CONFIDENCE LIMITS FOR THE EXPECTED TELEPHONE TRAFFIC IN SIMULATION MODELS USING ARMA-MODELS

K. STORDAHL
Norwegian Telecommunications Administration, Research Establishment Kjeller, Norway

B. H. SOLLIE, E. DAMSLETH
Norwegian Computing Center, Blindern, Oslo, Norway

1 ABSTRACT

Frequently it is necessary to build simulation models of telephone exchanges to analyse the telephone traffic. The object of the simulation is to produce estimates of the traffic within a given confidence interval. However, due to the dependence of the observations, one problem has been to calculate the length of the simulation period necessary to obtain a given precision.

One method of producing estimates of the confidence limits and thereby the necessary simulation time is to remove parts of the time series formed by the simulation in such a manner that the small time series (of equal length) resulted are assumed to be independent. Hence, it is possible to construct a confidence interval for the expected number of outgoing junctors occupied, based on the means of the independent time series. This method is, however, not satisfactory because some of the information in the time series is not taken into account when the confidence limits are estimated.

This paper introduces another method to construct confidence limits for the expected traffic in simulation models. By using the Box-Jenkins method [1], the output series from the simulations are identified as ARMA-models.

ARMA-models are a class of models containing the auto-regressive (AR) models and the moving average (MA) models. An important characteristic of the class is the principle of parsimony, i.e. the models are expressed by the smallest number of parameters for adequate representation.

In the stationary state of the time series, the process is identified obtaining the number of AR-parameters and the number of MA-parameters. Then the parameters, including the mean, are estimated using least square methods and the confidence interval for the mean is computed. The estimation utilizes all the observations in the time series.

The result of the modelling shows that the number of outgoing junctors occupied can be represented as auto-regressive processes of the first order, while the number of registers occupied have a more complicated structure.

Finally the traditional method of estimating confidence limits is compared with the new method using ARMA-models showing that substantial gain may be achieved.

2 INTRODUCTION

2.1 General

During the last years the Norwegian Telecommunications Administration has been strongly involved in traffic measurements and traffic simulations. The data technology and modern measurement equipment generates comprehensive and accurate traffic measurements. By using simulation models different traffic loads are studied in telephone network and exchanges.

2.2 Traffic simulation

Suppose a simulation of a telephone exchange is carried out. The number of occupied junctors in a junctor route as a function of time may be as shown in Figure 2.1.

After starting the simulation, the system needs some time to reach a stationary state. The time elapsed while the system is in a non stationary state is called the transient period.

The transient period has to be eliminated before estimation of the expected traffic. Determination of the length of the transient period is in general a difficult problem. It is often convenient to use the length of three mean calls as an approximation of the transient period. In this analysis, however, we have chosen to appoint the transient period after inspection of the behaviour of the traffic.

Observations of a process at equidistant points in time gives a sequence of observations which is called a time series. By analyzing the time series it is possible to construct a model for the time series. Then the expected value and other interesting parameters for the process can be estimated.

The objective is to estimate the expected traffic with the highest precision possible and at the same time to find an unbiased estimate of the standard deviation of the estimate.

One way to achieve an unbiased estimate of the deviation is to make a lot of simulations of the system each giving an estimate of the expected traffic. Then, since the simulations are independent, an estimate of the deviation is easily constructed. This method implies that a lot of transient periods have to be eliminated.

Simulations of large systems are often time consuming and expensive. Methods are therefore developed which are not based on repeated simulations.

A long simulation may be divided into separate simulation periods in such a way that the process is supposed to be independent in the different periods. This is done either by making large gaps between the simulation periods or by making the simulation periods so long, that their dependence is neglectable.

Then confidence limits of the expected traffic based on these simulation periods can be constructed using standard methods.

An alternative method is to identify the traffic process as an ARMA-model and then construct the confidence limits within this model.

Figure 2.1 Number of occupied junctors in a route as a function of time in a simulation experiment.

STORDAHL / SOLLIE / DAMSLETH-1
3 ARIMA-MODELS

3.1 Box-Jenkins theory

The ARIMA-(Autoregressive Integrated Moving Average) model is a class of models which describes the covariance structure of a time series. The class of models consists of the classical autoregressive and moving average models. The first unified approach to identification, estimation and forecasting ARIMA-models was given by Box and Jenkins (1970), and the ARIMA-models have since then become widely used.

3.2 Model representation

3.2.1 Autoregressive model, AR(p).

Let \( \{X_t\} \) be a time series observed at equidistant times. Suppose \( \{X_t\} \) is an autoregressive time series with mean \( \mu \).

Then the model may be written

\[
X_t - \mu = \varphi_1 (X_{t-1} - \mu) + \varphi_2 (X_{t-2} - \mu) + \cdots + \varphi_p (X_{t-p} - \mu) + \sigma_t
\]

where \( \{\sigma_t\} \) is a white noise process with \( \text{E} \sigma_t = 0 \), \( \text{Var} \sigma_t = \sigma^2 \).

By introducing the backward operator \( B \) and \( \varphi_p (B) \) where

\[
R_X = X_{t-1}
\]

\[
\varphi_p (B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_p B^p
\]

an equivalent form for the autoregressive model may be written

\[
\varphi_p (B) (X_t - \mu) = \sigma_t
\]

3.2.2 Moving average model, MA(q).

If the time series \( \{X_t\} \) is expressed as output from a finite linear filter, whose input is white noise \( \sigma_t \) as input, then the time series is a moving average process given by

\[
X_t - \mu = \theta_q (B) \sigma_t
\]

where

\[
\theta_q (B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q
\]

3.2.3 Mixed autoregressive - moving average process ARMA(p,q).

Every linear process might be expressed either as an autoregressive or as a moving average process. However, it will often be necessary to use a lot of parameters to obtain a satisfactory model representation.

The principle of parsimony is central in the Box-Jenkins approach. It is important, in practice, to employ the smallest number of parameters possible for adequate representation. This is carried out by combining the autoregressive and the moving average models. The ARMA \((p,q)\) model may be written

\[
\varphi_p (B) (X_t - \mu) = \theta_q (B) \sigma_t
\]

\( \{X_t\} \) is a stationary process if all the roots of \( \varphi_p (B) = 0 \) have moduls greater than 1. If all roots in \( \theta_q (B) = 0 \) have moduls greater than 1 the process is said to be invertible.

3.2.4 ARIMA(p,d,q) model.

Suppose that the time series is not a stationary process. Then it is necessary to transform the process into a stationary one. Usually non-stationarity of the mean can be removed by differentiating the time series a sufficient number of times using the difference operator defined by

\[
(1-B)X_t = X_t - X_{t-1}
\]

By introducing an additional parameter, \( \theta_0 \), in the model, it is possible to describe deterministic polynomial trends in the time series. The general model equation is then

\[
\varphi_p (B) (1-B)^d X_t = \theta_0 + \theta_q (B) \sigma_t
\]

3.3 Identification

The autocorrelation function and the partial autocorrelation function is used to identify the ARIMA \((p,d,q)\) model.

The autocorrelation function consists of the autocorrelations at lags \(0,1,2,\ldots\) where the autocorrelation at lag \(k\) is defined as the correlation of observations \(k\) units apart.

The partial autocorrelation function consists of the partial autocorrelations at lags \(0,1,2,\ldots\) where the partial autocorrelation at lag \(k\) is defined as the correlation of observations \(k\) units apart given the observations in between.

A non-stationary time series is characterized by a very slowly decreasing autocorrelation function. After differentiating the time series a sufficient number of times, the autocorrelation function will die out rather quickly. Then the number of differenciations, \(d\), is determined.

The structure of the estimated autocorrelation function and partial autocorrelation function gives proposals for number of autoregressive and number of moving average parameters. Hence a proposal for the orders \(p\), \(d\) and \(q\) of the model is given. Due to the uncertainty of the estimation of the two functions, several ARIMA models may some times be identified for the time series.

3.4 Estimation

After identification of the model the next step is to estimate the unknown parameters. A non-linear least square method is used in the estimation. The method estimates all the parameters simultaneously by minimizing the sum of squares of the calculated residuals \(\{\hat{\sigma}_t\}\).

3.5 Model diagnostic checking and improvements of the model

After identification and estimating of the model, an estimate of the white noise process, \(\{\hat{\sigma}_t\}\), is obtained. By assumption this process shall be white. If analysis shows that the white noise process is dependent, then this implies that a wrong model is fitted.

The analysis, however, gives proposals for improving the model. This improved model can then be fitted. The procedure is repeated until an adequate model is found.

A useful overall measure of the dependence in the estimated residuals is the statistic \(Q_{m-s}\) defined by

\[
Q_{m-s} = m \sum_{i=1}^{s} r_i^2
\]

where \(r_i\) is the estimated autocorrelation of the residuals at lag \(i\), \(m\) the number of estimated autocorrelations, \(n\) the number of observations and \(s\) the number of estimated parameters. If \(\{\sigma_t\}\) is normally distributed and the noise is really white, then \(Q_{m-s}\) is approximately Chi square distributed with \(m-s\) degrees. Therefore the statistic \(Q_{m-s}\) can be used to test the white noise hypothesis.

4 TELETRAFFIC SIMULATIONS

4.1 Data

In Norway an advanced telephone exchange simulation model, TETRASIM [2], has been constructed. The simulation
analyses in this paper are generated by the TETRASIM-model. There are two simulations of PABX exchanges, AKD 791 and GROSSCITOMAT, and three simulations by different load structure in a 88 Cross-Bar exchange.

In all simulations the number of outgoing junctors occupied and the number of registers occupied are analysed. The total simulation periods are mainly 1750 sec. The approximately first 250 sec are removed to assure that the processes are in a stationary state. The simulation output series are transformed to discrete time series which contain observations made at equidistant time intervals. The different time series are analysed with a time resolution of 5 sec, 10 sec and 15 sec.

4.2 The ARMA-models fitting ability

Earlier work, [3], [4], has shown that the ARMA-models can be used to describe both real traffic processes and simulated traffic processes.

We will now show how well the ARMA-models are able to describe the simulated process.

In the AKD 91-exchange we have been considering three routes and four registergroups. The registergroups have been merged and treated as one group because of low traffic.

For the junctors time series measuring the numbers of occupied outgoing junctors every 15 second have been studied for several routes. Inspection of the autocorrelation functions show for every route a nearly exponential decay, which is typical for an AR(1)-model. For route 1 the autocorrelation function tends to die out slower indicating that the process may be non-stationary. But the estimation of the autoregressive parameters shown in table 4.1, give parameter estimates significant less than 1. The Q-statistic indicates that the model fit is satisfactory.

Examination of the residuals and their autocorrelation functions give no limits to possible improvements of the model.

Table 4.1 Estimation results from the routes in AKD 91. Model AR(1): \( \varphi(X_{t-1}) + \mu \) = \( \psi(X_{t-1}) + a_t \)

<table>
<thead>
<tr>
<th>Routes</th>
<th>Capacity</th>
<th>Time resolution</th>
<th>Number of observations</th>
<th>( \psi )</th>
<th>( \mu )</th>
<th>( a_x )</th>
<th>( Q_1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Route 1</td>
<td>30</td>
<td>5 sec</td>
<td>300</td>
<td>.945</td>
<td>.044</td>
<td>1.1</td>
<td>35.9</td>
</tr>
<tr>
<td>Route 2</td>
<td>30</td>
<td>5 sec</td>
<td>300</td>
<td>.950</td>
<td>.045</td>
<td>1.1</td>
<td>31.9</td>
</tr>
<tr>
<td>Route 3</td>
<td>30</td>
<td>5 sec</td>
<td>300</td>
<td>.639</td>
<td>1.449</td>
<td>.045</td>
<td>33.6</td>
</tr>
</tbody>
</table>

By observing the number of occupied registers every second we get a time series with 1500 observations over the stationary period. Inspection of the autocorrelation function and the partial autocorrelation function suggest an ARIMA(1,2)-model. The result is given in table 4.2.

Table 4.2 Estimation results for the register group in AKD 91 with 1500 observations every second Model: ARIMA(1,2,2)

<table>
<thead>
<tr>
<th>Register group</th>
<th>Variance</th>
<th>( \varphi )</th>
<th>( \mu )</th>
<th>( \sigma )</th>
<th>( Q_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>.593</td>
<td>.909</td>
<td>.189</td>
<td>.087</td>
<td>2.0</td>
<td>41.2</td>
</tr>
</tbody>
</table>

In the GROSSCITOMAT-exchange nine routes are simulated. For the ASKIM 9a-exchange we have three different simulations with different traffic conditions for the exchange. Fourteen routes have been studied.

The results from the model fitting, estimation and model-checking have for all routes in the different exchanges and different simulations been unique: The best fit to the time series have been the ARIMA(1)-model. It is not surprising that such simulated queue-processes are well described by ARIMA(1)-models. We give a more detailed treatment of this question in the next section.

4.3 A queueprocess as an autoregressive process

Many queueprocesses can be described by the model M/M/N, i.e. a queuesystem with N servers, exponential distributed interarrivaltimes with expectation 1/\( \lambda \) and exponential distributed servicetimes with expectation 1/\( \mu \). With a large number of servers the queueprocess can be approximated by a Poisson process where the expected number of occupied servers is \( \mu = \lambda/\beta \).

We will now find what relations there are between the Poisson model and the AR(1)-model.

Let the number of occupied servers (e.g. junctors) at the time \( t \) be denoted \( X_t \). Then:

(i) \( X_t \) Poisson distributed with expectation \( \lambda/\beta = \mu \)

(ii) \( \text{Corr}(X_t, X_{t-k}) = e^{-k\beta} = \phi^k \)

(iii) \( E(X_t | X_{t-k} = x_{t-k}) = \mu + \psi(x_{t-k} - \mu) \)

(iv) \( \text{Var}(X_t | X_{t-k} = x_{t-k}) = \sigma_x^2 (1 - \phi^k) + \varphi^k (x_{t-k} - \mu)^2 \)

If, on the other hand, \( X_t \) follows an AR(1)-process, then

(i) \( X_t \) Normally distributed with expectation \( \mu \) and variance \( \sigma_x^2 \)

(ii) \( \text{Corr}(X_t, X_{t-k}) = \phi^k \)

(iii) \( E(X_t | X_{t-k} = x_{t-k}) = \mu + \psi(x_{t-k} - \mu) \)

(iv) \( \text{Var}(X_t | X_{t-k} = x_{t-k}) = \sigma_x^2 (1 - \phi^k) \)

The two models differ at (i) and (iv). Concerning (i) we know that the Poisson distribution is asymptotical normal if \( \mu = \lambda/\beta \) and that the normalapproximation is satisfactory for \( \mu > 5 \).

If the traffic is not too low, the marginal distributions for the two situations are therefore not too different.

The Poisson model has the property that

\( \text{Var}(X_t) = \text{E}(X_t) = \mu \)

That is not the case for the AR(1)-model which has a general variance \( \sigma_x^2 \).

The difference in (iv) can be more crucial. If \( \sigma_x^2 = \mu \) the conditional variance for the two models will differ by a factor

\( \phi^k (1 - \phi^k) (X_{t-k} - \mu) \)

while the unconditional variances will be equal. The ratio between the two conditional variances is, however:

\( \text{Var}(X_t | X_{t-k} = x_{t-k} | \text{poisson}) / \text{Var}(X_t | X_{t-k} = x_{t-k} | \text{AR} (1)) = 1 + \phi^k \mu / \phi^k \)

which tends to 1 with probability 1 as \( \mu \to \infty \).

We conclude that the AR(1)-model can give a satisfactory representation of a Poisson model if the traffic is not too low.

If the Poisson model can be treated as the "true model", we will have the following relations:

\( \phi = \lambda/\beta \Rightarrow \beta = \lambda/\phi \)

\( \mu = \lambda/\beta = \lambda - \mu \lambda \phi \)

(4.1)

\( \sigma_x^2 = (1 - \phi^2) \mu \)

By using (4.1) the interarrivaltimes 1/\( \lambda \) and service-times 1/\( \mu \) can be calculated by means of the estimated parameters in the AR(1)-model and vice versa. In table 4.3 we give such results for the routes presented in table 4.1. The columns \( \sigma_x^2 \) and \( \lambda/\beta \) should be equal if the Poisson model is the true model. The table shows that this may be so for the Routes 1 and 2 while Route 3
4.4 Different time resolutions

It is generally not the case that the best ARMA-model at one time resolution necessarily gives the best fit for other time resolutions. It is therefore of interest to study the influence of different time resolutions to the model fit.

If \( X_t \), \( t = 1, 2, \ldots, n \) is a time series with expectation 0 generated by an AR(1)-process, then \( X_t \) is also AR(1). If the parameters for the \( X_t \) series are \( \phi \) and \( \sigma_x^2 \), the corresponding parameters for the \( X_t \) series will be \( \phi^k \) and \( \sigma_x^2 \phi^2 k(1-\phi^2) \).

We have studied how this relation compares with the results from our simulation experiments. Time series measuring the number of occupied juncors every 5., 10, and 15. second have been analysed for the three exchanges. For all series the AR(1)-model gave the best fit.

The autoregressive parameter, \( \phi \), decreased close to \( \phi^2 \) and \( \phi^3 \) as the time resolution increased to 10 and 15 seconds. Some examples are given in table 4.4. The routes presented are randomly chosen and representative for the results.

Table 4.4 Autoregressive parameters for different time resolutions

<table>
<thead>
<tr>
<th>Routes</th>
<th>AR-parameter</th>
<th>AR-parameter</th>
<th>AR-parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>at 3. sec</td>
<td>at 10. sec</td>
<td>at 15. sec</td>
</tr>
<tr>
<td>Route 1 GROSS</td>
<td>0.9500.07</td>
<td>0.8650.11</td>
<td></td>
</tr>
<tr>
<td>Route 4 GROSS</td>
<td>0.9400.07</td>
<td>0.8350.11</td>
<td></td>
</tr>
<tr>
<td>Route 9 GROSS</td>
<td>0.6609.09</td>
<td>0.3505.14</td>
<td></td>
</tr>
<tr>
<td>ASK BB-GSTGD B</td>
<td>0.9020.05</td>
<td>0.7220.14</td>
<td></td>
</tr>
<tr>
<td>ASK BB-ASK 2</td>
<td>0.9750.03</td>
<td>0.9110.08</td>
<td></td>
</tr>
<tr>
<td>OSLO PI-ASK BB</td>
<td>0.9330.04</td>
<td>0.7880.12</td>
<td></td>
</tr>
</tbody>
</table>

From the table the exponential decreasing course of the parameters is clear.

For the registergroup in the AKD-exchange we found the ARMA(1,1)-model for the time resolution of 1 second. Observing the time series every 3. second gives 500 observations and we would expect an AR(1)-model. Because of low autoregressive parameter-value the autocorrelation function was significantly nonzero only at lag 1. This led us to fit a MA(1)-model to this time series. The estimation results are given in table 4.5.

Table 4.5 Estimation results for the registergroup in the AKD-exchange. Time resolution is 3 seconds

<table>
<thead>
<tr>
<th>Model</th>
<th>( \theta )</th>
<th>( \mu )</th>
<th>( \sigma_\theta )</th>
<th>( \omega_\varphi )</th>
</tr>
</thead>
<tbody>
<tr>
<td>MA(1)</td>
<td>-2.820.08</td>
<td>7.9020.34</td>
<td>3.0</td>
<td>34.82</td>
</tr>
</tbody>
</table>

4.5 Time series consisting of average traffic from short simulation periods

Instead of observing the number of occupied servers at given times, we can consider the sequence of observations consisting of the average traffic in short periods. For the routes in the ASKIM BB-exchange ARMA-models for such time series have been fitted. We chose a period of 30 seconds and the autocorrelation function showed that the AR(1)-model could be improved by introducing one MA-parameter, i.e. the ARMA(1,1)-model. In table 4.6 we give the fitted ARMA-models for three of the routes in ASKIM BB.

Table 4.6 Fitting of an ARMA(1,1)-model the time series consisting of the average traffic in periods of 30 seconds

<table>
<thead>
<tr>
<th>Route</th>
<th>Number of observations</th>
<th>( \omega )</th>
<th>( \theta )</th>
<th>( \mu )</th>
<th>( \sigma_\theta )</th>
<th>( \omega_\varphi )</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASK BB-OSLO PI 50</td>
<td>0.6920.26</td>
<td>-2.86132</td>
<td>47.021442</td>
<td>3.3</td>
<td>8.42</td>
<td></td>
</tr>
<tr>
<td>ASK BB-ASK 50</td>
<td>0.76195</td>
<td>-0.43229</td>
<td>18.024166</td>
<td>2.4</td>
<td>15.06</td>
<td></td>
</tr>
<tr>
<td>OSLO PI-ASK BB 50</td>
<td>0.56030</td>
<td>-2.90139</td>
<td>15.780161</td>
<td>2.0</td>
<td>11.42</td>
<td></td>
</tr>
</tbody>
</table>

In section 5.2 it is shown theoretically that if the originated process actually is AR(1), then the process of the means will be ARMA(1,1), confirming the findings of this section.

5 CONFIDENCE LIMITS FOR THE EXPECTED TRAFFIC IN SIMULATION MODELS

5.1 Estimation of the confidence limits based on the assumption that the observations are independent

One problem in simulation is to estimate the confidence of the estimated parameters in the simulated process. Since the output from a simulation model partly consists of highly correlated observations, it is difficult to find an unbiased confidence interval, since the methods to obtain such an interval is based on the assumption of independence.

The traditional methods to find the confidence interval is either based on repeated simulations or on long simulations where parts of the simulation are eliminated due to the dependence of the observations. As mentioned earlier a lot of simulation time is lost in repeated simulation since all the transient periods have to be eliminated. In long simulations, too, information is removed before the confidence limits are estimated.

Suppose we want to use the traditional method by making a long simulation to find the confidence limits of the expected traffic. Then the simulation is divided into simulation periods where every second period containing \( m_1 \) observations is eliminated. Let the number of intervals retained be \( n \) and the number of observations within the periods be \( m_1 \). Figure 5.1 shows the partition of the simulation.

Figure 5.1 A long simulation divided in \( n \) simulation periods

Suppose that \( X_t \) is a stochastic variable assigning the traffic at time \( t \). Let \( \{X_t\} \) be Normally distributed \( N(\mu, \sigma^2) \) and observed at equidistant times \( t = 0, 1, 2, \ldots \) Then the mean of the \( k \)-th simulation period, \( Y_k \), is given by

\[
Y_k = \frac{1}{m_1} \sum_{i=0}^{m_1-1} X(k-1)+i = k + 1, 2, \ldots , n
\]

where
\[ m = m_1 + m_2 \]

An unbiased estimated of \( \mu \) is then

\[ \bar{Y} = \frac{1}{n} \sum_{k=1}^{n} Y_k \]  
(5.1)

The best unbiased estimate \( \bar{Y} \) of \( \mu \) is, however, based on all the observations.

\[ \bar{Y} = \frac{1}{N} \sum_{t=0}^{N} X_t \]
(5.2)

where \( N \) is the total number of observations in the long simulation.

An estimate of \( \text{Var} \bar{Y} \) based on the assumption that \( Y_1, Y_2, \ldots, Y_n \) are independent, is

\[ \sigma^2_1 = \frac{1}{n(n-1)} \sum_{k=1}^{n} (Y_k - \bar{Y})^2 \]
(5.3)

Then a (1-\( \alpha \)) confidence interval of \( \mu \) is given by

\[ \bar{Y} \pm t_{1-\alpha/2, n-1} \frac{\sigma_1}{\sqrt{n}} \]
(5.4)

where \( t_{1-\alpha/2, n-1} \) is the \( (1-\alpha/2) \) fractile in the Student's \( t \)-distribution with \( n-1 \) degrees of freedom.

In chapter 4 it is shown that the simulated number of occupied junctors are autoregressive processes of the first order. This can be utilized to show how much the estimated variance (5.3) deviates from the real variance of \( Y \). Let \( Y_k \) be given by

\[ Y_t = \varphi Y_{t-1} + \sigma \]

Then

\[ \text{Var} X_t = \sigma^2 \]
(5.5)

\[ \text{Cov}(X_t, X_{t+k}) = \varphi^k \sigma^2 \]  
(5.6)

The variance of \( Y_k \) then is

\[ \text{Var} Y_k = (\varphi^2)^{m_1-1} \text{Var}(X(k-1)m+1) + 2 \sum_{1 \leq j < k} \text{Cov}(X(k-1)m+1, X(k-1)m+j) \]
(5.7)

where

\[ \text{Var}(X(k-1)m+1) = (m_1 + \text{Var}(X(k-1)m+1)) \frac{\sigma^2}{m_1} \]

\[ \text{Cov}(X(k-1)m+1, X(k-1)m+j) = \varphi^j \sigma^2 \]

The covariance between \( Y_t \) and \( Y_{t+k} \) is given by

\[ \text{Cov}(Y_t, Y_{t+k}) = (\varphi^2)^{m_1-1} \frac{\sigma^2}{m_1} \]  
(5.8)

where

\[ \text{Var}(X(k-1)m+1) = (m_1 + \text{Var}(X(k-1)m+1)) \frac{\sigma^2}{m_1} \]

\[ \text{Cov}(X(k-1)m+1, X(k-1)m+j) = \varphi^j \sigma^2 \]

and \( \varphi = \varphi \)

Now the exact variance of \( \bar{Y} \) given by (5.1), can be obtained. The variance, \( \sigma^2_2 \), of \( \bar{Y} \) taking into account the dependence of the observations is

\[ \sigma^2_2 = \frac{1}{n} \sum_{k=1}^{n} \text{Var} Y_k + 2 \sum_{1 \leq k < n} \text{Cov}(Y_1, Y_k) \]

By using (5.7) and (5.8) we obtain

\[ \sigma^2_2 = \frac{1}{n} \sum_{k=1}^{n} \text{Var} Y_k + 2 \sum_{1 \leq k < n} \text{Cov}(Y_1, Y_k) \]

Substituting the estimates of the autocorrelation \( \hat{\rho} \) and the variance of the white noise \( \sigma^2 \) in (5.9), we obtain an estimate of \( \text{Var} \bar{Y}, \hat{\sigma}^2 \). Using the asymptotic normality of \( \bar{Y} \), approximate confidence intervals may be constructed.

The process is supposed to be independent in the \( n \) simulation periods, then the expression

\[ \frac{1}{n} \sum_{k=1}^{n} \text{Var} Y_k = n(m_1 + \text{Var}(X(k-1)m+1)) \frac{\sigma^2}{m_1} \]

is the variance of \( \bar{Y} \). If, however, the process is dependent then:

\[ 2 \sum_{1 \leq k < n} \text{Cov}(Y_1, Y_k) \]

\( \leftrightarrow \frac{1}{n} \sum_{k=1}^{n} \text{Var} Y_k = n(m_1 + \text{Var}(X(k-1)m+1)) \frac{\sigma^2}{m_1} \]

(5.10)

expresses the increase of the variance of \( \bar{Y} \). In chapter 5.3 the relation between (5.10) and (5.11) are studied for different values of \( m_1, m_2 \) and \( n \).

Suppose that the gap lengths, \( m_2 \), are zero. Then \( m = m_1 \) and \( N = nm \). Since

\[ n \varphi + C(m_1, \varphi) V(n, \varphi^2) = V(N, \varphi^2) \]

the equation (5.9) gives:

\[ \sigma^2_2 = (n+\varphi(n, \varphi^2)) \frac{1}{n} \frac{\sigma^2}{1-\varphi^2} \]

(5.12)

and the variance, as expected, is independent of \( m_1 \).

The variance, \( \sigma^2_2 \), can in this case also be found directly by using (5.2), (5.5) and (5.6).

Using the method now pointed out several difficult problems due to the estimation of the confidence limits are eliminated.

1. It is not necessary to throw away observations that are making gaps in the simulations to obtain independent observations.

2. It is not necessary to use the observations from one simulation period as one observation to estimate the confidence limits.

3. It is not necessary to adjust for the (week) dependence between the mean values before estimating the confidence limits.

The equation (5.8) shows that the autocovariance between the mean traffic in two simulation periods \( m \) observations apart decreases with the factor \( \varphi^2 = \varphi^2 \). It is pointed out in chapter 4.4 that the same is true for the autocovariance between two observations in an autoregressive process \( m \) observations apart.

There exists an alternative method to find the autocovari­ance expression (5.11) when the original process is autoregressive of first order. Instead of analyzing the originally process \( X_t \), the process of the means \( Y_k \) can be analyzed directly. The autocovariances of the process of traffic means are given in (5.7) and (5.8). As mentioned in chapter 3.3 the structure of the autoco­variances or autocorrelations identifies the process. In [1] it is shown that an autocorrelation function with a spike at lag 1 followed by an exponentially decreasing behaviour is typical for an ANMA(1,1) process. The time series of the means is consequently an ANMA(1,1) process.

Hence the parameters \( \varphi \) and \( \theta \) in the model:

\[ Y_k = \varphi Y_{k-1} + b_k + \theta^2 \]

(5.13)

\[ k = 1, 2, \ldots, n \]

can be estimated. \( b_k \) is the white noise process in this model and \( \sigma^2 \) is the variance of \( b_k \). It is shown in [1] that the covariances of an ANMA(1,1)-model is given by

\[ \text{Var} Y_k = \frac{1+\theta^2-2\varphi\theta}{1-\varphi^2} b_k^2 \]

STORDAHL / SOLLE / DAMSLETH-5
where \( \{s_{ij}\} = \text{Cov} \)

\[ \text{Cov}(X, Y) = \text{Cov}(X, Y) + \text{Cov}(X, Y) \]

\[ j > 1 \]

Then the variance of \( \bar{Y} \) can be found substituting (5.13) in (5.9). Hence an alternative estimate of the variance of \( \bar{Y} \) is given.

It has, however, to be emphasized that this method can only be used when the process is autoregressive of first order. The analysis in this paper shows that the Junctor traffic is autoregressive, while the register traffic is not.

5.2 Construction of a confidence interval for the expected value of the process using the methods of Box and Jenkins

As stated in 3.4 the unknown parameters in an ARMA-model are estimated by the maximum likelihood principle.

For the approximately parameterized ARMA-model, the log likelihood will be approximately quadratic in the parameters \( \beta \) is the parameter vector consisting of \( \Psi \)'s, \( \theta \)'s and \( \mu \) so that

\[ \ell(\beta) = \ell(\beta, \theta) \]

\[ \ell(\beta) = \frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{k} \left( 8_s 8_t - S_{ij} \right) \]

where, to the approximation considered, the derivatives

\[ \ell_{ij} = \frac{\partial \ell(\beta, \theta)}{\partial 8_i 8_j} \]

are constant.

For moderate and large samples, where the local quadratic approximation is adequate, useful approximations to the variances and covariances of the estimates and approximate confidence regions may be obtained.

The \((kxk)\) matrix \( -E[\ell_{ij}]^{-1} \) is usually referred to as the information matrix for \( \beta \). For a given value of \( \theta \), the variance - covariance matrix \( \text{Var} \) for the maximum likelihood is, for large samples given by the inverse of this information matrix. That is

\[ \text{Var}(\beta) \approx [-E[\ell_{ij}]]^{-1} \]

(5.14)

It can be shown that

\[ \ell_{ij} = \frac{S_{ij}}{8_i 8_j} \]

where

\[ S_{ij} = \frac{3}{2} \sum_{t=0}^{k} \left( 8_s 8_t \right)^2 \]

and

\[ S(\beta) = \frac{1}{n} \sum_{t=0}^{k} \left( 8_s 8_t \right)^2 \]

The \((kxk)\) matrix \( -E[\ell_{ij}] \) is usually referred to as the expected information matrix of \( \ell_{ij} \) or of \( S_{ij} \) by the values actually observed then using (5.14)

\[ \text{Var}(\beta) \approx [-E[\ell_{ij}]]^{-1} \approx 2\text{Var}(\beta) \approx [S_{ij}]^{-1} \]

\( S \) is usually estimated by

\[ \delta_8 = \frac{S(\beta)}{n} \]

and for large samples \( \delta_8 \) and \( \delta \) is uncorrelated.

Finally, the elements of \( \text{Var}(\beta) \) can be estimated from

\( \text{Cov}(\beta, \beta) = \delta_8 \delta^{-1} \)

where \( S(\beta) = (\delta_s)^{-1} \)

In particular, these results allow us to obtain the approximate variances of our estimates. By taking the square root of these variances, we obtain approximate standard deviations, which are usually called the standard errors of the estimates, \( 8 \text{SE}(\delta) \).

A 95% confidence interval for the mean \( \mu \) is then approximated by

\[ \hat{\mu} \pm 2\text{SE}(\hat{\mu}) \]

5.3 Comparison

In table 5.1 we give point estimates and confidence intervals for the mean (expected) traffic on four of the routes in the ASKIM BB-exchange. The confidence level is approximately 95% if the underlying assumptions are fulfilled.

In the first column the estimates are given assuming an AR(1)-model for the time series measuring the number of occupied outgoing junctors every 15. sec. The second column gives the estimates based upon the time series consisting of the average traffic in 30. sec. intervals modelled as an ARMA(1,1)-process. For the last columns, 3-5, the intervals are computed on the basis of the average traffic in shorter simulation periods, considered to be uncorrelated, as given in equation (5.4).

These last estimates suffers from two weaknesses:

1) A tendency of underestimation caused by not taking into account the autocorrelation when computing the variance. This results in an actual confidence level less than 95%.

2) A tendency of overestimation due to the loss of information caused by the simulation gaps.

From equations (5.10) and (5.11) it is possible to compute correction terms for the underestimated variance when the \( \beta \)-parameter in the AR(1)-model is known. Estimation of \( \beta \) for the four routes considered here gave the results given in table 5.2 below.

Table 5.1 Estimated mean traffic of four routes in the ASKIM BB-exchange. The confidence levels of the intervals are approximately 95% when the underlying assumption are fulfilled.

<table>
<thead>
<tr>
<th>Route</th>
<th>Time series observed: 15 sec, 30 sec, intervals AR(1)</th>
<th>Time series observed: 15 sec, intervals ARIMA(1,1)</th>
<th>Simulation period: 30 sec</th>
<th>Simulation period: 15 sec</th>
<th>Simulation period: 225 sec</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASK BB - SPB BB</td>
<td>23.901±1.29</td>
<td>23.410±1.76</td>
<td>23.815±1.96</td>
<td>23.811±1.96</td>
<td>23.812±1.96</td>
</tr>
<tr>
<td>ASK BB - ASK 1</td>
<td>21.19±1.94</td>
<td>20.75±1.99</td>
<td>20.77±1.97</td>
<td>20.77±1.97</td>
<td>20.77±1.97</td>
</tr>
<tr>
<td>MY 7D - ASK BB</td>
<td>18.34±4.16</td>
<td>18.94±4.40</td>
<td>18.32±4.14</td>
<td>18.32±4.14</td>
<td>18.32±4.14</td>
</tr>
</tbody>
</table>

Table 5.2 Autoregressive parameter for the four routes in the ASKIM BB-exchange.

<table>
<thead>
<tr>
<th>Route</th>
<th>( \beta )</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASK BB - SPB BB</td>
<td>0.93±0.05</td>
</tr>
<tr>
<td>ASK BB - ASK 1</td>
<td>0.93±0.05</td>
</tr>
<tr>
<td>ASK BB - ASK 3</td>
<td>0.93±0.05</td>
</tr>
<tr>
<td>MY 7D - ASK BB</td>
<td>0.93±0.05</td>
</tr>
</tbody>
</table>

Under the assumption that the 95% confidence intervals can be approximated by "two times the standard error of the mean" we give correction factors for the confidence intervals presented in table 5.1. They give an indication of the amount of error due to 1) and are presented in table 5.3.

The estimates in column 1 and 2 of table 5.1 are given within the model assumption and can be treated as "the best we can achieve" for these processes. If we multiply the confidence intervals in table 5.1 with their correction factors given in table 5.3. We obtain intervals with approximately the same confidence level. We can
Table 5.3 Correction factors for the underestimated confidence intervals in table 5.1

<table>
<thead>
<tr>
<th>Route</th>
<th>Sim. per. 30 sec, Sim. gap 0 s</th>
<th>Sim. per. 30 sec, Sim. gap 120 s</th>
<th>Sim. per. 115 sec, Sim. gap 120 s</th>
<th>Sim. per. 225 sec, Sim. gap 200 s</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASK 8B - SPB 8B</td>
<td>1.93 1.39 1.05 1.02 1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASK 8B - ASK 1</td>
<td>2.28 1.60 1.12 1.06 1.01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASK 8B - ASK 3</td>
<td>2.59 1.84 1.24 1.13 1.03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MY 7D - ASK 8B</td>
<td>3.21 2.28 1.48 1.30 1.10</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.4 Estimated mean traffic of four routes in the ASKIM 8B-exchange. The confidence levels of the intervals are approximately 95%. Corrections are made according to eq. (5.10) and (5.11)

<table>
<thead>
<tr>
<th>Route</th>
<th>Time series observed 15 sec average</th>
<th>Time series based upon average traffic volume (AAR)</th>
<th>Simulation periods 30 sec. intervals</th>
<th>Simulation periods 115 sec</th>
<th>Simulation periods 225 sec. intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASK 8B - SPB 8B</td>
<td>23.901±5.3</td>
<td>23.901±5.3</td>
<td>23.811±3.5</td>
<td>23.811±3.5</td>
<td>23.811±3.5</td>
</tr>
<tr>
<td>ASK 8B - ASK 1</td>
<td>21.15±1.94</td>
<td>21.15±1.94</td>
<td>20.77±1.05</td>
<td>20.77±1.05</td>
<td>20.77±1.05</td>
</tr>
<tr>
<td>MY 7D - ASK 8B</td>
<td>18.64±3.5</td>
<td>18.64±3.5</td>
<td>19.32±1.30</td>
<td>19.32±1.30</td>
<td>19.32±1.30</td>
</tr>
</tbody>
</table>

Table 5.5 Estimated standard deviation $\sigma_x$ and $\sigma_1$ for different values of $m_2$ of the traffic of four routes

<table>
<thead>
<tr>
<th>Jumctor group</th>
<th>$m_2$</th>
<th>6</th>
<th>6</th>
<th>12</th>
<th>18</th>
<th>24</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASK 8B - SPB 8B</td>
<td>0.38</td>
<td>0.58</td>
<td>0.68</td>
<td>0.53</td>
<td>0.83</td>
<td>0.77</td>
</tr>
<tr>
<td>ASK 8B - ASK 1</td>
<td>0.45</td>
<td>0.72</td>
<td>0.80</td>
<td>0.98</td>
<td>1.28</td>
<td>0.97</td>
</tr>
<tr>
<td>ASK 8B - ASK 3</td>
<td>0.50</td>
<td>0.76</td>
<td>0.84</td>
<td>1.03</td>
<td>1.42</td>
<td>1.19</td>
</tr>
<tr>
<td>MY 7D - ASK 8B</td>
<td>0.67</td>
<td>1.05</td>
<td>1.17</td>
<td>1.59</td>
<td>1.92</td>
<td>2.18</td>
</tr>
</tbody>
</table>

$\sigma_x$ is obtained in (5.12):

$$\sigma_x^2 = \left( \frac{N + V(N, \psi) }{N} \right) \frac{\sigma^2}{N^2}$$

By setting

$$E_{\psi}^2(m_2|m_1, \psi) = \sigma^2_x$$

$m_2$ can be found by solving this unlinear equation.

Table 5.5 shows the estimated standard deviation $\sigma_x$ and $\sigma_1$ for different values of $m_2$ of some simulated traffic processes.

$\rho_1$ is the number of observations in the simulation periods. $m_2$ is the number of observations in the eliminated periods.

Table 5.5 shows that the number of observations $m_2$, necessary to obtain $\hat{\theta}_x$, $\hat{\theta}_1$ naturally increases as a function of $\psi$. If $\psi$ is known and $m_1$ is given, then $m_2$ can be found by solving the equation (5.16).

However, it must be emphasized that it is difficult to find the number of observations to eliminate, $m_2$, if the autoregressive parameter $\psi$ is not known.

REFERENCES


