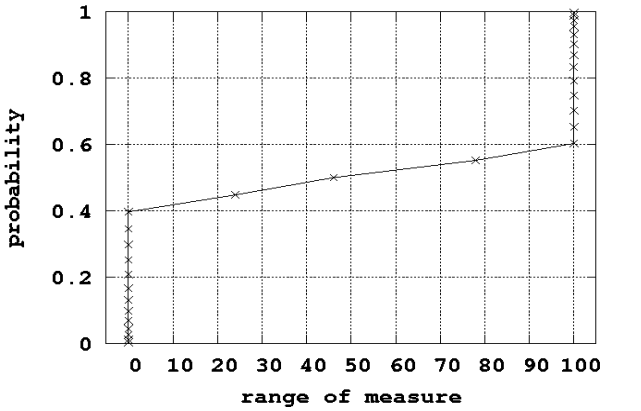
(a) $1000 \cdot 1000$ observations



(b) $1000 \cdot 50000$ observations



(c) empirical cumulative prob. distribution function at index 50



(d) empirical cumulative prob. distribution function at index 50000

Figure 4: Constant mean value, but other quantiles are changing over time.

[Hamilton, 1994] Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.

[Heidelberger and Lewis, 1984] Heidelberger, P. and Lewis, P. (1984). Quantile estimation in dependent sequences. *Operations Research*, 32(1):185–209.

[Jain and Chlamtac, 1985] Jain, R. and Chlamtac, I. (1985). The P^2 algorithm for dynamic calculations of quantiles and histograms without storing observations. *Communications of the ACM*, 28(10):1076–1085.

[Law and Kelton, 2000] Law, A. M. and Kelton, W. D. (2000). *Simulation Modeling and Analysis*. McGraw-Hill Higher Education, New York.

[L'Ecuyer et al., 2002] L'Ecuyer, P., Simard, R., Chen, E. J., and Kelton, W. D. (2002). An object-oriented random-number package with many long streams and substreams. *Operations Research*, 50(6):1073–1075.

[Lee et al., 1999] Lee, J.-S. R., McNickle, D., and Pawlikowski, K. (1999). Quantile estimation in sequential steady-state simulation. *Proceedings of the 13th European Simulation Multiconference*, pages 168–174.

[Raatikainen, 1987] Raatikainen, K. E. E. (1987). Simultaneous estimation of several percentiles. *SIMULATION*, 49(4):159–164.

AUTHOR BIOGRAPHIES

MIRKO EICKHOFF holds a Diploma degree in Computer Science from the University of Dortmund. His research interests are in the area of output analysis of discrete event simulation using multiple replications. Currently he is a Ph.D. Candidate of the Simulation Research Group of the University of Canterbury. For details see http://www.cosc.canterbury.ac.nz/research/PG/mei16. His e-mail address is m.eickhoff@cosc.canterbury.ac.nz.

DONALD C. MCNICKLE is an Associate Professor of Management Science in the Management Department at the University of Canterbury. His research interests include queueing theory, networks of queues and statistical aspects of stochastic simulation. He is a memeber of INFORMS and the Operational Research Society. His e-mail address is don.mcnickle@canterbury.ac.nz.

KRZYSZTOF PAWLIKOWSKI is a Professor in Computer Science at the University of Canterbury, in Christchurch, New Zealand. The author of over 130 research papers and four books; has given invited lectures at over 80 universities and research institutes in Asia, Australia, Europe and North America. Alexander-von-Humboldt Research Fellow (Germany) in 1983-84 and 1999. His research interests include performance modelling of telecommunication networks, discrete-event simulation and distributed processing. Senior Member of IEEE, member of ACM and SMSI. His e-mail address is krys.pawlikowski@canterbury.ac.nz.