A New Confidence Interval Method for the Estimation of Quantiles

Johann Christoph Strelen
Rheinische Friedrich–Wilhelms–Universität Bonn

Abstract

Confidence intervals for the median of estimators or other quantiles were proposed as a substitute for usual confidence intervals in terminating and steady-state simulation. This is adequate since for many estimators the median and the expectation are close together or coincide, particularly if the sample size is large. The novel confidence intervals are easy to obtain, the variance of the estimator is not used. They are well suited for correlated simulation output data, apply to functions of estimators, and in simulation they seem to be particularly accurate. For the estimation of quantiles by order statistics, the new confidence intervals are exact.

1 Introduction

In simulation, confidence intervals tend to be inaccurate, since assumptions concerning the sample and the estimator are not fulfilled literally, more precisely, the assumed confidence level is not the probability that the real parameter value lies within the calculated confidence interval.

We proposed in [4, 5] an alternative technique for confidence intervals which we call *median confidence intervals*, and *min-max confidence intervals* which are more general: Median confidence intervals (MCI) are a special case of min-max confidence intervals (MMCI), but they seem to be particularly useful. Min-max confidence intervals are suitable for the estimation of quantiles, and they have some potential for further development. These confidence intervals are obtained by means of a small number of replications. They have attractive features and some minor disadvantages compared to classical confidence intervals.

The variance of estimators is not needed for them, but this is usually a main difficulty when confidence intervals are constructed because "simulation output data are always correlated" [1]. Even the variance of an estimator may not exist, for example in the case of some heavy-tailed distributions [3]. Nevertheless, min-max confidence intervals can be constructed, whereas classical confidence intervals cannot.

It is easy to obtain median confidence intervals for functions of two or more estimators whereas it is difficult to get confidence intervals with other known methods, in general, except for jackknife intervals [2]; an example is given in [6, section 5].

In realistic models which involve dependent simulation output with unknown distribution, median confidence intervals seem to be more accurate, i.e. the coverages are closer to the predefined confidence level. This is shown in [6].

Some independent estimations of a measure of interest are used for a min-max or median confidence interval, say w, e.g. 5 or 6. The confidence level depends on w, only values like $1 - 0.5^w,\ w = 1, 2, \ldots$, or similar are possible – this becomes clear later in Theorem 1.

This paper deals with the estimation of quantiles with order statistics. Theorem 2 states that here min-max confidence intervals are exact when the sample is iid.

2 Min-max and Median Confidence Intervals

Now we explain the technique in detail. The random variables of the sample $X_{1,1}, ..., X_{1,n}$ may have the common distribution function $F_{X,\theta}(x)$ when they stem from a steady-state simulation run of length n, or from n independent terminating simulation runs. $\theta,\ \theta \in \Theta$, is a parameter, for example the mean or a quantile, and Θ a set of possible parameter values. Or the sample stems form a terminating simulation and has a common distribution with the parameter θ.

Let $T(X_{1,1}, ..., X_{1,n})$ denote an estimator for the parameter θ with the distribution function $F_\theta(x),\ \theta \in \Theta$.

We consider a novel kind of confidence interval

$$[T^{\min},\ T^{\max}) \tag{1}$$

where $T^{\min} = \min_{1 \le i \le w} T_i$ and $T^{\max} = \max_{1 \le i \le w} T_i$. Here, the $T_i = T(X_{i,1}, ..., X_{i,n}),\ i = 1, \ldots, w$, are estimators for w independent replications $X_{i,1}, ..., X_{i,n}$ of the sample $X_{1,1}, ..., X_{1,n}$. We call (1) a "min-max confidence interval".

For $F = F_\theta(\theta)$, the value of the estimator distribution function at θ, the unknown parameter value, the following theorem holds.

Theorem 1 *The interval (1) is a confidence interval for the parameter θ with the confidence level $CL = 1 - F^w - (1 - F)^w$, i.e.*

$$P\{T^{min} \le \theta < T^{max}\} = 1 - F^w - (1 - F)^w \tag{2}$$

holds.

The **proof** is very simple. The probability that T_i is less than or equal to θ is $P\{T_i \le \theta\} = F_\theta(\theta) = F$, the probability that $T^{\max}$ is less than or equal to θ is $P\{T^{\max} \le \theta\} = P\{\text{all } T_i \le \theta\} = F^w$ due to the independency. Similarly, $P\{T_i > \theta\} = 1 - F_\theta(\theta) = 1 - F$, $P\{T^{\min} > \theta\} = P\{\text{all } T_i > \theta\} = (1 - F)^w$. Hence $P\{T^{\min} > \theta \text{ or } T^{\max} \le \theta\} = F^w + (1 - F)^w$, and (2) follows. □

Remarks

1. The distribution function $F_\theta(x)$ of the estimator may not be known, only the value $F_\theta(\theta)$ is needed. We term this value the *confidence-level-determining (CLD) probability.*

2. The variance of the estimator is not needed, the question whether the random variables $X_{i,1}, \ldots, X_{i,n}$ are independent does not arise.

3. The confidence level cannot be chosen arbitrarily, only the values $CL = 1 - F^w - (1-F)^w,\ w = 2, 3, \ldots$ are allowed. If a specific confidence level $1-\alpha$ is required, the number w of replications is the smallest number for which $CL \geq 1-\alpha$ holds.

Given the number w of replications, the confidence level CL is obviously a function of the CLD probability F which has a single maximum:

Lemma 1 *The function* $1-F^w-(1-F)^w$ *has the maximum* $1-0.5^{w-1}$ *at* $F = 0.5$.

Now we consider the most important special case where $F_\theta(\theta) = 1/2$, i.e. the unknown parameter is the median of the estimator. Therefore we use the term "median confidence intervals". This is the case for all unbiased estimators with symmetric distributions, for example the normal distribution. Here holds

$$P\{T^{\min} \leq \theta < T^{\max}\} = 1 - 0.5^{w-1}. \tag{3}$$

Symmetry of the estimator distribution, the absence of skewness, is a sufficient condition for median conficence intervals to be exact. It is not necessary, there are unsymmetric distributions for which the mean and the median coincide.

If the median is merely close to the expectation of an unbiased estimator, only $F \approx 0.5$ holds, and the median confidence interval is only approximate. The error of the conficence level is the difference between (2) and (3). This happens quite often, due to the central limit theorem, when the summed random variables are not normally distributed, but n, the number of summands, is large. Then the distribution function of the estimator is approximately a normal distribution, hence approximately symmetric, and the median is near to the expectation.

A min-max confidence interval is exact if the w replications are independent, even the estimator may be biased. This sounds very interesting, but the serious problem is the CLD probability $F = F_\theta(\theta)$. We do not know how to calculate this F in general. But there is an interesting application where F can be calculated: Order statistics as estimates for quantiles. Consider samples $X_1, \ldots, X_n$ and the according ordered sequence $X_{(1)}, \ldots, X_{(n)},\ X_{(k)} \leq X_{(j)}$ if $k < j$, where the X_k are iid. with the strictly increasing distribution function $F(x)$. The q-quantile $\theta = x_q,\ q \in (0,1),\ F(x_q) = q$, is estimated by $X_{(r)},\ r \in \{1, 2, \ldots, n\}$. Let $F_\theta(x)$ denote the distribution function of the estimator, namely $X_{(r)}$. Here, $F = F_\theta(x)$ is known:

Theorem 2 *If the q-quantile x_q is estimated by $X_{(r)}$, the min-max confidence interval (1) has precisely the confidence level (2) with*

$$F = \sum_{i=r}^{n} \binom{n}{i} q^i (1-q)^{n-i}. \tag{4}$$

Proof For any k, $0 < k < n$, the probability $P\{X_{(k)} \le x < X_{(k+1)}\}$ equals the probability that k of the random variables X_i of the sample are less or equal x, hence $P\{X_{(k)} \le x < X_{(k+1)}\} = \binom{n}{k} F^k(x)[1-F(x)]^{n-k}$, $k = 1, \ldots, n-1$, and $P\{X_{(n)} \le x\} = F^n(x)$ hold. Using this we conclude $F_\theta(x) = P\{X_{(r)} \le x\} = P\{X_{(r)} \le x < X_{(r+1)}\} + P\{X_{(r+1)} \le x < X_{(r+2)}\} + \ldots + P\{X_{(n)} \le x\} = \sum_{i=r}^{n} \binom{n}{i} F^i(x)[1-F(x)]^{n-i}$. With $F(x_q) = q$, (4) follows. □

Remarks

1. Here the CLD probability $F = F_\theta(x_q)$ is independent of the actual distribution function of the sample elements X_i.

2. Theorem 2 is not useful for the simulation of the extremes, $q = 0$ or $q = 1$. Here we get the confidence level 0.

3. Usually, $r \approx qn$ is chosen.

It is common to take $r = \lceil nq \rceil$. Here $\lceil y \rceil$, y real, denotes the largest integer i, $i \le y$, hence $i = y + \delta$ where $0 \le \delta < 1$ holds. But for our min-max confidence intervals, we take r slightly different, see section 3. It is interesting to note that unbiasedness of the estimator is not so important here, a bias only decreases the confidence level.

For $n > 9$, the CLD probability F can be approximated with the standard normal distribution function Φ, $F \approx \Phi\left(\frac{n-nq+0.5}{\sqrt{nq(1-q)}}\right) - \Phi\left(\frac{r-nq-0.5}{\sqrt{nq(1-q)}}\right)$.

The confidence level can be kept near to the maximum which is assumed for $F = 0.5$ (Lemma 1), at least when the sample size n is not too small:

Lemma 2 *For a fixed probability q, $0 < q < 1$, and the estimator $X_{(r)}$, $r = \lceil qn \rceil$, of the q-quantile, $\lim_{n\to\infty} F = 0.5$ holds.*

Proof $\lim_{n\to\infty} F = \lim_{n\to\infty} \Phi\left(\frac{n-nq}{\sqrt{nq(1-q)}}\right) - \lim_{n\to\infty} \Phi\left(\frac{r-nq}{\sqrt{nq(1-q)}}\right)$ holds according to the De Moivre-Laplace limit theorem. This is 0.5 since $r = nq + \delta$ where $0 \le \delta < 1$. □

Also for smaller sample sizes n, the confidence level can be kept near to the maximum $1 - 0.5^{w-1}$. This can be seen in section 3 and in the following corollary.

Corollary 1 *If the sample size n is odd, $r = \lceil n/2 \rceil$ and $q = 0.5$, i.e. the median is estimated, $F = 0.5$ holds.*

Proof Here $F = \sum_{i=\lceil n/2 \rceil}^{n} \binom{n}{i} 0.5^i (1-0.5)^{n-i} = 0.5^n \sum_{i=\lceil n/2 \rceil}^{n} \binom{n}{i}$. Due to $\binom{n}{i} = \binom{n}{n-i}$, we have also $F = 0.5^n \sum_{i=\lceil n/2 \rceil}^{n} \binom{n}{n-i} = 0.5^n \sum_{j=0}^{n-\lceil n/2 \rceil} \binom{n}{j} = 0.5^n \sum_{j=0}^{\lfloor n/2 \rfloor} \binom{n}{j}$. We add these two equations: $2F = 0.5^n \sum_{j=0}^{n} \binom{n}{j} = 0.5^n 2^n = 1$. $F = 0.5$ follows. □

The CLD probability F is also known for some toy simulations, see [4]. Moreover, it can be estimated in a very long and expensive simulation, see [6, section 6].

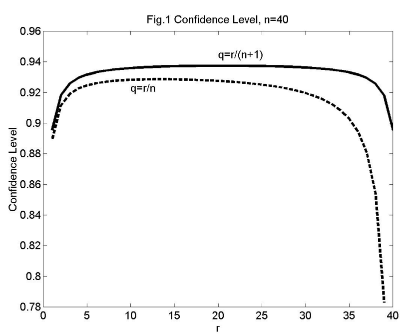
3 Numerical Studies on the Confidence Level

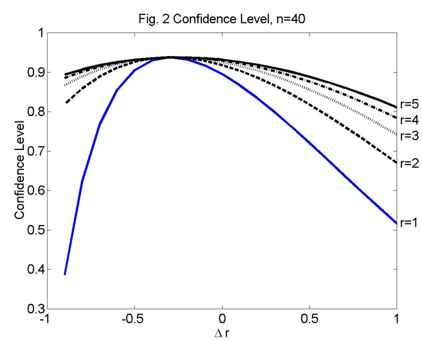
We present numerical values for the confidence level according to Theorem 2. The main conclusions from this are as follows:

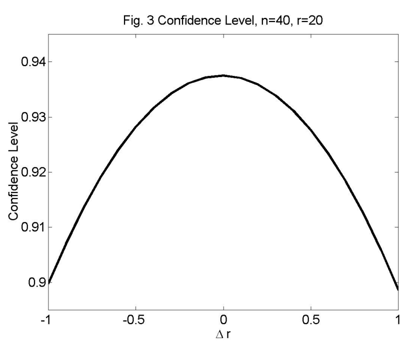
1. The conficence level can be kept near to the maximum $1 - 0.5^{w-1}$.
2. The common estimator should be modified.

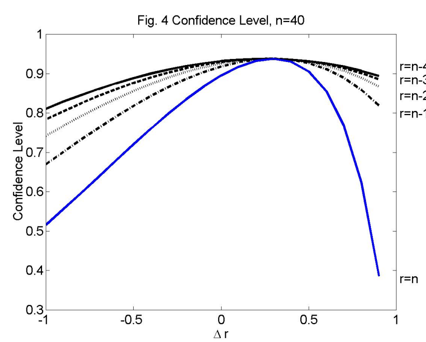
First, we consider $r = \lceil qn \rceil$ versus $r = \lceil q(n+1) \rceil$. In Figure 1, for $q = r/n$, $r = 1, \ldots, n-1$, and for $q = r/(n+1)$, $r = 1, \ldots, n$, the confidence level is presented. Obviously, one should prefer $q = r/(n+1)$, hence the estimator $X_{(r)}$ with $r \approx (n+1)q$, since for all values r the confidence level is near to the possible maximum; in the other case with the common estimator this is not true.

Now we consider the case where nq and $(n+1)q$ are not integer. In the following figures we draw the confidence level for $q = (r + \Delta r)/(n+1)$ where $\Delta r = -0.9, \ldots, 1.0$, $r = 1, \ldots, 5$ (small q, Figure 2), $\Delta r = -0.9, \ldots, 0.9$, $r = \lfloor n/2 \rfloor$ (medium q, Figure 3), and $\Delta r = -0.9, \ldots, 0.9$, $r = n-4, \ldots, n$ (large q, Figure 4).









We first conclude from the figures that $r = \lfloor q(n+1) \rfloor$ is reasonable for medium and large q, but for small q, $r = \lceil q(n+1) \rceil$ is better. Secondly, r should be at least 4 or 5, and at most $n-4$ or $n-3$. This can be accomplished with sufficiently large sample sizes n. A closer look indicates the following. For small values $q \approx 4/n$, $r \approx \lfloor (n+1)q + 0.75 \rfloor$ is good, for medium values $q \approx 0.5$, $r \approx \lfloor (n+1)q + 0.4 \rfloor$ is good, for large values $q \approx 1 - 4/n$, $r \approx \lfloor (n+1)q + 0.1 \rfloor$ is good. Accordingly, we tried the estimator $X_{(r)}$, $r = \lfloor (n+1)q + 0.1 + 0.65(1 - q - 4/n)/(1 - 8/n) \rfloor$ where $n \geq 4/q$ and $n \geq 4/(1-q)$.

4 Numerical Experience with the Accuracy

We performed simulation studies where we looked at the accuracy of min-max confidence intervals for order statistics, and compared it with two other techniques. As measure for the accuracy we considered the difference betwee the predefined confidence level CL and the observed coverage. The coverage C is the fraction of a number of simulations where an estimated parameter whose exact value must be known in advance here lies within the confidence interval.

We applied the following models: For independent samples Pareto distributed random variables with the cumulative disribution function (CDF) $F_P(x) = 1 - x^{-a},\ 1 < a \le 2,\ x \ge 1$, or random variables $Z_i,\ i = 1, \ldots, n$, with a uniform distribution and some probability at $x = 0$ with the CDF $F_U(x) = P_0 + x(1 - P_0),\ 0 \le P_0 < 1,\ 0 \le x \le 1$.

For dependent samples we considered the stochastic process $Y_i = Y_{i-1}$ if $U_i \le corr$, $Y_i = Z_i$ otherwise, $i = 1, \ldots, n$, where Y_0 has the CDF $F_U(x)$, U_i has a 0-1-uniform distribution, and $corr$, $0 \le corr < 1$, determines the correlation.

An other stochastic process are the delays in an M/M/1 queueing system with load $\rho,\ 0 < \rho < 1$.

Welch [7] proposes a confidence interval technique (WelchCI) for q-quantiles which are taken from an ordered iid. sample $X_{(1)}, \ldots, X_{(N)},\ P\{X_{(l)} \le x_q \le X_{(h)}\} = \sum_{i=l}^{h-1} \binom{N}{i} q^i (1-q)^{N-i} \ge 1 - \alpha$ where l and h are chosen symmetrically about qN until the sum first exceeded $1 - \alpha$. Usually, if $N > 9$, the approximations $l = \lfloor qN + 1/2 + \Phi^{-1}(\alpha/2)\sqrt{Nq(1-q)} \rfloor$ and $h = \lceil qN + 1/2 + \Phi^{-1}(1 - \alpha/2)\sqrt{Nq(1-q)} \rceil$ are applied.

For dependent samples, Welch proposes to consider w replicated ordered samples $X_{i,(1)}, \ldots, X_{i,(n)},\ i = 1, \ldots, n$. The q-quantile is estimated with the mean of the idependent $X_{i,(r)},\ r = \lceil qn \rceil$, and a confidence interval is calculated as usual with the Student distribution (MVCI). Min-max confidence intervals are $\left[\min_i X_{i,(r)},\ \max_i X_{i,(r)}\right)$ with the confidence level CL according to Theorem 2. For our comparative studies we adopted this confidence level CL and the same overall sample size for all three techniques, i.e. $N = wn$.

Broadly speaking, the result is as follows. Welch confidence intervals are very accurate for iid. samples, only for small samples, min-max confidence intervals are often slightly more accurate. But the independence assumption is essential. For small sample sizes n, quite large or quite small values q, skewed distributions of the sample random variables, or high correlation of them, min-max confidence intervals are more accurate than mean value confidence intervals; for large sample sizes, both seem to be similarly accurate. In the M/M/1 model, mean value confidence intervals are quite inaccurate, min-max confidence intervals are better but also not really good.

These differences can be seen in the following examples.

In each example, a model with some specific parameter values was simulated 100000 (Tables 1,2,3) or 40000 (Table 4) times, and the error $CL - C$, namely the mean difference between the confidence level and the observed coverage, the mean width of the confidence intervals, and the empirical variance of this was calculated. For the errors $CL - C$, 90% confidence intervals $\pm\epsilon$ were calculated. The reader may note that $CL - C = 0.08$ could mean: The predefined confidence level is 90% but the coverage is only 82%, the confidence intervals are too optimistic. And $CL - C = -0.05$ is not accuate as well, but pessimistic, the confidence intereval is too wide but safe.

In Table 1, independent Pareto distributed random variables are the model, and the half width ϵ of the confidence intervals for the errors $CL-C$ is smaller than 0.002. In Table 2, independent $F_U(x)$ distributed random variables are the model, and ϵ is smaller than 0.002, in Table 3, dependent $F_U(x)$ distributed random variables are the model, and ϵ is smaller than 0.0021, and in Table 4, the delays in an M/M/1 queueing system are the model, and ϵ is smaller than 0.004.

Table 1, Independent Pareto: Errors $CL - C$, avg. Half Length, var. Half Length

a	q	n	MMCI			MVCI			WelchCI		
2	0.9	40	0.001	0.813	0.163	0.064	0.617	0.094	-0.030	0.665	0.051
2	0.05	100	0.000	0.013	0.000	0.017	0.012	0.000	-0.023	0.010	0.000
1.1	0.5	20	0.000	0.438	0.047	0.029	0.369	0.033	-0.025	0.342	0.010
1.1	0.01	401	0.002	0.005	0.000	0.040	0.005	0.000	-0.019	0.003	0.000

Table 2, Independent $F_U(x)$: Errors $CL - C$, avg. Half Length, var. Half Length

P_0	q	n	MMCI			MVCI			WelchCI		
0	0.5	10	0.001	0.168	0.003	0.062	0.129	0.002	-0.037	0.128	0.001
0.7	0.8	20	0.001	0.252	0.006	0.089	0.199	0.004	-0.003	0.204	0.002
0.7	0.9	40	0.001	0.193	0.006	0.064	0.146	0.003	-0.030	0.124	0.001

Table 3, Dependent $F_U(x)$: Errors $CL - C$, avg. Half Length, var. Half Length

P_0	$corr$	q	n	MMCI			MVCI		
0	0.8	0.5	10	-0.036	0.313	0.008	0.015	0.245	0.005
0.4	0.8	0.5	10	-0.036	0.326	0.015	0.028	0.272	0.010
0.8	0.5	0.9	40	-0.001	0.342	0.008	0.095	0.271	0.006

Table 4, M/M/1 Delays: Errors $CL - C$, avg. Half Length, var. Half Length

ρ	q	n	MMCI			MVCI		
0.5	0.7	20	0.040	0.687	0.239	0.094	0.587	0.177
0.5	0.99	400	0.024	1.227	0.366	0.160	0.937	0.224
0.2	0.8	20	-0.086	0.080	0.007	-0.054	0.070	0.006
0.2	0.95	80	-0.032	0.212	0.012	0.077	0.148	0.006
0.9	0.7	20	0.082	0.997	0.359	0.153	0.846	0.263
0.9	0.95	1000	0.184	1.153	0.357	0.241	1.068	0.313

References

[1] A. M. Law and W. D. Kelton. *Simulation Modeling and Analysis.* McGraw-Hill, New York, third edition, 2000.

[2] R.G. Miller. The jackknife - a review. *Biometrika*, 61:1–15, 1974.

[3] K. Sigman. A primer on heavy-tailed distributions. *Queueing Systems*, 33:261–275, 1999.

[4] J. Ch. Strelen. Median confidence intervals. In E.J.H. Kerkhoffs and M. Snorek, editors, *Modelling and Simulation 2001 - Proceedings of the ESM 2001*, pages 771–775. Society for Computer Simulation, 2001.

[5] J. Ch. Strelen. Median confidence intervals - grouping data into batches and comparison with other techniques. In M. Ades and L.M. Deschaine, editors, *Proceedings of the Business and Industry Symposium - ASTC 2002*, pages 169–175. Society for Computer Simulation, San Diego, 2002.

[6] J. Ch. Strelen. The accuracy of a new confidence interval method. In R.G. Ingalls, M.D. Rossetti, J.S. Smith, and B.A. Peters, editors, *Proceedings of the 2004 Winter Simulation Conference*, pages 654–662. Omnipress, Washington DC., 2004. www.informs-sim.org/wsc04papers/079.pdf.

[7] P.D. Welch. The statistical analysis of simulation results. In S.S. Lavenberg, editor, *Computer Performance Modeling Handbook*, pages 267–329. Academic Press, New York, 1983.