Markov Chain Models: Computations and Applications

Wai-Ki CHING
AMACL, Department of Mathematics
The University of Hong Kong

Abstract: Markov chain models are popular mathematical tools for studying many different kinds of real world systems such as queueing networks (continuous time) and categorical data sequences (discrete time). In this talk, I shall first present some efficient numerical algorithms for solving Markovian queueing networks. I shall then introduce some other parsimonious high-order and multivariate Markov chain models for categorical sequences with applications. Efficient estimation methods for solving the model parameters will also be discussed. Practical problems and numerical examples will then be given to demonstrate the effectiveness of our proposed models.

Research Supported in Part by HK RGC Grant Nos. HKU 712602, 701707, and 17301214.

The Outline

(1) Motivations and Objectives.

(2) Queueing Systems.

(3) High-order Markov Chain Models.

(4) Multivariate Markov Chain Models.

(5) Concluding Remarks.

1. Motivations and Objectives.

• Markov chain models are popular tools for modeling many real world systems including

• **Continuous Time** Markov Chain Process
-Queueing Systems.
-Telecommunication Networks.
-Manufacturing Systems.
-Re-manufacturing Systems.

• **Discrete Time** Markov Chain Process
-Categorical Time Series.
-Financial and Credit Risk Models.
-Genetic Networks and DNA Sequencing.
-Inventory Prediction and Control.

2. Queueing Systems

- [**Queueing systems with Batch Arrival**] R. Chan and W. Ching, *Toeplitz-circulant Preconditioners for Toeplitz Systems and Their Applications to Queueing Networks with Batch Arrivals*, SIAM Journal of Scientific Computing, 1996.

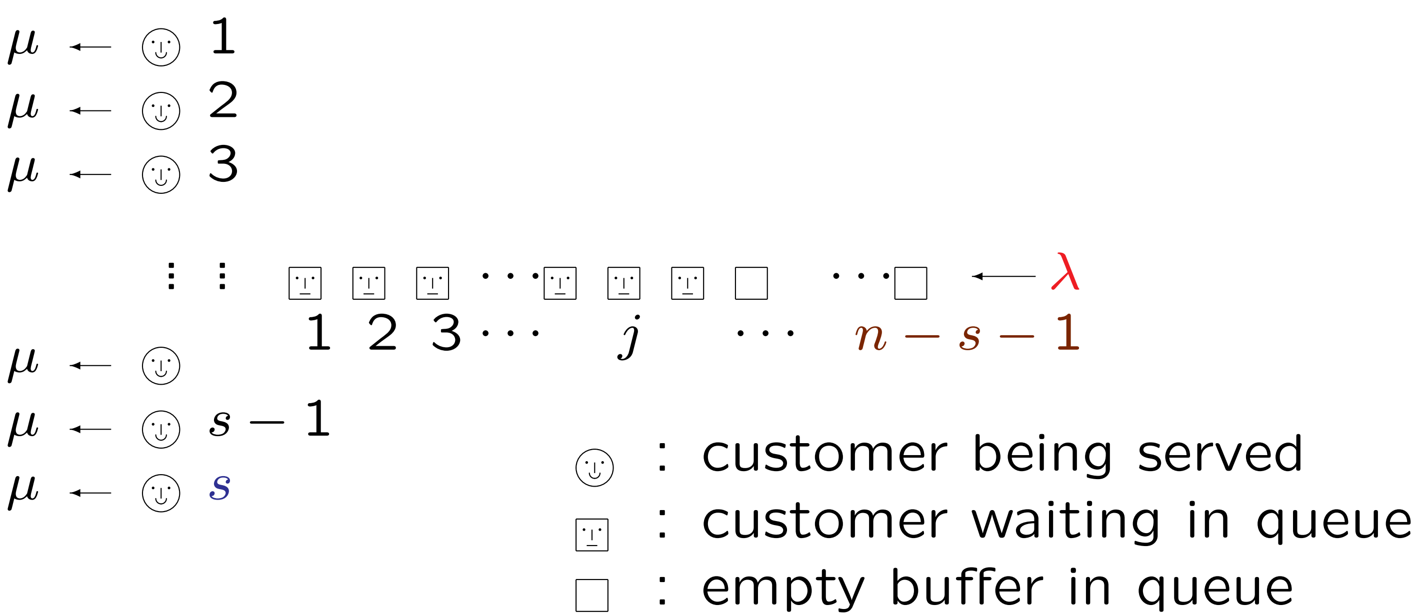
- [**Manufacturing Systems with Production Control**] W. Ching et al., *Circulant Preconditioners for Markov Modulated Poisson Processes and Their Applications to Manufacturing Systems*, SIAM Journal on Matrix Analysis and Its Applications, 1997.

- [**Queueing systems with Negative Customers**] W. Ching, *Iterative Methods for Queuing Systems with Batch Arrivals and Negative Customers*, BIT, 2003.

- [**Hybrid Method (Genetic Algorithm & SOR Iterative Method)**] W. Ching, *A Hybrid Algorithm for Queueing Systems*, CALCOLO, 2004.

2.1.1 Single Markovian Queue (M/M/s/n-s-1)

- λ input rate (Arrival Rate),
- μ output rate (Service Rate, Production Rate).
- s number of servers.
- $n-s-1$ number of waiting space.
- Set of possible states (number of customers): $\{0, 1, \ldots, n-1\}$.



2.1.2 *The Steady-state Distribution*

- Let p_i be the steady-state probability that i customers in the queueing system and $\mathbf{p} = (p_0, \ldots, p_{n-1})^t$ be the **steady-state probability vector**.

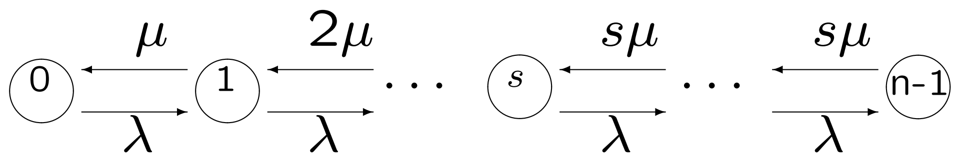
- Important for system performance analysis, e.g. average waiting time of the customers in long run.

- B. Bunday, *Introduction to Queueing Theory*, Arnold, N.Y., 1996.

- p_i governed by the **Kolmogorov equations**:

Out Going Rate **Incoming Rate**

$$p_{i-1} \xleftarrow[i\mu]{} p_i \xrightarrow{\lambda} p_{i+1} \qquad p_{i-1} \xrightarrow{\lambda} p_i \xleftarrow[(i+1)\mu]{} p_{i+1}$$



The Markov Chain of the M/M/s/n-s-1 Queue

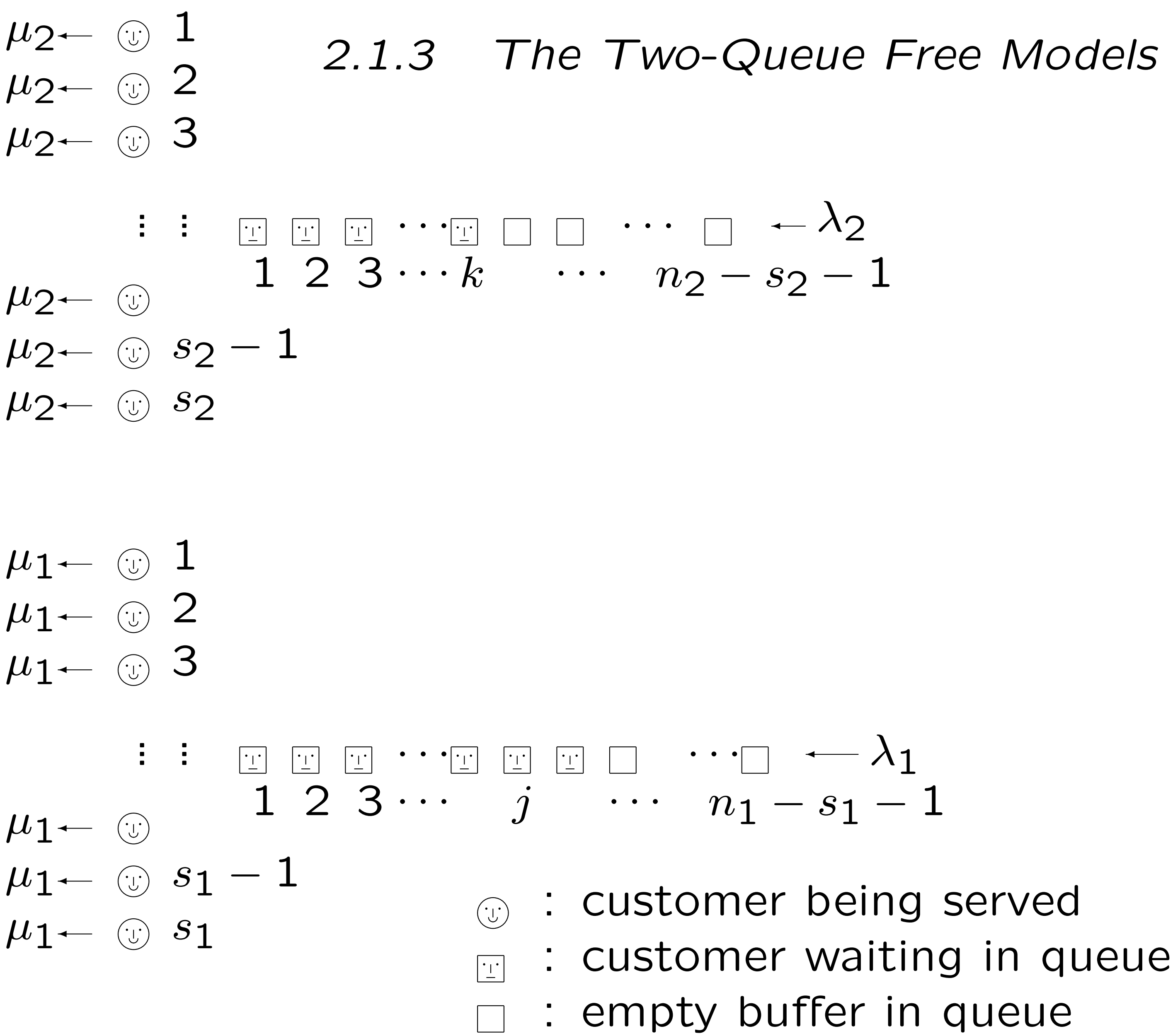
- We are solving:

$$\begin{cases} A_0 \mathbf{p_0} & = \mathbf{0}, \\ \sum p_i & = 1, \\ p_i & \geq 0. \end{cases}$$

- A_0, the **generator matrix**, is given by the $n \times n$ tridiagonal matrix:

$$A_0 = \begin{pmatrix} \lambda & -\mu & & & & & \\ -\lambda & \lambda+\mu & -2\mu & & & 0 & \\ & -\lambda & \lambda+2\mu & -3\mu & & & \\ & & \cdot & \cdot & \cdot & & \\ & & -\lambda & \lambda+s\mu & -s\mu & & \\ & & & \cdot & \cdot & \cdot & \\ & 0 & & & -\lambda & \lambda+s\mu & -s\mu \\ & & & & & -\lambda & s\mu \end{pmatrix}.$$

2.1.3 The Two-Queue Free Models



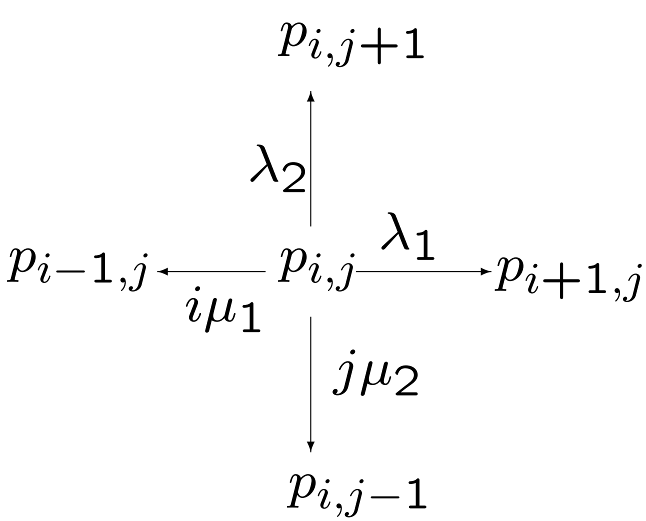
- Set of possible states (number of customers in each queue):

$$S = \{(i,j) : i,j \in \mathbf{N}, 0 \le i \le n_1 - 1, 0 \le j \le n_2 - 1\}.$$

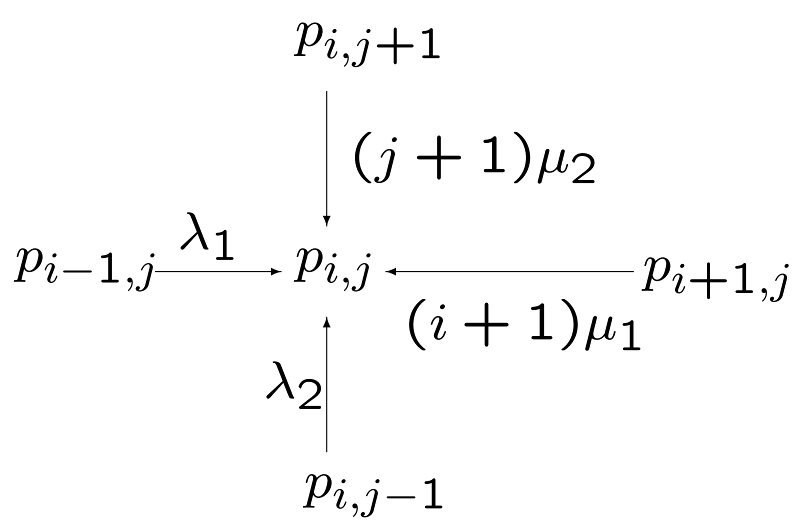
- Let $p_{i,j}$ be the probability that i customers in queue 1 and j customers in queue 2.

- The Kolmogorov equations for the two-queue network:

Out Going Rate



Incoming Rate



• Again we have to solve

$$\begin{cases} A_1\mathbf{p} & = \mathbf{0}, \\ \sum p_{ij} & = 1, \\ p_{ij} & \geq 0. \end{cases}$$

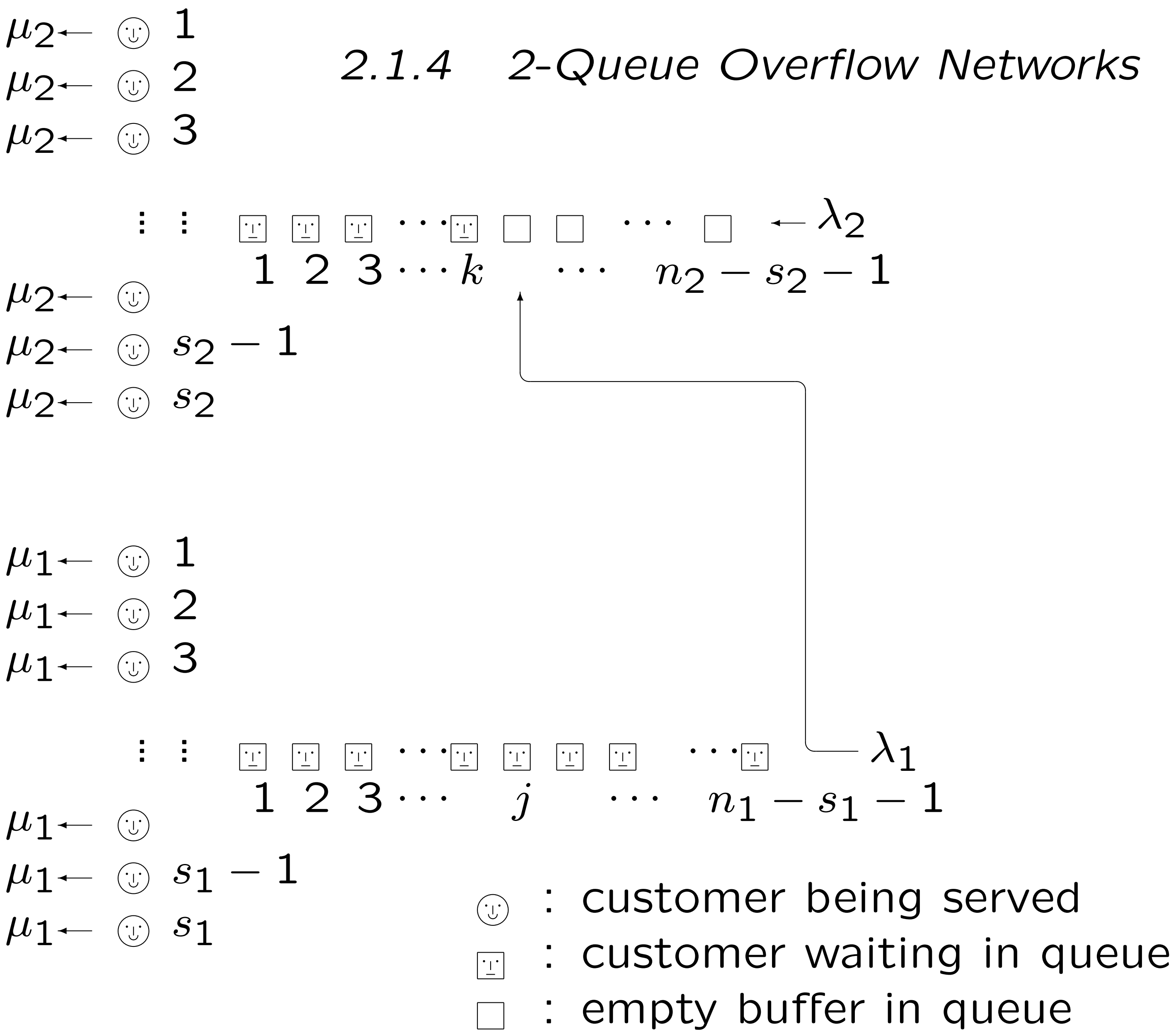
• The generator matrix A_1 is separable (no interaction between the queues): $A_1 = A_0 \otimes I + I \otimes A_0$.

• Kronecker tensor product of two matrices $C_{n\times r}$ and $B_{m\times k}$:

$$C_{n\times r} \otimes B_{m\times k} = \begin{pmatrix} c_{11}B & \cdots & \cdots & c_{1r}B \\ c_{21}B & \cdots & \cdots & c_{2r}B \\ \vdots & \vdots & \vdots & \vdots \\ c_{n1}B & \cdots & \cdots & c_{nr}B \end{pmatrix}_{nm\times rk} .$$

• It is easy to check that the Markov chain of the queueing system is irreducible and the unique solution is $\mathbf{p} = \mathbf{p_0} \otimes \mathbf{p_0}$.

2.1.4 2-Queue Overflow Networks



- L. Kaufman, *Matrix Methods for Queuing Problems*, SIAM J. Sci. Statist. Comput., 1983.

- The generator matrix A_2 is given by

$$A_2 = A_0 \otimes I + I \otimes A_0 + \begin{pmatrix} \mathbf{0} & \\ & 1 \end{pmatrix} \otimes R_0,$$

where

$$R_0 = \lambda_1 \begin{pmatrix} 1 & & & & & \\ -1 & 1 & & & 0 & \\ & \ddots & \ddots & & & \\ & & \ddots & \ddots & & \\ & 0 & & -1 & 1 & \\ & & & & -1 & 0 \end{pmatrix}$$

describes the overflow discipline of the queueing system.

- In fact, we may write

$$A_2 = A_1 + \begin{pmatrix} \mathbf{0} & \\ & 1 \end{pmatrix} \otimes R_0,$$

- Unfortunately **analytic solution** for the steady-state distribution $\mathbf{p}$ is **not** available.

- The generator matrices are sparse and have block structures.

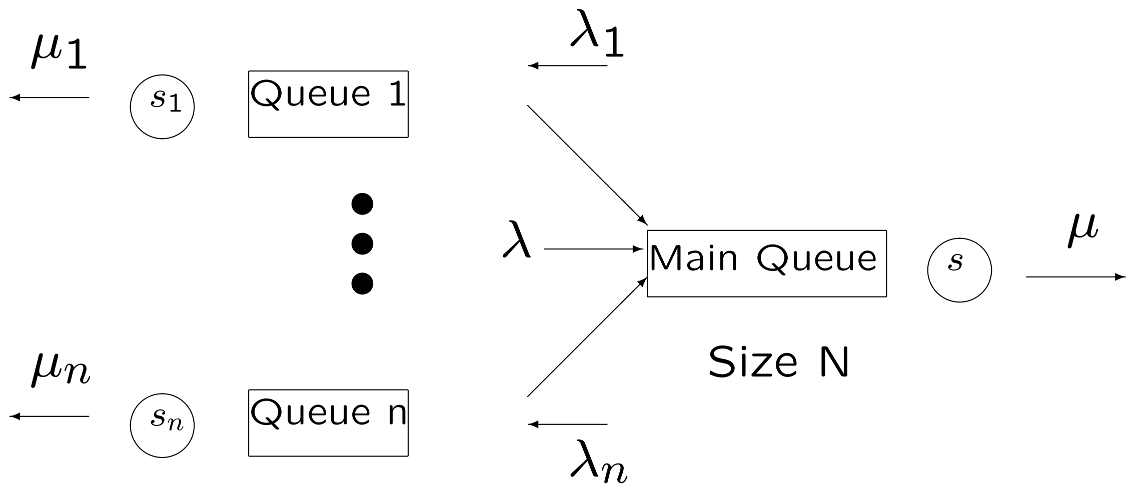
- Direct method (LU decomposition will result in dense matrices L and U) is not efficient in general.

- Fast algorithm should make use of the block structures and the sparsity of the generator matrices.

- Block Gauss-Seidel (BGS) is an usual approach for mentioned queueing problems. Its convergence rate is not fast and increase linearly with respect to the size of the generator matrix in general.

- R. Varga, *Matrix Iterative Analysis*, Prentice-Hall, N.J., 1963.

2.2 The Telecommunication System



- W. Ching, *Iterative Methods for Queuing and Manufacturing System*, Springer Monograph, 2001.

- W. Ching et al., *Circulant Preconditioners for Markov Modulated Poisson Processes and Their Applications to Manufacturing Systems*, SIAM Journal on Matrix Analysis and Its Applications, 1997.

- The generator matrix is given by:

$$A_3 = \begin{pmatrix} Q+\Gamma & -\mu I & & & & & 0 \\ -\Gamma & Q+\Gamma+\mu I & -2\mu I & & & & \\ & \ddots & \ddots & \ddots & & & \\ & & -\Gamma & Q+\Gamma+s\mu I & -s\mu I & & \\ & & \ddots & \ddots & \ddots & & \\ & & & -\Gamma & Q+\Gamma+s\mu I & -s\mu I \\ 0 & & & & -\Gamma & Q+s\mu I \end{pmatrix}$$

($(N+1)$-block by $(N+1)$-block), where

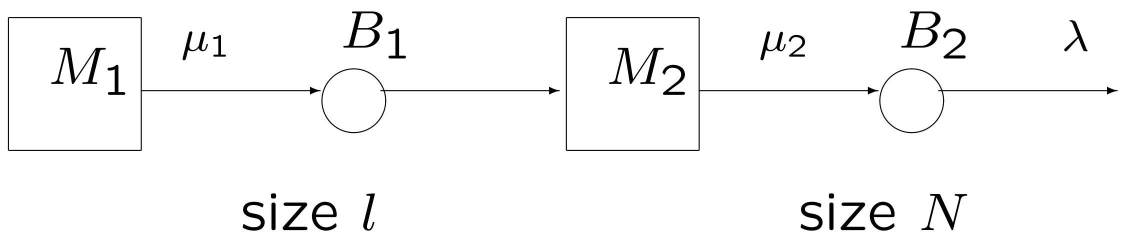
$$\Gamma = \Lambda + \lambda I_{2^n},$$

$$Q = (Q_1 \otimes I_2 \otimes \cdots \otimes I_2) + (I_2 \otimes Q_2 \otimes I_2 \otimes \cdots \otimes I_2) + \cdots + (I_2 \otimes \cdots \otimes I_2 \otimes Q_n),$$

$$\Lambda = (\Lambda_1 \otimes I_2 \otimes \cdots \otimes I_2) + (I_2 \otimes \Lambda_2 \otimes I_2 \otimes \cdots \otimes I_2) + \cdots + (I_2 \otimes \cdots \otimes I_2 \otimes \Lambda_n),$$

$$Q_j = \begin{pmatrix} \sigma_{j1} & -\sigma_{j2} \\ -\sigma_{j1} & \sigma_{j2} \end{pmatrix} \quad \text{and} \quad \Lambda_j = \begin{pmatrix} \lambda_j & 0 \\ 0 & 0 \end{pmatrix}.$$

2.3 The Manufacturing System of Two Machines in Tandem



• Search for optimal buffer sizes l and N ($N >> l$), which minimizes (1) the average running cost, (2) maximizes the throughput, or (3) minimizes the blocking and the starving rate.

• W. Ching, *Iterative Methods for Manufacturing Systems of Two Stations in Tandem*, Applied Mathematics Letters, 1998.

• W. Ching, *Markovian Approximation for Manufacturing Systems of Unreliable Machines in Tandem*, Naval Research Logistics, 2001.

The generator matrix is of the form:

$$A_4 = \begin{pmatrix} \Lambda + \mu_1 I & -\Sigma & & & 0 \\ -\mu_1 I & \Lambda + D + \mu_1 I & -\Sigma & & \\ & \ddots & \ddots & \ddots & \\ & & -\mu_1 I & \Lambda + D + \mu_1 I & -\Sigma \\ 0 & & & -\mu_1 I & \Lambda + D \end{pmatrix},$$

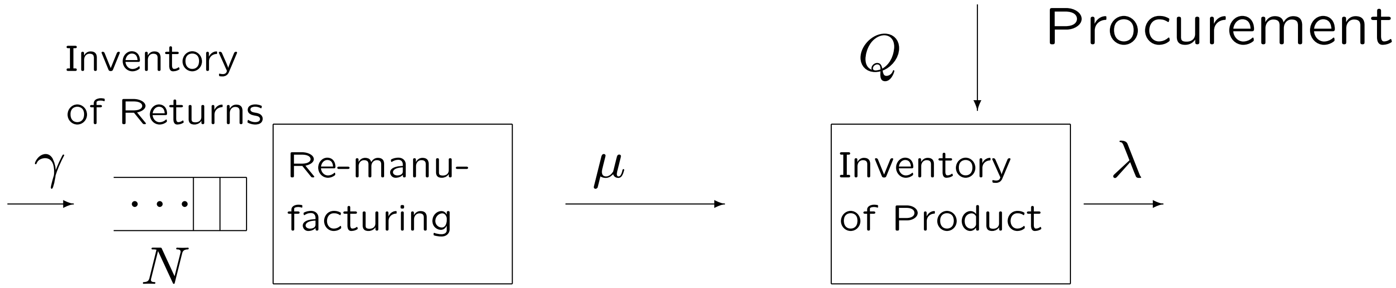
($(l+1)$-block by $(l+1)$-block), where

$$\Lambda = \begin{pmatrix} 0 & -\lambda & & 0 \\ & \lambda & \ddots & \\ & & \ddots & -\lambda \\ 0 & & & \lambda \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 0 & & & 0 \\ \mu_2 & \ddots & & \\ & \ddots & \ddots & \\ 0 & & \mu_2 & 0 \end{pmatrix},$$

and

$$D = \mathrm{Diag}(\mu_2, \cdots, \mu_2, 0).$$

2.4 The Re-Manufacturing System



• There are two types of inventory to manage: the serviceable product and the returned product. The re-cycling process is modelled by an M/M/1/N queue.

• The product inventory level and outside procurements are controlled by an (Q, r) continuous review policy. Here r is the outside procurement level and Q is the procurement quantity. We assume that $N >> Q$.

• W. Ching et al., *An Inventory Model with Returns and Lateral Transshipments*, Journal of the Operational Research Society, 2003.

• W. Ching et al., *A Direct Method for Solving Block-Toeplitz with Near-Circulant-Block Systems with Applications to Hybrid Manufacturing Systems*, Journal of Numerical Linear Algebra with Applications, 2005.

- The generator matrix is given by

$$A_5 = \begin{pmatrix} B & -\lambda I & & & 0 \\ -L & B & -\lambda I & & \\ & \ddots & \ddots & \ddots & \\ & & -L & B & -\lambda I \\ -\lambda I & & & -L & B_Q \end{pmatrix},$$

where

$$L = \begin{pmatrix} 0 & \mu & & & 0 \\ & 0 & \ddots & & \\ & & \ddots & \ddots & \\ & & & \ddots & \mu \\ 0 & & & & 0 \end{pmatrix}, \quad B = \lambda I_{N+1} + \begin{pmatrix} \gamma & & & & 0 \\ -\gamma & \gamma+\mu & & & \\ & \ddots & \ddots & & \\ & & \ddots & \gamma+\mu & \\ 0 & & & -\gamma & \mu \end{pmatrix},$$

and

$$B_Q = B - \mathrm{Diag}(0, \mu, \ldots, \mu).$$

2.5 Numerical Algorithm (Preconditioned Conjugate Gradient (PCG) Method)

- Conjugate Gradient (CG) Method to solve $A\mathbf{x} = \mathbf{b}$.

- Need preconditioning to accelerate convergence rate.

- O. Axelsson, *Iterative Solution Methods*, Cambridge University Press, 1996.

- In the Preconditioned Conjugate Gradient (PCG) method with preconditioner C, CG method is applied to solve

$$C^{-1}A\mathbf{x} = C^{-1}\mathbf{b}$$

instead of

$$A\mathbf{x} = \mathbf{b}.$$

- A **Good preconditioner** C is a matrix satisfying:

(a) Easy and fast to construct;

(b) The preconditioner system $C\mathbf{x} = \mathbf{f}$ can be solved very fast;

(c) The preconditioned matrix $C^{-1}A$ has singular values clustered around 1[1].

Note:
(1) One sufficient condition for a sequence of matrices B_n (size $n \times n$) has singular values clustered around 1 : the number of singular values of B_n different from 1 is bounded above and independent of the matrix size of B_n.

2.5.1 Circulant-based Preconditioners

- Circulant matrices are Toeplitz matrices (constant diagonal entries) such that each column is a cyclic shift of its preceding column. It is characterized by its first column.

- The class of circulant matrices is denoted by $\mathcal{F}$.

– $C \in \mathcal{F}$ implies C can be diagonalized by Fourier matrix F:

$$C = F^* \Lambda F.$$

Hence

$$C^{-1}\mathbf{x} = F^* \Lambda^{-1} F\mathbf{x}.$$

– Eigenvalues of a circulant matrix has analytic form, therefore enhance the spectrum analysis of the preconditioned matrix.

– $C^{-1}\mathbf{x}$ can be done in $O(n \log n)$.

- P. Davis, *Circulant Matrices*, John Wiley and Sons, N.J. 1985.

2.5.2 Circulant-based Preconditioner

The idea of circulant approximation:

$$A = \begin{pmatrix} \lambda & -\mu & & & & & 0 \\ -\lambda & \lambda+\mu & -2\mu & & & & \\ & \cdot & \cdot & \cdot & & & \\ & & -\lambda & \lambda+s\mu & -s\mu & & \\ & & & \cdot & \cdot & \cdot & \\ & & & & -\lambda & \lambda+s\mu & -s\mu \\ 0 & & & & & -\lambda & s\mu \end{pmatrix}.$$

$$s(A) = \begin{pmatrix} \lambda+s\mu & \color{red}{-s\mu} & & & & & \color{red}{-\lambda} \\ -\lambda & \lambda+s\mu & \color{red}{-s\mu} & & & & \\ & \cdot & \cdot & \cdot & & & \\ & & -\lambda & \lambda+s\mu & -s\mu & & \\ & & & \cdot & \cdot & \cdot & \\ & & & & -\lambda & \lambda+s\mu & -s\mu \\ \color{red}{-s\mu} & & & & & -\lambda & \lambda+s\mu \end{pmatrix}.$$

We have $\text{rank}(A - s(A)) = s + 1$.

• $A_3 = I \otimes Q + A \otimes I + R \otimes \Lambda$, where

$$R = \begin{pmatrix} 1 & & & & 0 \\ -1 & 1 & & & \\ & -1 & \ddots & & \\ & & \ddots & 1 & \\ 0 & & & -1 & 0 \end{pmatrix}.$$

• $s(A_3) = s(I) \otimes Q + s(A) \otimes I + s(R) \otimes \Lambda$, where

$$s(I) = I \quad \text{and} \quad s(R) = \begin{pmatrix} 1 & & & & -1 \\ -1 & 1 & & & \\ & -1 & \ddots & & \\ & & \ddots & 1 & \\ 0 & & & -1 & 1 \end{pmatrix}.$$

2.5.4 *The Manufacturing System of Two Machines in Tandem*

- Circulant-based approximation of A_4 : $s(A_4) =$

$$
\begin{pmatrix}
s(\Lambda)+\mu_1 I & -s(\Sigma) & & & 0 \\
-\mu_1 I & s(\Lambda)+s(D)+\mu_1 I & -s(\Sigma) & & \\
 & \ddots & \ddots & \ddots & \\
 & & -\mu_1 I & s(\Lambda)+s(D)+\mu_1 I & -s(\Sigma) \\
0 & & & -\mu_1 I & s(\Lambda)+s(D)
\end{pmatrix},
$$

($(l+1)$-block by $(l+1)$-block), where

$$
s(\Lambda) = \begin{pmatrix}
\lambda & -\lambda & & 0 \\
 & \lambda & \ddots & \\
 & & \ddots & -\lambda \\
-\lambda & & & \lambda
\end{pmatrix}, \quad
s(\Sigma) = \begin{pmatrix}
0 & & & \mu_2 \\
\mu_2 & \ddots & & \\
 & \ddots & \ddots & \\
0 & & \mu_2 & 0
\end{pmatrix},
$$

and

$$
s(D) = \mathrm{Diag}(\mu_2, \cdots, \mu_2, \mu_2).
$$

- circulant-based approximation of A_5:

$$s(A_5) = \begin{pmatrix} s(B) & -\lambda I & & & 0 \\ -s(L) & s(B) & -\lambda I & & \\ & \ddots & \ddots & \ddots & \\ & & -s(L) & s(B) & -\lambda I \\ -\lambda I & & & -s(L) & s(B_Q) \end{pmatrix},$$

where

$$s(L) = \begin{pmatrix} 0 & \mu & & & 0 \\ & 0 & \ddots & & \\ & & \ddots & \ddots & \\ & & & \ddots & \mu \\ \mu & & & & 0 \end{pmatrix}, \quad s(B) = \lambda I_{N+1} + \begin{pmatrix} \gamma+\mu & & & -\gamma \\ -\gamma & \gamma+\mu & & \\ & \ddots & \ddots & \\ 0 & & -\gamma & \gamma+\mu \end{pmatrix},$$

and

$$s(B_Q) = s(B) - \mu I.$$

- In fact, all the generator matrices A take the following form:

$$A = \sum_{i=1}^{m} \bigotimes_{j=1}^{n} A_{ij},$$

where A_{i1} is relatively huge in size.

- Our preconditioner is defined as

$$C = \sum_{i=1}^{m} s(A_{i1}) \bigotimes_{j=2}^{n} A_{ij}.$$

- We note that

$$\left[F \bigotimes_{j=2}^{n} I\right]^{*} C \left[F \bigotimes_{j=2}^{n} I\right] = \sum_{i=1}^{m} \Lambda_{i1} \bigotimes_{j=2}^{n} A_{ij} = \bigoplus_{k=1}^{\ell} \left[\sum_{i=1}^{m} \lambda_{i1}^{k} \bigotimes_{j=2}^{n} A_{ij}\right]$$

which is a block-diagonal matrix.

• One of the advantages of our preconditioner is that it can be inverted in parallel by using a parallel computer easily. This would therefore save a lot of computational cost.

• Theorem: If all the parameters stay fixed then the preconditioned matrix has singular values clustered around one. Thus we expect our PCG method converges very fast.

• $A_{i1} \approx$ Toeplitz except for rank $(s+1)$ perturbation
$\approx s(A_{i1})$ except for rank $(s+1)$ perturbation.

• R. Chan and W. Ching, *Circulant Preconditioners for Stochastic Automata Networks*, Numerise Mathematik, (2000).

2.6 **Numerical Results**

• Since generator A is non-symmetric, we used the generalized CG method, the Conjugate Gradient Squared (CGS) method. This method does not require the multiplication of $A^T\mathbf{x}$.

• Our proposed method is applied to the following systems.

(1) The Telecomunications System.

(2) The Manufacturing Systems of Two Machines in Tandem.

(3) The Re-Manufacturing System.

• P. Sonneveld, *A Fast Lanczos-type Solver for Non-symmetric Linear Systems*, SIAM J. Sci. Comput., 1989.

• Stopping Criteria: $\frac{||\mathbf{r}_n||_2}{||\mathbf{r}_0||_2} < 10^{-10}$; $||\mathbf{r}_n||_2 = n$th step residual.

2.6.1 The Telecomunications System

- n, number of external queues; N, size of the main queue.
- Cost per Iteration:

I	C	BGS
$O(n2^nN)$	$O(n2^nN\log N)$	$O((2^n)^2N)$

- Number of Iterations:

$s=2$	$n=1$			$n=4$		
N	I	C	BGS	I	C	BGS
32	155	8	171	161	13	110
64	**	7	242	**	13	199
128	**	8	366	**	14	317
256	**	8	601	**	14	530
512	**	8	**	**	14	958

- '**' means greater than 1000 iterations.

2.6.2 The Manufacturing Systems of Two Machines in Tandem

- l, size of the first buffer; N, size of the second buffer.

- Cost per Iteration:

I	C	BGS
$O(lN)$	$O(lN \log N)$	$O(lN)$

- Number of Iterations:

	$l = 1$			$l = 4$		
N	I	C	BGS	I	C	BGS
32	34	5	72	64	10	72
64	129	7	142	139	11	142
128	**	8	345	**	12	401
256	**	8	645	**	12	**
1024	**	8	**	**	12	**

- '**' means greater than 1000 iterations.

2.6.3 The Re-Manufacturing System

- Q, size of the serviceable inventory; N, size of the return inventory.

- Cost per iteration:

I	C	BGS
$O(QN)$	$O(QN \log N)$	$O(QN)$

- Number of Iterations:

	$Q=2$			$Q=3$			$Q=4$		
N	I	C	BGS	I	C	BGS	I	C	BGS
100	246	8	870	**	14	1153	**	19	1997
200	**	10	1359	**	14	**	**	19	**
400	**	10	**	**	14	**	**	19	**
800	**	10	**	**	14	**	**	19	**

- '**' means greater than 2000 iterations.

2.7 **Concluding Remarks**

• We proposed circulant-based preconditioners in conjunction with CG type methods for solving Markovian queueing systems. The approach has the following advantages for hugh size problems:

– The preconditioner can be inverted efficiently via FFT and also in parallel.

– The proposed method is a matrix-free method. Since the generator matrix is very sparse, and in each CG iteration, the major cost is the matrix-vector multiplications. We don't need to store the whole matrix but the values and the positions of those non-zero entries.

3. High-order Markov Chain Models.

• Markov chains are popular models for solving many practical systems including **categorical time series**.

• It is also very easy to construct a Markov chain model. Given the observed time series data sequence (Markov chain) $\{X_t\}$ of m **states**, one can count the **transition frequency** F_{jk} (one step) in the observed sequence from State k to State j in one step.

• Hence one can construct the **one-step transition frequency matrix** for the observed sequence $\{X_t\}$ as follows:

$$F = \begin{pmatrix} F_{11} & \cdots & \cdots & F_{1m} \\ F_{21} & \cdots & \cdots & F_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ F_{m1} & \cdots & \cdots & F_{mm} \end{pmatrix}. \tag{1}$$

• From F, one can get the estimates for P_{kj} (**column normalization**) as follows:

$$P = \begin{pmatrix} P_{11} & \cdots & \cdots & P_{1m} \\ P_{21} & \cdots & \cdots & P_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ P_{m1} & \cdots & \cdots & P_{mm} \end{pmatrix} \tag{2}$$

where

$$\boxed{P_{kj} = \frac{F_{kj}}{\sum_{r=1}^{m} F_{rj}}}$$

is the **maximum likelihood estimator**.

• Example 1: Consider a data sequence of **3 states**/**categories**:

$$\{1, 1, 2, 2, 1, 3, 2, 1, 2, 3, 1, 2, 3, 1, 2, 3, 1, 2, 1, 2\}. \quad (3)$$

We adopt the following **notation**. The sequence $\{X_t\}$ can be written in **vector form** (canonical representation by **Robert J. Elliott**):

$\mathbf{X}_0 = (1, 0, 0)^T$, $\mathbf{X}_1 = (1, 0, 0)^T$, $\mathbf{X}_2 = (0, 1, 0)^T$, $\ldots$, $\mathbf{X}_{19} = (0, 1, 0)^T$.

• The **First-order** Markov Chain Model:
By counting the transition frequency from State k to State j in the sequence, one can construct the transition frequency matrix F (then the transition probability matrix $\hat{Q}$) for the sequence.

$$F = \begin{pmatrix} 1 & 3 & 3 \\ 6 & 1 & 1 \\ 1 & 3 & 0 \end{pmatrix} \quad \text{and} \quad \hat{Q} = \begin{pmatrix} 1/8 & 3/7 & 3/4 \\ 6/8 & 1/7 & 1/4 \\ 1/8 & 3/7 & 0 \end{pmatrix} \quad (4)$$

and the first-order Markov chain model is

$$\boxed{\mathbf{X}_{t+1} = \hat{Q}\mathbf{X}_t.}$$

Robert J. Elliott (University of Calgary).
Hidden Markov models and financial engineering.
Ph.D. and Sc.D. (Cambridge University)

3.1 The Steady-state Probability Distribution

Proposition 3.1: Given an **irreducible** and **aperiodic** Markov chain of m states, then for any initial probability distribution $\mathbf{X}_0$

$$\lim_{t \to \infty} ||\mathbf{X}_t - \boldsymbol{\pi}|| = \lim_{t \to \infty} ||P^t \mathbf{X}_0 - \boldsymbol{\pi}|| = 0.$$

where $\boldsymbol{\pi}$ is the **steady-state probability distribution** of the transition probability matrix P of the underlying Markov chain and $\boldsymbol{\pi} = P\boldsymbol{\pi}$.

• We remark that a non-negative probability vector $\boldsymbol{\pi}$ satisfies $\boldsymbol{\pi} = P\boldsymbol{\pi}$ is called a **stationary probability vector**. A stationary probability vector is not necessary the **steady-state probability distribution vector**.

Proposition 3.2: [Perron (1907) - Frobenius (1912) Theorem] Let A be a **non-negative**, **irreducible** and **aperiodic** square matrix of size m. Then

(i) A has a positive real eigenvalue λ, equal to its spectral radius,

$$\lambda = \max_{1 \leq k \leq m} |\lambda_k(A)|$$

where $\lambda_k(A)$ denotes the kth eigenvalue of A.

(ii) There corresponds an eigenvector $\mathbf{z}$, its entries being **real and positive**, such that $A\mathbf{z} = \lambda\mathbf{z}$.

(iii) The eigenvalue λ is a **simple eigenvalue** of A.

This theorem is important in Markov chains and dynamical systems.

• We remark that requirement **aperiodic** is important. The following matrix is non-negative and irreducible but not aperiodic:

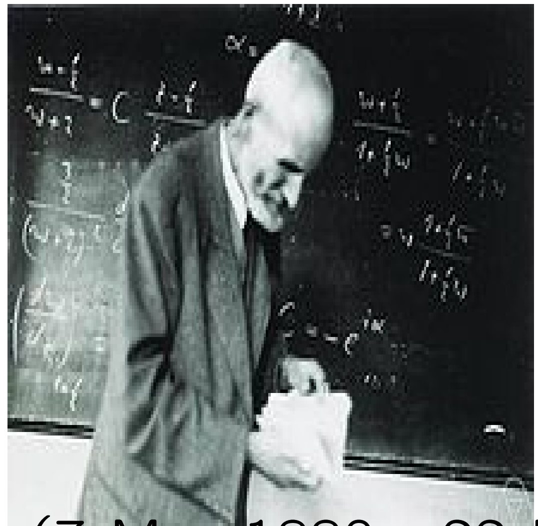
$$Q = \begin{pmatrix} 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & \ddots & & 0 \\ 0 & 1 & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix} \tag{5}$$

Now we see that all eigenvalues of Q are given by

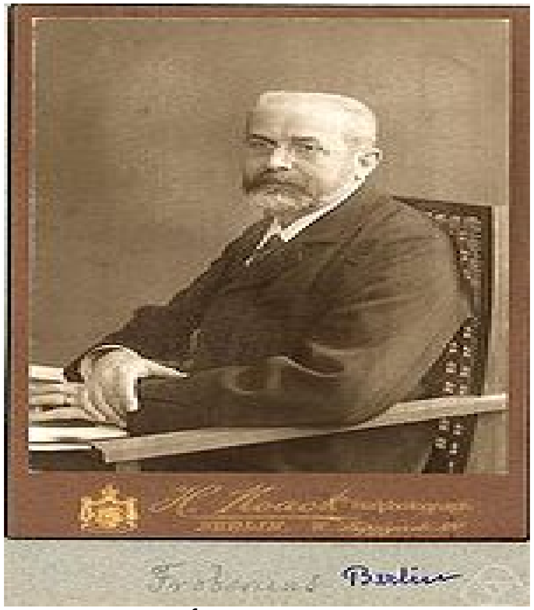
$$\lambda_j = e^{2\pi ij/n} \quad j = 0, 1, \ldots, n-1$$

and they all satisfy $|\lambda_j| = 1$. They are all on the unit circle. We see that $\boldsymbol{\pi} = (1, 1, \cdots, 1)^T$ is the **stationary probability vector** but it is not the **steady-state distribution vector**. In fact, begin with $\mathbf{x}_0 = (1, 0, 0, \cdots, 0)^T$, $\mathbf{x}_t$ does not converge to $\boldsymbol{\pi}$. But we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i = \boldsymbol{\pi}.$$



Oskar Perron (7 May 1880 - 22 February 1975).
(Taken from Wikipedia, the free encyclopedia)
Differential equations and partial differential equations.



Ferdinand Georg Frobenius (October 26, 1849 - August 3, 1917).

(Taken from Wikipedia, the free encyclopedia)

Differential equations and group theory.

He gave the first full proof for the **Cayley-Hamilton theorem**.

3.2 Motivations for High-order Markov Chain Models.

• **Categorical data sequences** occur frequently in many real world applications. The delay effects occur in many applications and data.

(i) **Inventory Control and Demand Forecasting.** W. Ching et al., *A Higher-order Markov Model for the Newsboy's Problem*, Journal of Operational Research Society, 54 (2003) 291-298.

(ii) **Genetic Networks and DNA Sequencing.** W. Ching et al., *Higher-order Markov Chain Models for Categorical Data Sequences*, Naval Research Logistics, 51 (2004) 557-574.

(iii) **Financial and Credit Risk Models.** T. Siu, W. Ching et al., *A Higher-Order Markov-Switching Model for Risk Measurement*, Computers & Mathematics with Applications, 58 (2009) 1–10.

3.3 High-order Markov Chain Models.

• High-order Markov chain can better model categorical data sequence (to capture the delay effect).

• Problem: A conventional n-th order Markov chain of m states has $O(m^n)$ states and therefore parameters. The number of transition probabilities (to be estimated) increases **exponentially** with respect to the order n of the model.

(i) P. Jacobs and P. Lewis, *Discrete Time Series Generated by Mixtures I : Correlational and Runs Properties*, J. R. Statist. Soc. B, 40 (1978) 94–105.

(ii) G. Pegram, *An Autoregressive Model for Multilag Markov Chains*, J. Appl. Prob., 17 (1980) 350–362.

(iii) **A.E. Raftery**, *A Model for High-order Markov Chains*, J. R. Statist. Soc. B, 47 (1985) 528–539.

Adrian E. Raftery, University of Washington.

Statistical methods for social, environmental and health sciences.

C. Clogg Award (1998) and P. Lazarsfeld Award (2003)

World's most cited researcher in mathematics for the decade 1995-2005

- Raftery proposed a **high-order Markov chain model** which involves only one additional parameter for **each extra lag**.

- The model:

$$\boxed{P(X_t = k \mid X_{t-1} = k_1, \ldots, X_{t-n} = k_n) = \sum_{i=1}^{n} \lambda_i q_{kk_i}} \tag{6}$$

where $k, k_1, \ldots, k_n \in M$. Here $M = \{1, 2, \ldots, m\}$ is the set of the possible states and

$$\sum_{i=1}^{n} \lambda_i = 1 \quad \text{and} \quad Q = [q_{ij}]$$

is a transition probability matrix with column sums equal to one, such that

$$0 \le \sum_{i=1}^{n} \lambda_i q_{kk_i} \le 1, \quad k, k_1, \ldots, k_n = 1, 2, \ldots, m. \tag{7}$$

• Raftery proved that Model (6) is analogous to the **standard AR(n) model** in time series.

• The parameters $q_{k_0 k_i}, \lambda_i$ can be estimated numerically by **maximizing the log-likelihood** of Model (6) subjected to the constraints (7).

• Problems:

(i) This approach involves solving a highly non-linear optimization problem (which is not easy to solve).

(ii) The proposed numerical method neither guarantees *convergence* nor a *global maximum*.

3.4 Our High-order Markov Chain Model.

- Raftery's model can be generalized as follows:

$$\boxed{\mathbf{X}_{t+n+1} = \sum_{i=1}^{n} \lambda_i Q_i \mathbf{X}_{t+n+1-i}.} \tag{8}$$

- We define Q_i to be the i**th step transition probability matrix** of the sequence.

- We also assume that λ_i are **non-negative** such that

$$\sum_{i=1}^{n} \lambda_i = 1$$

so that the right-hand-side of Eq. (8) is a probability distribution.

3.5 A Property of the Model.

Proposition 3.1: If Q_i is **irreducible**, $\lambda_i > 0$ for $i = 1, 2, \ldots, n$ and

$$\sum_{i=1}^{n} \lambda_i = 1$$

then the model in Eq. (8) has a **stationary distribution** $\bar{\mathbf{X}}$ when $t \to \infty$ independent of the initial state vectors

$$\mathbf{X}_0, \mathbf{X}_1, \ldots, \mathbf{X}_{n-1}.$$

The proof is based on Perron-Frobenius Theorem.

• The stationary distribution $\bar{\mathbf{X}}$ is the unique solution of the linear system of equations

$$\boxed{(I - \sum_{i=1}^{n} \lambda_i Q_i)\bar{\mathbf{X}} = \mathbf{0} \quad \text{and} \quad \mathbf{1}^T \bar{\mathbf{X}} = 1} \qquad (9)$$

where I is the m-by-m identity matrix (m is the number of possible states taken by each data point).

3.6 Parameter Estimation.

- Estimation of Q_i, the ith step transition probability matrix. One can count the transition frequency $f_{jk}^{(i)}$ in the sequence from State k to State j in the ith step. We get

$$F^{(i)} = \begin{pmatrix} f_{11}^{(i)} & \cdots & \cdots & f_{m1}^{(i)} \\ f_{12}^{(i)} & \cdots & \cdots & f_{m2}^{(i)} \\ \vdots & \vdots & \vdots & \vdots \\ f_{1m}^{(i)} & \cdots & \cdots & f_{mm}^{(i)} \end{pmatrix}, \quad \text{for} \quad i = 1, 2, \ldots, n. \qquad (10)$$

- From $F^{(i)}$, we get by **column normalization**:

$$\hat{Q}_i = \begin{pmatrix} \hat{q}_{11}^{(i)} & \cdots & \cdots & \hat{q}_{m1}^{(i)} \\ \hat{q}_{12}^{(i)} & \cdots & \cdots & \hat{q}_{m2}^{(i)} \\ \vdots & \vdots & \vdots & \vdots \\ \hat{q}_{1m}^{(i)} & \cdots & \cdots & \hat{q}_{mm}^{(i)} \end{pmatrix} \quad \text{where} \quad \hat{q}_{kj}^{(i)} = \frac{f_{kj}^{(i)}}{\sum_{r=1}^{m} f_{rj}^{(i)}} \qquad (11)$$

- Linear Programming Formulation for the Estimation of λ_i.

Note: Proposition 3.1 gives a sufficient condition for the sequence $\mathbf{X}_t$ to converge to a stationary distribution $\mathbf{X}$.

- We assume $\mathbf{X}_t \to \bar{\mathbf{X}}$ as $t \to \infty$.

- $\bar{\mathbf{X}}$ can be estimated from the sequence $\{X_t\}$ by computing the proportion of the occurrence of each state in the sequence and let us denote it by $\hat{\mathbf{X}}$.

- From Eq. (9) one would expect that

$$\boxed{\sum_{i=1}^{n} \lambda_i \hat{Q}_i \hat{\mathbf{X}} \approx \hat{\mathbf{X}}.} \qquad (12)$$

• This suggests one possible way to estimate the parameters

$$\lambda = (\lambda_1, \ldots, \lambda_n).$$

We consider the following optimization problem:

$$\min_{\lambda} \left\| \sum_{i=1}^{n} \lambda_i \widehat{Q}_i \hat{\mathbf{X}} - \hat{\mathbf{X}} \right\|_{\infty} = \min_{\lambda} \max_{k} \left| \left[\sum_{i=1}^{n} \lambda_i \widehat{Q}_i \hat{\mathbf{X}} - \hat{\mathbf{X}} \right]_k \right|$$

subject to

$$\sum_{i=1}^{n} \lambda_i = 1, \quad \text{and} \quad \lambda_i \geq 0, \quad \forall i.$$

Here $[\cdot]_k$ denotes the kth entry of the vector.

Remark: The optimization problem can be re-formulated as an **Linear Programming** (LP) problem. Thus it can be solved by using **EXCEL** for instance.

• W. Ching et al., *Building Higher-order Markov Chain Models with EXCEL*, International Journal of Mathematical Education in Science and Engineering, 35, 2004.

Remark: Other norms such as $||.||_2$ and $||.||_1$ can also be considered. The former will result in a **quadratic programming problem** while $||.||_1$ will still result in a **linear programming problem**.

• It is known that in approximating data by a linear function $||.||_1$ gives the **most robust answer**.

• $||.||_\infty$ **avoids gross discrepancies** with the data as much as possible.

• If the errors are known to be **normally distributed** then $||.||_2$ is the best choice.

- The linear programming formulation:

$$\min_{\lambda} \; w$$

subject to

$$\begin{pmatrix} w \\ w \\ \vdots \\ w \end{pmatrix} \geq \hat{\mathbf{X}} - \left[\widehat{Q}_1\hat{\mathbf{X}} \mid \widehat{Q}_2\hat{\mathbf{X}} \mid \; \cdots \; \mid \widehat{Q}_n\hat{\mathbf{X}}\right] \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{pmatrix},$$

$$\begin{pmatrix} w \\ w \\ \vdots \\ w \end{pmatrix} \geq -\hat{\mathbf{X}} + \left[\widehat{Q}_1\hat{\mathbf{X}} \mid \widehat{Q}_2\hat{\mathbf{X}} \mid \; \cdots \; \mid \widehat{Q}_n\hat{\mathbf{X}}\right] \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{pmatrix},$$

$$w \geq 0, \quad \sum_{i=1}^{n} \lambda_i = 1, \quad \text{and} \quad \lambda_i \geq 0, \quad \forall i.$$

- V. Chvatal, *Linear Programming*, Freeman, New York, 1983.

• For a numerical demonstration, we refer to Example 1. A **2nd-order** ($n = 2$) model for the **3-states** ($m = 3$) categorical sequence.

We have the transition frequency matrices

$$F^{(1)} = \begin{pmatrix} 1 & 3 & 3 \\ 6 & 1 & 1 \\ 1 & 3 & 0 \end{pmatrix} \quad \text{and} \quad F^{(2)} = \begin{pmatrix} 1 & 4 & 1 \\ 2 & 2 & 3 \\ 3 & 1 & 0 \end{pmatrix}. \tag{13}$$

From (13) we have the i-step transition probability matrices:

$$\hat{Q}_1 = \begin{pmatrix} 1/8 & 3/7 & 3/4 \\ 3/4 & 1/7 & 1/4 \\ 1/8 & 3/7 & 0 \end{pmatrix} \quad \text{and} \quad \hat{Q}_2 = \begin{pmatrix} 1/6 & 4/7 & 1/4 \\ 1/3 & 2/7 & 3/4 \\ 1/2 & 1/7 & 0 \end{pmatrix} \tag{14}$$

and the stationary distribution

$$\hat{\mathbf{X}} = (\frac{2}{5}, \frac{2}{5}, \frac{1}{5})^T.$$

Hence we have

$$\hat{Q}_1\hat{\mathbf{X}} = (\frac{13}{35}, \frac{57}{140}, \frac{31}{140})^T, \quad \text{and} \quad \hat{Q}_2\hat{\mathbf{X}} = (\frac{29}{84}, \frac{167}{420}, \frac{9}{35})^T.$$

• To estimate λ_i we consider the optimization problem:

$$\min_{\lambda_1,\lambda_2} w$$

subject to

$$\begin{cases} w \geq \frac{2}{5} - \frac{13}{35}\lambda_1 - \frac{29}{84}\lambda_2 \\ w \geq -\frac{2}{5} + \frac{13}{35}\lambda_1 + \frac{29}{84}\lambda_2 \\ w \geq \frac{2}{5} - \frac{57}{140}\lambda_1 - \frac{167}{420}\lambda_2 \\ w \geq -\frac{2}{5} + \frac{57}{140}\lambda_1 + \frac{167}{420}\lambda_2 \\ w \geq \frac{1}{5} - \frac{31}{140}\lambda_1 - \frac{9}{35}\lambda_2 \\ w \geq -\frac{1}{5} + \frac{31}{140}\lambda_1 + \frac{9}{35}\lambda_2 \\ w \geq 0, \quad \lambda_1 + \lambda_2 = 1, \quad \lambda_1, \lambda_2 \geq 0. \end{cases}$$

• The **optimal solution** is $(\lambda_1^*, \lambda_2^*, w^*) = (1, 0, 0.0286)$. Since λ_2 is zero, the “optimal” model is a first-order model.

Proposition 3.2 If Q_n is irreducible and aperiodic, $\lambda_1, \lambda_n > 0$ and

$$\sum_{i=1}^{n} \lambda_i = 1$$

then the model has a stationary distribution $\mathbf{X}$ satisfying

$$(I - \sum_{i=1}^{n} \lambda_i Q_i)\bar{\mathbf{X}} = \mathbf{0} \quad \text{and} \quad \mathbf{1}^T \bar{\mathbf{X}} = 1$$

and

$$\lim_{t \to \infty} \mathbf{X}_t = \mathbf{X}.$$

• We remark that if $\lambda_n = 0$ then it is not an nth model and if $\lambda_1 = 0$ then the model is clearly reducible.

3.7 The Newsboy's Problem.

• A newsboy sells newspaper (**perishable product**) every morning. The cost of each newspaper remains at the end of the day is C_o (**overage cost**) and the cost of each unsatisfied demand is C_s (**shortage cost**).

• Suppose that the (stationary distribution) probability density function of the **demand** D is given by

$$\boxed{\text{Prob } (D = d) = p_d \geq 0, \quad d = 1, 2, \ldots, m.} \tag{15}$$

• To determine the best amount r^* (order size) of newspaper to be ordered such that the **expected cost** is minimized.

Proposition 3.3: The optimal order size r^* is the one which satisfies

$$\boxed{F(r^* - 1) < \frac{C_s}{C_s + C_o} \leq F(r^*).} \tag{16}$$

Here $F(x) = \sum_{i=1}^{x} p_i$.

Example 2: Suppose that the demand $(1, 2, \ldots, 2k)$ $(m = 2k)$ follows an Markov process with the transition probability matrix Q of size $2k \times 2k$ given by

$$Q = \begin{pmatrix} 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & \ddots & & 0 \\ 0 & 1 & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix} \quad (17)$$

• Assume that $C_o = C_s$. Clearly the next "demand" can be determined by the state of the current demand, and hence the optimal expected cost is equal to **zero** when the **first-order Markov model** is used.

• When the classical Newsboy model is used, we note that the stationary distribution of Q is given by

$$\frac{1}{2k}(1, 1, \ldots, 1)^T.$$

The optimal ordering size is equal to k by Proposition 2.2 and therefore the optimal expected cost is $C_o k$.

Example 3: A large soft-drink company in Hong Kong faces an in-house problem of production planning and inventory control.

- Products are labeled as either **very high sales volume** (state 1), **high sales volume** (state 2), **standard sales volume** (state 3), **low sales volume** (state 4) or **very low sales volume** (state 5). ($m = 5$).

- For simplicity, we assume the following **symmetric cost matrix**:

$$C = \begin{pmatrix} 0 & 100 & 300 & 700 & 1500 \\ 100 & 0 & 100 & 300 & 700 \\ 300 & 100 & 0 & 100 & 300 \\ 700 & 300 & 100 & 0 & 100 \\ 1500 & 700 & 300 & 100 & 0 \end{pmatrix} \qquad (18)$$

where $[C]_{ij}$ is the assigned **penalty cost** when the production plan is for sales volume of State i and the actual sales volume is State j.

• We employ high-order Markov model for modeling the sales demand data sequence.

• Optimal production policy can then be derived based on Proposition 2.2.

• The following table shows the optimal costs based on three different models for three different products.

Products	A	B	C
High-order Markov Model ($n = 3$)	11200	9300	10800
First-order Markov Model ($n = 1$)	27600	18900	11100
Stationary Model	31900	18900	16300

• W. Ching et al., *A Higher-order Markov Model for the Newsboy's Problem*, Journal of Operational Research Society, 54, 2003.

4. Multivariate Markov Model for multiple Categorical Data Sequences.

• In many situations, there is a need to consider a number of categorical data sequences (having same number of categories/states) **together at the same time** because they are related to each other.

(i) **Gene Expression Sequences.** W. Ching et al., *On Construction of Stochastic Genetic Networks Based on Gene Expression Sequences*, International Journal of Neural Systems, 2005.

(ii) **Credit Risk Measurements.** T. Siu, W. Ching et al, *On Multivariate Credibility Approach for Portfolio Credit Risk Measurement*, Quantitative Finance, 2005.

(iii) **Sales Demand.** W. Ching et al., *A Multivariate Markov Chain Model for Categorical Data Sequences and Its Applications in Demand Prediction*, IMA Journal of Management Mathematics, 2002.

• **Problem:** The conventional first-order Markov chain model for s categorical data sequences of m states has m^s states.

4.1 The Multivariate Markov Chain Model.

• We propose a multivariate Markov chain model which can capture both the **intra- and inter-transition** probabilities among the sequences and the number of model parameters is $O(s^2m^2)$.

• Given s sequences with fixed m, the minimum number of parameters required is $\begin{pmatrix} s \\ 2 \end{pmatrix} = O(s^2)$.

• We assume that there are s categorical sequences and each has m possible states in

$$M = \{1, 2, \ldots, m\}.$$

• Let $\mathbf{X}_n^{(k)}$ be the state probability distribution vector of the kth sequence at time n. If the kth sequence is in State j at time n then

$$\mathbf{X}_n^{(k)} = \mathbf{e}_j = (0, \ldots, 0, \underbrace{1}_{j\text{th entry}}, 0 \ldots, 0)^T.$$

• In our proposed multivariate Markov chain model, we assume the following relationship:

$$\boxed{\mathbf{X}_{n+1}^{(j)} = \sum_{k=1}^{s} \lambda_{jk} P^{(jk)} \mathbf{X}_n^{(k)}, \quad \text{for} \quad j = 1, 2, \cdots, s} \tag{19}$$

where

$$\lambda_{jk} \geq 0, \quad 1 \leq j, k \leq s \quad \text{and} \quad \sum_{k=1}^{s} \lambda_{jk} = 1, \quad \text{for} \quad j = 1, 2, \cdots, s. \tag{20}$$

• The state probability distribution of the **kth sequence** at **time $(n+1)$** depends on the weighted average of $P^{(jk)}\mathbf{X}_n^{(k)}$.

• Here $P^{(jk)}$ is a one-step transition probability matrix from the states in the **kth sequence** to the states in the **jth sequence**. Here recall that $\mathbf{X}_n^{(k)}$ is the state probability distribution of the **kth sequences** at time n.

• In matrix form, we have the following block structure matrix equation (a compact representation):

$$\mathbf{X}_{n+1} \equiv \begin{pmatrix} \mathbf{X}_{n+1}^{(1)} \\ \mathbf{X}_{n+1}^{(2)} \\ \vdots \\ \mathbf{X}_{n+1}^{(s)} \end{pmatrix} = \begin{pmatrix} \lambda_{11}P^{(11)} & \lambda_{12}P^{(12)} & \cdots & \lambda_{1s}P^{(1s)} \\ \lambda_{21}P^{(21)} & \lambda_{22}P^{(22)} & \cdots & \lambda_{2s}P^{(2s)} \\ \vdots & \vdots & \vdots & \vdots \\ \lambda_{s1}P^{(s1)} & \lambda_{s2}P^{(s2)} & \cdots & \lambda_{ss}P^{(ss)} \end{pmatrix} \begin{pmatrix} \mathbf{X}_{n}^{(1)} \\ \mathbf{X}_{n}^{(2)} \\ \vdots \\ \mathbf{X}_{n}^{(s)} \end{pmatrix} \equiv Q\mathbf{X}_n$$

or

$$\boxed{\mathbf{X}_{n+1} = Q\mathbf{X}_n.}$$

Remark: We note that $\mathbf{X}_n^{(j)}$ is a probability distribution vector.

4.2 Some Properties of the Model.

Proposition 4.1: If $\lambda_{jk} > 0$ for $1 \leq j, k \leq s$, then the matrix Q has an eigenvalue equal to one and the eigenvalues of Q have modulus less than or equal to one.

Proposition 4.2: Suppose that $P^{(jk)}$ $(1 \leq j, k \leq s)$ are **irreducible** and $\lambda_{jk} > 0$ for $1 \leq j, k \leq s$. Then there is a vector $\mathbf{X} = [\mathbf{X}^{(1)}, \cdots, \mathbf{X}^{(s)}]^T$ such that

$$\boxed{\mathbf{X} = Q\mathbf{X} \quad \text{and} \quad \sum_{i=1}^{m} [\mathbf{X}^{(j)}]_i = 1, \ 1 \leq j \leq s.}$$

Moreover,

$$\boxed{\lim_{n \to \infty} \mathbf{X}_n = \mathbf{X}.}$$

Again the proof depends on Perron-Frobenius theorem.

Remark: Propositions 4.1 and 4.2 still hold if $[\Lambda_{ij}]$ is irreducible and aperiodic .

4.3 Parameter Estimation.

4.3.1 Estimations of $P^{(jk)}$.

• We can construct the **transition frequency matrix** from the observed data sequences. More precisely, we count the **transition frequency** $f_{i_j i_k}^{(jk)}$ from the state i_k in the sequence $\{X_n^{(k)}\}$ to the state i_j in the sequence $\{X_n^{(j)}\}$.

• Therefore we construct the transition frequency matrix for the sequences as follows:

$$F^{(jk)} = \begin{pmatrix} f_{11}^{(jk)} & \cdots & \cdots & f_{m1}^{(jk)} \\ f_{12}^{(jk)} & \cdots & \cdots & f_{m2}^{(jk)} \\ \vdots & \vdots & \vdots & \vdots \\ f_{1m}^{(jk)} & \cdots & \cdots & f_{mm}^{(jk)} \end{pmatrix}.$$

- From $F^{(jk)}$, we get the estimates for $P^{(jk)}$ as follows:

$$\widehat{P}^{(jk)} = \begin{pmatrix} \widehat{p}_{11}^{(jk)} & \cdots & \cdots & \widehat{p}_{m1}^{(jk)} \\ \widehat{p}_{12}^{(jk)} & \cdots & \cdots & \widehat{p}_{m2}^{(jk)} \\ \vdots & \vdots & \vdots & \vdots \\ \widehat{p}_{1m}^{(jk)} & \cdots & \cdots & \widehat{p}_{mm}^{(jk)} \end{pmatrix}$$

where

$$\widehat{p}_{i_j i_k}^{(jk)} = \frac{f_{i_j i_k}^{(jk)}}{\sum_{i_j=1}^{m} f_{i_j i_k}^{(jk)}}$$

4.4 Estimation of λ_{jk}

• We have seen that the multivariate Markov chain has a **stationary vector** (joint probability distribution vector) $\mathbf{X}$.

• The vector $\mathbf{X}$ can be estimated from the sequences by computing the proportion of the occurrence of each state in each of the sequences, and let us denote it by

$$\hat{\mathbf{X}} = (\hat{\mathbf{X}}^{(1)}, \hat{\mathbf{X}}^{(2)}, \ldots, \hat{\mathbf{X}}^{(s)})^T.$$

One would expect that

$$\begin{pmatrix} \lambda_{11}\hat{P}^{(11)} & \lambda_{12}\hat{P}^{(12)} & \ldots & \lambda_{1s}\hat{P}^{(1s)} \\ \lambda_{21}\hat{P}^{(21)} & \lambda_{22}\hat{P}^{(22)} & \ldots & \lambda_{2s}\hat{P}^{(2s)} \\ \vdots & \vdots & \vdots & \vdots \\ \lambda_{s1}\hat{P}^{(s1)} & \lambda_{s2}\hat{P}^{(s2)} & \ldots & \lambda_{ss}\hat{P}^{(ss)} \end{pmatrix} \hat{\mathbf{X}} \approx \hat{\mathbf{X}}. \tag{21}$$

• If $||.||_\infty$ is chosen to minimize the discrepancies then we have the following optimization problem:

$$\left\{\begin{array}{ll} & \min\limits_{\lambda}\left\{\max\limits_{i}\left|\left[\sum_{k=1}^{m}\lambda_{jk}\widehat{P}^{(jk)}\widehat{\mathbf{X}}^{(k)}-\widehat{\mathbf{X}}^{(j)}\right]_i\right|\right\} \\ \text{subject to} & \\ & \sum_{k=1}^{s}\lambda_{jk}=1, \quad \text{and} \quad \lambda_{jk}\geq 0, \quad \forall k. \end{array}\right. \tag{22}$$

Remark: Again other norms such as $||.||_2$ and $||.||_1$ can also be considered. The former will result in a **quadratic programming problem** while $||.||_1$ will still result in a **linear programming problem**.

Problem (22) can be formulated as s **linear programming problems**. For each j:

$$\min_{\lambda} w_j$$

subject to

$$\begin{cases} \begin{pmatrix} w_j \\ w_j \\ \vdots \\ w_j \end{pmatrix} \geq \hat{\mathbf{x}}^{(j)} - B_j \begin{pmatrix} \lambda_{j1} \\ \lambda_{j2} \\ \vdots \\ \lambda_{js} \end{pmatrix} \\ \begin{pmatrix} w_j \\ w_j \\ \vdots \\ w_j \end{pmatrix} \geq -\hat{\mathbf{x}}^{(j)} + B_j \begin{pmatrix} \lambda_{j1} \\ \lambda_{j2} \\ \vdots \\ \lambda_{js} \end{pmatrix}, \\ w_j \geq 0, \quad \sum_{k=1}^{s} \lambda_{jk} = 1, \quad \lambda_{jk} \geq 0, \quad \forall k, \end{cases} \tag{23}$$

where

$$B_j = [\hat{P}^{(j1)}\hat{\mathbf{X}}^{(1)} \mid \hat{P}^{(j2)}\hat{\mathbf{X}}^{(2)} \mid \cdots \mid \hat{P}^{(js)}\hat{\mathbf{X}}^{(s)}].$$

4.5 A Numerical Demonstration

- Consider $s = 2$ sequences of $m = 4$ states:

$$S_1 = \{4, 3, 1, 3, 4, 4, 3, 3, 1, 2, 3, 4\}$$

and

$$S_2 = \{1, 2, 3, 4, 1, 4, 4, 3, 3, 1, 3, 1\}.$$

- By counting the transition frequencies

$S_1 :$ 4 $\rightarrow$ 3 $\rightarrow$ 1 $\rightarrow$ 3 $\rightarrow$ 4 $\rightarrow$ 4 $\rightarrow$ 3 $\rightarrow$ 3 $\rightarrow$ 1 $\rightarrow$ 2 $\rightarrow$ 3 $\rightarrow$ 4

and

$S_2 :$ 1 $\rightarrow$ 2 $\rightarrow$ 3 $\rightarrow$ 4 $\rightarrow$ 1 $\rightarrow$ 4 $\rightarrow$ 4 $\rightarrow$ 3 $\rightarrow$ 3 $\rightarrow$ 1 $\rightarrow$ 3 $\rightarrow$ 1

we have

$$F^{(11)} = \begin{pmatrix} 0 & 0 & 2 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 2 \\ 0 & 0 & 2 & 1 \end{pmatrix} \quad \text{and} \quad F^{(22)} = \begin{pmatrix} 0 & 0 & 2 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{pmatrix}.$$

- Moreover by counting the inter-transition frequencies

$$
\begin{array}{cccccccccccccccccccccccc}
S_1: 4 & & 3 & & 1 & & 3 & & 4 & & 4 & & 3 & & 3 & & 1 & & 2 & & 3 & & 4 \\
 & \searrow & & \searrow & & \searrow & & \searrow & & \searrow & & \searrow & & \searrow & & \searrow & & \searrow & & \searrow & & \searrow & \\
S_2: 1 & & 2 & & 3 & & 4 & & 1 & & 4 & & 4 & & 3 & & 3 & & 1 & & 3 & & 1
\end{array}
$$

and

$$
\begin{array}{cccccccccccccccccccccccc}
S_1: 4 & & 3 & & 1 & & 3 & & 4 & & 4 & & 3 & & 3 & & 1 & & 2 & & 3 & & 4 \\
 & \nearrow & & \nearrow & & \nearrow & & \nearrow & & \nearrow & & \nearrow & & \nearrow & & \nearrow & & \nearrow & & \nearrow & & \nearrow & \\
S_2: 1 & & 2 & & 3 & & 4 & & 1 & & 4 & & 4 & & 3 & & 3 & & 1 & & 3 & & 1
\end{array}
$$

We have

$$
F^{(21)} = \begin{pmatrix} 1 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 3 & 0 \\ 1 & 0 & 0 & 2 \end{pmatrix}, \qquad F^{(12)} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 2 & 0 & 1 & 2 \\ 1 & 0 & 1 & 1 \end{pmatrix}.
$$

- After normalization we have the transition probability matrices:

$$\hat{P}^{(11)} = \begin{pmatrix} 0 & 0 & \frac{2}{5} & 0 \\ \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{2} & 1 & \frac{1}{5} & \frac{2}{3} \\ 0 & 0 & \frac{2}{5} & \frac{1}{3} \end{pmatrix}, \qquad \hat{P}^{(12)} = \begin{pmatrix} 0 & 1 & \frac{1}{4} & 0 \\ 0 & 0 & \frac{1}{4} & 0 \\ \frac{2}{3} & 0 & \frac{1}{4} & \frac{2}{3} \\ \frac{1}{3} & 0 & \frac{1}{4} & \frac{1}{3} \end{pmatrix},$$

$$\hat{P}^{(21)} = \begin{pmatrix} \frac{1}{2} & 0 & \frac{2}{5} & 0 \\ 0 & 0 & 0 & \frac{1}{3} \\ 0 & 1 & \frac{3}{5} & 0 \\ \frac{1}{2} & 0 & 0 & \frac{2}{3} \end{pmatrix}, \qquad \hat{P}^{(22)} = \begin{pmatrix} 0 & 0 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & 1 & \frac{1}{4} & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{1}{4} & \frac{1}{3} \end{pmatrix}.$$

- Moreover, we also have

$$\hat{\mathbf{X}}_1 = (\frac{1}{6}, \frac{1}{12}, \frac{5}{12}, \frac{1}{3})^T \quad \text{and} \quad \hat{\mathbf{X}}_2 = (\frac{1}{3}, \frac{1}{12}, \frac{1}{3}, \frac{1}{4})^T$$

Solving the corresponding linear programming problems, the multivariate Markov models of the two categorical data sequences S_1 and S_2 are then given by

$$\begin{cases} \mathbf{X}_{n+1}^{(1)} = 0.00\hat{P}^{(11)}\mathbf{X}_n^{(1)} + 1.00\hat{P}^{(12)}\mathbf{X}_n^{(2)} \\ \mathbf{X}_{n+1}^{(2)} = 0.89\hat{P}^{(21)}\mathbf{X}_n^{(1)} + 0.11\hat{P}^{(22)}\mathbf{X}_n^{(2)}. \end{cases}$$

Proposition 4.3 [Zhu and Ching (2011)] If $\lambda_{ii} > 0$, P_{ii} is **irreducible** (for $1 \leq i \leq s$), the matrix $[\lambda_{ij}]$ is also **irreducible** and at least one of P_{ii} is **aperiodic** then the model has a stationary joint probability distribution

$$\mathbf{X} = (\mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(s)})^T$$

satisfying

$$\mathbf{X} = Q\mathbf{X}.$$

Moreover, we have

$$\lim_{t \to \infty} \mathbf{X}_t = \mathbf{X}.$$

• D. Zhu and W. Ching, *A Note on the Stationary Property of High-dimensional Markov Chain Models*, International Journal of Pure and Applied Mathematics, 66, 2011.

5. Concluding Remarks.

5.1 Extension to High-order Multivariate Markov Chain Models

• We assume that there are s **categorical sequences** with **order** n and each has m **possible states** in M. In the proposed model, we assume that the jth sequence at time $t = r + 1$ depends on all the sequences at times $t = r, r - 1, \ldots, r - n + 1$.

• Using the same notations, our proposed high-order (nth-order) multivariate Markov chain model takes the following form:

$$\mathbf{x}_{r+1}^{(j)} = \sum_{k=1}^{s} \sum_{h=1}^{n} \lambda_{jk}^{(h)} P_h^{(jk)} \mathbf{x}_{r-h+1}^{(k)}, \quad j = 1, 2, \ldots, s \tag{24}$$

where

$$\lambda_{jk}^{(h)} \geq 0, \quad 1 \leq j, k \leq s, \quad 1 \leq h \leq n$$

and

$$\sum_{k=1}^{s} \sum_{h=1}^{n} \lambda_{jk}^{(h)} = 1, \quad j = 1, 2, \ldots, s.$$

In fact, if we let

$$\mathbf{X}_r^{(j)} = ((\mathbf{x}_r^{(j)})^T, (\mathbf{x}_{r-1}^{(j)})^T, \ldots, (\mathbf{x}_{r-n+1}^{(j)})^T)^T$$

for $j = 1, 2, \ldots, s$ be the $nm \times 1$ vectors.

Then the model can be written as the following matrix form:

$$\mathbf{X}_{r+1} \equiv \begin{pmatrix} \mathbf{X}_{r+1}^{(1)} \\ \mathbf{X}_{r+1}^{(2)} \\ \vdots \\ \mathbf{X}_{r+1}^{(s)} \end{pmatrix} = \begin{pmatrix} B^{(11)} & B^{(12)} & \ldots & B^{(1s)} \\ B^{(21)} & B^{(22)} & \ldots & B^{(2s)} \\ \vdots & \vdots & \vdots & \vdots \\ B^{(s1)} & B^{(s2)} & \ldots & B^{(ss)} \end{pmatrix} \begin{pmatrix} \mathbf{X}_r^{(1)} \\ \mathbf{X}_r^{(2)} \\ \vdots \\ \mathbf{X}_r^{(s)} \end{pmatrix} \equiv J\mathbf{X}_r$$

where

$$B^{(ii)} = \begin{pmatrix} \lambda_{ii}^{(1)} P_1^{(ii)} & \lambda_{ii}^{(2)} P_2^{(ii)} & \dots & \lambda_{ii}^{(n-1)} P_{n-1}^{(ii)} & \lambda_{ii}^{(n)} P_n^{(ii)} \\ I & 0 & \dots & 0 & 0 \\ 0 & I & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & I & 0 \end{pmatrix}_{mn \times mn}$$

and if $i \neq j$ then

$$B^{(ij)} = \begin{pmatrix} \lambda_{ij}^{(1)} P_1^{(ij)} & \lambda_{ij}^{(2)} P_2^{(ij)} & \dots & \lambda_{ij}^{(n-1)} P_{n-1}^{(ij)} & \lambda_{ij}^{(n)} P_n^{(ij)} \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & 0 & 0 \end{pmatrix}_{mn \times mn} .$$

• W. Ching et al., *Higher-order Multivariate Markov Chains and their Applications*, Linear Algebra and Its Applications, 428 (2-3), 2008.

• We define an $s \times s$ matrix $\tilde{B}$.

We let $\tilde{B}_{ii} = 1$ for all $i = 1, 2, \ldots, s$, and if $i \neq j$

$$\tilde{B}_{ij} = \begin{cases} 1 & \text{if } \lambda_{ij}^{(k)} > 0 \quad \text{for all } k \\ 0 & \text{otherwise.} \end{cases}$$

Here $j = 1, 2, \ldots, s, k = 1, 2, \ldots, n$.

Proposition 4.4 If $\lambda_{ii}^{(1)}, \lambda_{ii}^{(n)} > 0$, $P_1^{(ii)}$ is irreducible and at least one of them is aperiodic ($i = 1, 2, \ldots, s$), additionally, $\tilde{B}$ is irreducible, then the model has a stationary probability distribution $\mathbf{X}$ satisfying

$$\mathbf{X} = J\mathbf{X}.$$

Moreover,

$$\lim_{t \to \infty} \mathbf{X}_t = \mathbf{X}.$$

- Theoretical results of the multivariate Markov chains were obtained when $\lambda_{ij} \geq 0$ (non-negative matrix). It is interesting to study the properties of the models when λ_{ij} are allowed to be **negative**.

- An example of a two-chain model:

$$\begin{pmatrix} \mathbf{x}_{n+1}^{(1)} \\ \mathbf{x}_{n+1}^{(2)} \end{pmatrix} = \underbrace{\begin{pmatrix} \lambda_{1,1}P^{(11)} & \lambda_{1,2}I \\ \lambda_{2,1}I & \lambda_{2,2}P^{(22)} \end{pmatrix} \begin{pmatrix} \mathbf{x}_n^{(1)} \\ \mathbf{x}_n^{(2)} \end{pmatrix}}_{Positive\ correlated\ part}$$
$$+ \underbrace{\frac{1}{m-1} \begin{pmatrix} \lambda_{1,-1}P^{(11)} & \lambda_{1,-2}I \\ \lambda_{2,-1}I & \lambda_{2,-2}P^{(2,2)} \end{pmatrix} \begin{pmatrix} \mathbf{1} - \mathbf{x}_n^{(1)} \\ \mathbf{1} - \mathbf{x}_n^{(2)} \end{pmatrix}}_{Negative\ correlated\ part}.$$

Here $\lambda_{i,j} \geq 0$ for $i = 1, 2$ and $j = \pm 1, \pm 2$ and $\sum_{j=-2}^{2} \lambda_{i,j} = 1$.

- W. Ching, et al., *An Improved Parsimonious Multivariate Markov Chain Model for Credit Risk*, Journal of Credit Risk, 5, 2009.

$$\mathbf{x}_{n+1} \equiv \begin{pmatrix} \mathbf{x}_{n+1}^{(1)} \\ \mathbf{x}_{n+1}^{(2)} \\ \vdots \\ \mathbf{x}_{n+1}^{(s)} \end{pmatrix} \tag{25}$$

$$= \begin{pmatrix} \lambda_{11}P^{(11)} & \lambda_{12}I & \cdots & \lambda_{1s}I \\ \lambda_{21}I & \lambda_{22}P^{(22)} & \cdots & \lambda_{2s}I \\ \vdots & \vdots & \vdots & \vdots \\ \lambda_{s1}I & \lambda_{s2}I & \cdots & \lambda_{ss}P^{(ss)} \end{pmatrix} \begin{pmatrix} \mathbf{x}_n^{(1)} \\ \mathbf{x}_n^{(2)} \\ \vdots \\ \mathbf{x}_n^{(s)} \end{pmatrix} \tag{26}$$

$$\equiv Q\mathbf{x}_n \tag{27}$$

- The models can still capture both the intra- and inter-transition probabilities among the sequences but the number of parameters in the proposed model is $O(sm^2 + s^2)$.

Proposition 5.1. Suppose that $P^{(jj)}$ $(1 \le j \le s)$ and $\Lambda = [\lambda_{jk}]^T$ are **irreducible**. Then there is a vector

$$\mathbf{x} = (\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(s)})^T$$

such that $\mathbf{x} = Q\mathbf{x}$ and

$$\sum_{i=1}^{m} [\mathbf{x}^{(j)}]_i = 1, \ 1 \le j \le s.$$

- The estimation methods described previously for the model parameters λ_{ij} and P_{ij} can still be applied.

- W. Ching et al., *On Multi-dimensional Markov Chain Models*, Pacific Journal of Optimization, 3 (2007) 235-243.

5.4 The Simplified Model with Short Sequences.

In this part, we consider the case when the length of the observed data sequences are very short.

In this case, we have two problems:

(i) The estimation of the transition probability matrices P_{ij} may have large error.

(ii) The steady-state of the process may not be reached.

• For Problem (i) we propose to replace the transition probability matrix $P^{(ii)}$ in the model by the following rank-one matrix

$$(\hat{\mathbf{x}}^{(i)})^T(1, 1, \ldots, 1). \tag{28}$$

In this case, the number of parameters will further reduce to $O(sm + s^2)$

• For Problem (ii), the weights λ_{ij} should be chosen such that the multivariate Markov process converges very fast to its stationary distribution.

The convergence rate of the process depends on the second largest eigenvalue in modulus of the matrix Q . From numerical experience, the second largest eigenvalue depends very much on the value of λ_{ii}. We then modified our simplified model for short sequences by adding the extra constraints

$$0 \le \lambda_{ii} \le \beta. \tag{29}$$

5.5 A Special Case Analysis.

• We give an analysis of the simplified model with the assumptions (28) and (29) by further assuming that

$$P = P^{(ii)} = (\hat{\mathbf{x}})^T(1, 1, \ldots, 1) \quad \text{for all } i = 1, 2, \ldots, s.$$

In this case, for $\lambda_{ij} > 0$ the steady-state probability distributions $\hat{\mathbf{x}}$ is an invariant.

• The problem here is how to assign λ_{ij} such that the second largest eigenvalue of Q is small. For simplicity of discussion, we assume one possible form for $[\Lambda]$ as follows:

$$\boxed{\lambda_{ij} = \begin{cases} \lambda & \text{if} \quad i = j \\ \frac{1-\lambda}{m-1} & \text{if} \quad i \neq j \end{cases}}$$

where $0 < \lambda < 1$.

- With these assumptions, we have the tensor product form for

$$Q = I \otimes \lambda P + (\Lambda - \lambda I) \otimes I.$$

The eigenvalues of Q are then then given by

$$1, \quad 1 - \lambda, \quad \frac{\lambda - 1}{m - 1}, \quad \text{and} \quad \frac{\lambda m - 1}{m - 1}$$

where 1 and $1 - \lambda$ are the two simple eigenvalues. The second largest eigenvalue of Q can be minimized by solving the following maxmin problem:

$$\min_{0<\lambda<1} \left\{ \max \left\{ 1 - \lambda, \frac{\lambda - 1}{m - 1}, \frac{\lambda m - 1}{m - 1} \right\} \right\}.$$

It is straight forward to check that the optimal value is

$$\lambda^* = \frac{m}{2m - 1}$$

and the optimal second largest eigenvalue in this case is

$$\frac{m}{2m - 1}.$$

5.6 Numerical Examples

- A soft-drink company in Hong Kong presently faces an in-house problem of production planning and inventory control. A pressing issue that stands out is the storage space of its central warehouse, which often finds itself in the state of overflow or near capacity.

- Products (A,B,C,D,E) at different time t can be categorized into five possible states according to sales volume. All products are labelled as either
(1) very fast-moving (very high sales volume),
(2) fast-moving,
(3) standard,
(4) slow-moving, or
(5) very slow-moving (very low sales volume).

• We then build the multivariate Markov model, to predict the next state $\hat{\mathbf{x}}_t$ at time t which can be taken as the state with the maximum probability, i.e.,

$$\hat{\mathbf{x}}_t = j, \quad \text{if } [\hat{\mathbf{x}}_t]_i \le [\hat{\mathbf{x}}_t]_j, \forall 1 \le i \le m.$$

• To evaluate the performance and effectiveness of our multivariate Markov chain model, a prediction result is measured by the prediction accuracy r defined as

$$\boxed{r = \frac{1}{T} \times \sum_{t=1}^{T} \delta_t \times 100\%,}$$

where T is the length of the data sequence and

$$\delta_t = \begin{cases} 1, & \text{if } \hat{\mathbf{x}}_t = \mathbf{x}_t \\ 0, & \text{otherwise.} \end{cases}$$

• Another way to compare the performance of the models is to use the BIC (Bayesian Information Criterion) which is defined as

$$\boxed{BIC = -2L + q \log n,}$$

where

$$\boxed{L = \sum_{j=1}^{s} \left\{ \sum_{i_0,k_1,\cdots k_s=1}^{m} n_{i_0,k_1,\cdots,k_s}^{(j)} \log\left(\sum_{l=1}^{m} \sum_{k=1}^{s} \lambda_{jk} p_{i_0,k_l}^{(jk)}\right) \right\},}$$

is the log-likelihood of the model,

$$n_{i_0,k_1,k_2,\cdots,k_s}^{(j)} = \sum x_{n+1}^{(j)}(i_0) x_n^1(k_1) x_n^2(k_2) \cdots x_n^s(k_s),$$

and q is the number of independent parameters, and n is the length of the sequence.

- The smaller the value of BIC, the better the model is.

Models	A	B	C	D	E
Multivariate Markov Model	50%	45%	63%	52%	55%
Simplified Model	46%	46%	63%	52%	54%
Simplified Model For Short Sequences ($\beta = 1.0$)	43%	40%	63%	36%	36%
Simplified Model For Short Sequences ($\beta = 0.9$)	42%	41%	63%	33%	36%
Simplified Model For Short Sequences ($\beta = 0.8$)	42%	38%	63%	37%	38%
Simplified Model For Short Sequences ($\beta = 0.7$)	42%	37%	45%	39%	39%
Simplified Model For Short Sequences ($\beta = 0.6$)	39%	37%	35%	39%	39%
Simplified Model For Short Sequences ($\beta = \frac{6}{2*6-1}$)	42%	38%	56%	24%	24%

Table 5.1 Prediction Accuracy for the Sales Demand Data Sequences of Full Length.

Models	BIC
Multivariate Markov Model	8.0215e+003
Simplified Model	3.9878e+003
Simplified Model For Short Sequences ($\beta = 1.0$)	4.5422e+003
Simplified Model For Short Sequences ($\beta = 0.9$)	4.5292e+003*
Simplified Model For Short Sequences ($\beta = 0.8$)	4.5747e+003
Simplified Model For Short Sequences ($\beta = 0.7$)	4.6673e+003
Simplified Model For Short Sequences ($\beta = 0.6$)	4.8074e+003
Simplified Model For Short Sequences ($\beta = \frac{6}{2*6-1}$)	4.9585e+003

Table 5.2 BIC for the Sales Demand Data Sequences in Full Length.

Models	A	B	C	D	E
Multivariate Markov Model	66%	59%	83%	55%	55%
Simplified Model	66%	55%	83%	55%	52%
Simplified Model For Short Sequences ($\beta = 1.0$)	48%	52%	83%	34%	31%
Simplified Model For Short Sequences ($\beta = 0.9$)	48%	52%	83%	41%	45%
Simplified Model For Short Sequences ($\beta = 0.8$)	52%	55%	83%	41%	45%
Simplified Model For Short Sequences ($\beta = 0.7$)	52%	55%	83%	41%	45%
Simplified Model For Short Sequences ($\beta = 0.6$)	52%	31%	72%	41%	45%
Simplified Model For Short Sequences ($\beta = \frac{6}{2*6-1}$)	55%	45%	83%	17%	10%

Table 5.3 Prediction Accuracy for the Sales Demand Data Sequences of Shorter Length of 30.

Models	BIC
Multivariate Markov Model	3.3714e+003
Simplified Model	934.3750
Simplified Model For Short Sequences ($\beta = 1.0$)	1.0138e+003
Simplified Model For Short Sequences ($\beta = 0.9$)	1.0079e+003*
Simplified Model For Short Sequences ($\beta = 0.8$)	1.0089e+003
Simplified Model For Short Sequences ($\beta = 0.7$)	1.0151e+003
Simplified Model For Short Sequences ($\beta = 0.6$)	1.0264e+003
Simplified Model For Short Sequences ($\beta = \frac{6}{2*6-1}$)	1.0749e+003

Table 5.4 BIC for the Sales Demand Data Sequences in Shorter Length of 30.


应用数学译丛
Markov Chains
Models, Algorithms and Applications Second Edition
马尔可夫链：模型、算法与应用
[美] Wai-Ki Ching, Ximin Huang, Michael K.Ng, Tak-Kuen Siu 著
陈曦 译
www.tup.tsinghua.edu.cn
清华大学出版社
International Series in
Operations Research & Management Science
Wai-Ki Ching
Ximin Huang
Michael K. Ng
Tak-Kuen Siu
Markov Chains
Models, Algorithms and Applications
Second Edition
Operations Research Management Science
Springer