Level process estimation and model selection

Azam Asanjarani, Yoni Nazarathy

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

ACEMS working seminar
12 May 2014
Devise, analyse and compare methods for fitting different stochastic models for level processes

Accuracy, Versatility, Computation time, Robustness, ...

Continuous time Markov chains with a countable state-space
Stable and Stationary
Underlying offered load $\rho$

Prediction for performance measures as a function of $\rho$
• **Structure of Observed Data**
• Parameter Estimation
  ▶ Different Models
  ▶ Different Techniques
• Model Selection
• Post Estimation Analysis
• Sub-projects
Structure of observed data: $\mathcal{D} = (\ell_0, t_i, b_i)$

- $\ell_0 \in \{0, 1, \ldots\}$ initial level
- $t_i \in \mathbb{R}_+ \setminus \{0\}$ and $\{t_i\}_{i=1}^{N}$ the times between level changes
- $b_i \in \{-1, +1\}$ and $\{b_i\}_{i=1}^{N}$ level changes

![Diagram showing level changes and times between changes](image-url)
Structure of observed data: \( \mathcal{D} = \{ (\ell_i, t_i) \}_{i=0}^{N} \)

- \( \ell_i = \ell_0 + \sum_{j=1}^{i} b_j \)
Structure of observed data: $\mathcal{D} = \{(\ell_i, T_i)\}_{i=0}^{N}$

- $T_0 = 0$, $T_i = \sum_{j=1}^{i} t_j$
- $t \in [T_i, T_{i+1})$ the level of the process is $\ell_{i+1}$
Structure of observed data: $\mathcal{D} = (\ell_0, \tau_i, s_i)$

- $\mathcal{T}_A = \{T_i : b_i = +1\}$ and $\{T_i^A\}_{i=1}^{N_A}$ arrival times
- $\mathcal{T}_D = \{T_i : b_i = -1\}$ and $\{T_i^D\}_{i=1}^{N_D}$ departure times
- $N_A + N_D = N$
Structure of observed data: $\mathcal{D} = (\ell_0, \tau_i, s_i)$

- $\tau_i = T_{i+1}^A - T_i^A$, $i = 1, \ldots, N^A - 1$
- $s_i = T_{i+1}^D - 1\{\ell_i > 0\} T_i^D - 1\{\ell_i = 0\} T_i^A$, $i = 1, \ldots, N^D - 1$
The length of our data may be characterized in three ways:

- \( N \) is fixed
- \( N^A \) is fixed
- \( N^D \) is fixed

\[ T_{\text{max}} := T_N \]

- \( T_{\text{max}} \) is fixed and \( N = \inf \{ n : T_n \geq T_{\text{max}} \} \)

- \( \tau_B \), the busy time is fixed, when

\[ \tau_B = \sum_{i=1}^{N} t_i \mathbb{1} \{ \ell_i > 0 \} \]

In general we have big \( N \) (alt. \( T_{\text{max}} \)) in mind.
Different observation/different estimation (M/M/∞)

Vaclav E Beneš.
A sufficient set of statistics for a simple telephone exchange model. 

- Initial state
- $N^A$
- $N^D$
- Mean number of jobs

Ronald W Wolff.
Problems of statistical inference for birth and death queuing models. 

- Number of arrivals to state $i$
- Number of services finished at state $i$
- Total time spend in state $i$
• Structure of Observed Data

• Parameter Estimation
  ▶ Different Models
  ▶ Different Techniques

• Model Selection

• Post Estimation Analysis

• Sub-projects
Parameter estimation

Nazarathy, Yoni and Pollet, Phil
Parameter and State Estimation in Queues and Related Stochastic Models: A Bibliography
Nov. 2012
Parameter estimation

Nazarathy, Yoni and Pollet, Phil
Parameter and State Estimation in Queues and Related Stochastic Models: A Bibliography
Nov. 2012

\[ D \text{ and } M \]

\[ \hat{\theta}_M \text{ and } \hat{\rho}_M \]
Parameter estimation

Nazarathy, Yoni and Pollet, Phil
Parameter and State Estimation in Queues and Related Stochastic Models: A Bibliography
Nov. 2012

$D$ and $M$

estimation procedure

$\hat{\theta}_M$ and $\hat{\rho}_M$

$\rho = \text{mean service time} \times \text{arrival rate}$
Outline

• Structure of Observed Data
• Parameter Estimation
  ▶ Different Models
  ▶ Different Techniques
• Model Selection
• Post Estimation Analysis
• Sub-projects
Parameter estimation based on different models

\[ QBD_r \]

\[ MAP_p/PH_q/1 \]

\[ PH_p/PH_q/1 \]
\[ MAP_p/M/1 \]

\[ M/PH_q/1 \]
\[ PH_p/M/1 \]

\[ M/M/1 \]
Parameter estimation based on different models

QBD\(_r\)

MAP\(_p\)/PH\(_q\)/1

MAP\(_p\)/M/1

PH\(_p\)/PH\(_q\)/1

PH\(_p\)/M/1

M/PH\(_q\)/1

M/M/1

Inter-arrival and service processes do not interact
Parameter estimation based on different models

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M/M/1$</td>
<td>$\lambda, \mu$</td>
</tr>
<tr>
<td>$PH_p/PH_q/1$</td>
<td>$(\alpha_{1\times p}^a, T_{p\times p}^a), (\alpha_{1\times q}^s, T_{q\times q}^s)$</td>
</tr>
<tr>
<td>$QBD_r$</td>
<td>$A_{-1}, A_0, A_1$</td>
</tr>
</tbody>
</table>
\( Q = \begin{pmatrix}
-\lambda & \lambda & 0 & 0 & 0 \\
\mu & -(\mu + \lambda) & \lambda & 0 & \cdots \\
0 & \mu & -(\mu + \lambda) & \lambda & \cdots \\
0 & 0 & \mu & -(\mu + \lambda) & \cdots \\
\cdots & \cdots & \cdots & \cdots & \cdots 
\end{pmatrix} \)
Parameter estimation for $M/M/1$

- Obvious estimates for $\lambda$ and $\mu$:

$$\hat{\lambda} = \frac{N^A}{T} \quad \hat{\mu} = \frac{N^D}{\tau_B}$$

- $N^A$ total number of arrivals
- $T$ fixed time of observation
- $N^D$ total number of departures
- $\tau_B$ busy time
When we consider the stationary case $\rho = \frac{\lambda}{\mu} < 1$:

$$\hat{\lambda} \approx \frac{N^A + \nu}{T_m} \quad \hat{\mu} \approx \frac{N^D - \nu}{\tau_B}$$

- $N^A$ total number of arrivals
- $\nu$ initial queue size
- $T_m$ time of $m$th(last) departure
- $N^D$ total number of departures
- $\tau_B$ busy time
\[ Q = \begin{pmatrix} 0 & 0 \\ -T_1 & T \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ t & T \end{pmatrix} \]
\( \mathbf{Q} = \begin{pmatrix} 
B_1 & B_0 & 0 & 0 & \cdots \\
B_2 & A_0 & A_1 & 0 & \cdots \\
0 & A_{-1} & A_0 & A_1 & \cdots \\
0 & 0 & A_{-1} & A_0 & \ddots \\
\vdots & \vdots & \vdots & \vdots & \ddots
\end{pmatrix} \)
$MAP_2$ (Markovian arrival process)
Outline

• Structure of Observed Data
• Parameter Estimation
  ▶ Different Models
  ▶ Different Techniques
• Model Selection
• Post Estimation Analysis
• Sub-projects
Method of moments

- \( f_X(x; \theta) \) the distribution of random variable \( X \)
- \( \theta = (\theta_1, \theta_2, \ldots, \theta_k) \) vector of unknown parameters
- The first \( k \) population moments:
  \[
  \mu_j = \mathbb{E}[X^j] = g_j(\theta_1, \theta_2, \ldots, \theta_k) \quad j = 1, \ldots, k
  \]
- The first \( k \) sample moments:
  \[
  \hat{\mu}_j = \frac{1}{n} \sum_{i=1}^{n} x_i^j \quad j = 1, \ldots, k
  \]
  where \( n \) is sample size
- \( \hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k) \) is the solution of the equations:
  \[
  \begin{align*}
  \hat{\mu}_1 &= g_1(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k) \\
  \hat{\mu}_2 &= g_2(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k) \\
  &\vdots \\
  \hat{\mu}_k &= g_k(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k)
  \end{align*}
  \]
Given a statistical model, 
\[ y = (y_1, y_2, \ldots, y_n) \] a set of observed data 
\[ \theta \] a vector of unknown parameters

**MLE** :
Likelihood function:

\[
L(\theta; y) = f_Y(y|\theta)
\]

\[
\hat{\theta}_{MLE} = \arg \max_{\theta} L(\theta; y)
\]
The EM algorithm

\( Z \) a set of missing values or hidden data

The EM algorithm tries to find the MLE of \( \theta \) by:

1. **Expectation step**: Calculate 
   \[
   h(\theta) = \mathbb{E}_{Z|y, \theta_0} [\log L(\theta; y, Z)]
   \]
2. **Maximization step**: Maximize \( h \) to find
   \[
   \hat{\theta} = \arg \max_{\theta} h(\theta|\theta_0)
   \]
3. Set \( \theta_0 = \hat{\theta} \) and go to step 1
Outline

- Structure of Observed Data
- Parameter Estimation
  - Different Models
  - Different Techniques
- Model Selection
- Post Estimation Analysis
- Sub-projects
Model selection

\( \mathcal{M} \) set of possible models

Purpose of model selection

- Identify a **useful model** \( M \in \mathcal{M} \) whether or not a true model exists

1. choose some estimation methods resulting in \( \hat{\theta}_M \)
2. calculate \( \hat{\theta}_M \) for all \( M \in \mathcal{M} \)
3. compare the different models

- Possible extension: Consider several estimators for the models
Selection criterion: AIC and BIC

- The Akaike information criterion:
  \[ AIC(M) = -2 \ln(L(\theta_M)) + 2k \]
  \( k = \text{length } \theta_M, \ L = \text{the maximized value of the likelihood function} \)

- Bayesian information criterion:
  \[ BIC(M) = -2 \ln(L(\theta_M)) + \ln(n)k \]
  \( n = \text{the sample size} \)
Selection criterion: AIC and BIC

- The Akaike information criterion:
  \[
  AIC(M) = -2 \ln(L(\theta_M)) + 2k
  \]
  \(k = \text{length } \theta_M, \ L = \text{the maximized value of the likelihood function}\)

- Bayesian information criterion:
  \[
  BIC(M) = -2 \ln(L(\theta_M)) + \ln(n)k
  \]
  \(n = \text{the sample size}\)

Samuel Müller, High dimensional data, 2014 AMSI Summer School
Selection criterion: AIC and BIC

- The Akaike information criterion:
  \[ AIC(M) = -2 \ln(L(\theta_M)) + 2k \]
  \[ k = \text{length } \theta_M, \ L = \text{the maximized value of the likelihood function} \]
- Bayesian information criterion:
  \[ BIC(M) = -2 \ln(L(\theta_M)) + \ln(n)k \]
  \[ n = \text{the sample size} \]

- The smaller the AIC/BIC the better the model
Outline

- Structure of Observed Data
- Parameter Estimation
  - Different Models
  - Different Techniques
- Model Selection
- Post Estimation Analysis
- Sub-projects
Post estimation analysis

1. $M^*$ selected model
2. $\hat{\theta}^*$ estimated parameters
3. $\hat{\rho}^*$ estimated offered load

Analyse key performance measures as a function of $\rho$

- The dependence of the steady state mean level on $\rho$. Denote it \( \overline{L} := \mathbb{E}[L] \).
- The same for the variance of the levels.
The more novel aspect of post-estimation analysis is giving **confidence bounds** for the performance measures.
Outline

- Structure of Observed Data
- Parameter Estimation
  - Different Models
  - Different Techniques
- Model Selection
- Post Estimation Analysis
- Sub-projects
$\mathcal{QBD}_2^{-}$

Phase $A = 0$

Level $A_0 = 0$

Level $A_1$

Phase $A_{-1}$

Level $A_{-1}$

Level $A_0$

Level $A_1$

35
Inference based on inter-arrivals and services

In some of the models, the parameters can be estimated directly from their inter-arrival sequence and service sequence. For example for $PH/PH/1$, we could consider

\[
\hat{\theta}_a = \arg \max L_A(\alpha^a, T^a; \text{Arrival data}) \quad \hat{\theta}_s = \arg \max L_S(\alpha^s, T^s; \text{Service data})
\]

\[
\hat{\theta} = (\hat{\theta}_a, \hat{\theta}_s)
\]
Thank you!