Queueing estimation insights from 200 papers applied to 200,000 patient journey observations.

Yoni Nazarathy,

INFORMS-APS, 2017.

informs-aps.smp.uq.edu.au

Home Register Submit an abstract View the program Accommodation details

INFORMS-APS 2019, Brisbane Australia

The 20'th INFORMS Applied Probability Society Conference, July 2-4, 2019, Brisbane Australia.

We are delighted to announce the 20'th INFORMS Applied Probability Society Conference taking place at the Brisbane Convention Centre, Brisbane Australia, July 2-4, 2019.

+ July 5,
Applied^2 Prob

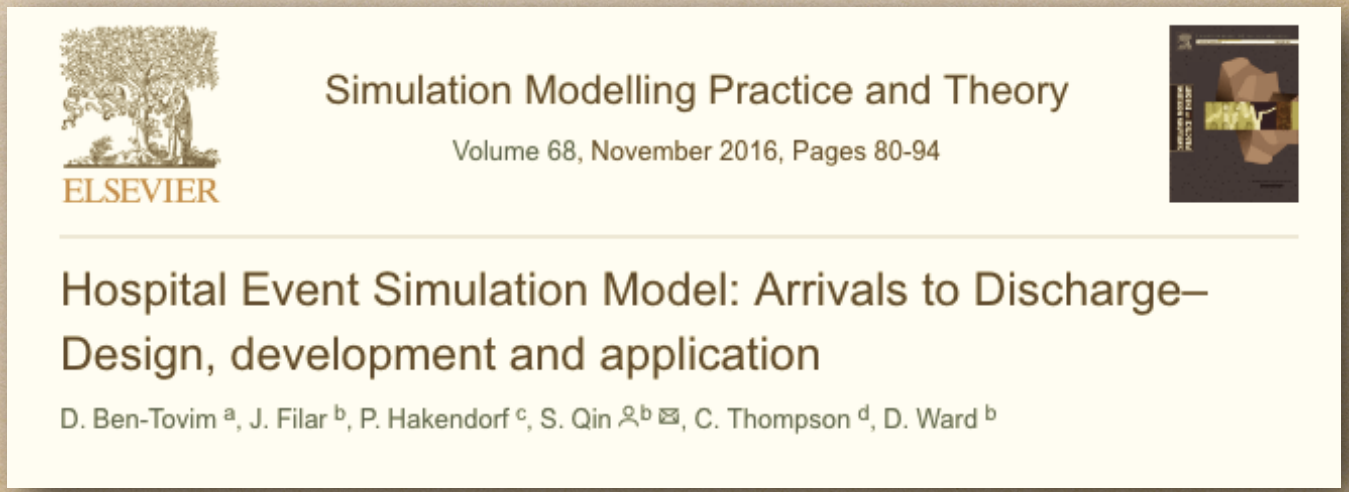
12-th International Conference on Monte Carlo Methods and Applications

or MCM2019 for short is scheduled to be held in July 2019 in Sydney, Australia and to run back-to-back with the 2019 INFORMS-APS Conference in Brisbane, Australia.
You may explore earlier editions of MCM2019
For more details contact the local chair Zdravko Botev.


Coral Sea
Australia
Great
Australian
Bight
Pomona
Noosa
North Shore
Noosa Heads
Cooroy
Eumundi
Belli Park
Coolum Beach
Nambour
Sunshine Coast
Montville
Mooloolaba
Maleny
Landsborough
Caloundra
Golden Beach
Beerwah

Sunshine Coast Hospital and Health Service

- About 50K patients arrivals per year emergency/elective
- 4 years of patient journey data
- Ethics approval
- Questions: Transitions from an old main hospital to new
- Related work:
  - Hospitals in Adelaide, South Australia
  - Data driven discrete event simulation:



ELSEVIER

Simulation Modelling Practice and Theory

Volume 68, November 2016, Pages 80-94

Hospital Event Simulation Model: Arrivals to Discharge–Design, development and application

D. Ben-Tovim [a], J. Filar [b], P. Hakendorf [c], S. Qin [b], C. Thompson [d], D. Ward [b]

TODAY: NO PATIENTS - ONLY PAPERS

J	K	L	M	N	O	P	
Arrival Date	Triaged At	Treating Clin Seen At	Senior ED Dr Seen At	Treat Nrs Seen At	Presenting	Presenting	Diagı
1/01/2016 12:19:00 AM	1/01/2016 12:19:00 AM	1/01/2016 1:00:00 AM	1/01/2016 1:15:00 AM		9000	INJURY	T16
1/01/2016 12:31:00 AM	1/01/2016 12:31:00 AM	1/01/2016 12:52:00 AM			9012	FALL	S13.
1/01/2016 12:47:00 AM	1/01/2016 12:47:00 AM	1/01/2016 2:02:00 AM			9033	LACERATIO	S01.
1/01/2016 1:04:00 AM	1/01/2016 1:04:00 AM	1/01/2016 1:06:00 AM			9000	INJURY	T20.
1/01/2016 1:23:00 AM	1/01/2016 1:23:00 AM	1/01/2016 3:47:00 AM			36000	SYMPTOMS	F10.
1/01/2016 1:31:00 AM	1/01/2016 1:31:00 AM	1/01/2016 1:40:00 AM			9000	INJURY	S93.
1/01/2016 1:57:00 AM	1/01/2016 1:57:00 AM	1/01/2016 2:19:00 AM	1/01/2016 2:15:00 AM		36016	POSSIBLE A	T78.
1/01/2016 2:03:00 AM	1/01/2016 2:03:00 AM			1/01/2016 2:05:00 AM	9000	INJURY	S83.
1/01/2016 2:09:00 AM	1/01/2016 2:09:00 AM	1/01/2016 3:01:00 AM	1/01/2016 2:47:00 AM	1/01/2016 2:28:00 AM	36000	SYMPTOMS	R55
1/01/2016 2:17:00 AM	1/01/2016 2:17:00 AM	1/01/2016 3:36:00 AM	1/01/2016 2:49:00 AM	1/01/2016 2:28:00 AM	36000	SYMPTOMS	I50.0
1/01/2016 2:21:00 AM	1/01/2016 2:21:00 AM	1/01/2016 2:56:00 AM			9000	INJURY	T00.
1/01/2016 2:31:00 AM	1/01/2016 2:31:00 AM	1/01/2016 3:48:00 AM	1/01/2016 2:51:00 AM		9000	INJURY	S13.
1/01/2016 2:36:00 AM	1/01/2016 2:36:00 AM	1/01/2016 3:00:00 AM	1/01/2016 2:47:00 AM		25000	PAIN (NOT A	K57.
1/01/2016 2:46:00 AM	1/01/2016 2:46:00 AM	1/01/2016 3:54:00 AM			30000	MENTAL HE	F43.
1/01/2016 2:57:00 AM	1/01/2016 2:57:00 AM	1/01/2016 4:06:00 AM	1/01/2016 3:20:00 AM		9000	INJURY	S00.
1/01/2016 3:28:00 AM	1/01/2016 3:28:00 AM	1/01/2016 5:00:00 AM		1/01/2016 3:50:00 AM	36000	SYMPTOMS	M75.
1/01/2016 3:45:00 AM	1/01/2016 3:45:00 AM	1/01/2016 5:11:00 AM		1/01/2016 4:16:00 AM	36000	SYMPTOMS	S00.
1/01/2016 4:01:00 AM	1/01/2016 4:01:00 AM	1/01/2016 6:33:00 AM	1/01/2016 5:01:00 AM		9000	INJURY	S02.
1/01/2016 4:22:00 AM	1/01/2016 4:22:00 AM	1/01/2016 5:16:00 AM	1/01/2016 4:31:00 AM		36000	SYMPTOMS	F10.
1/01/2016 4:46:00 AM	1/01/2016 4:46:00 AM	1/01/2016 5:37:00 AM			36000	SYMPTOMS	E16.
1/01/2016 4:51:00 AM	1/01/2016 4:51:00 AM	1/01/2016 5:41:00 AM			9000	INJURY	T00.
1/01/2016 4:53:00 AM	1/01/2016 4:53:00 AM	1/01/2016 6:55:00 AM		1/01/2016 5:07:00 AM	9000	INJURY	Z53.
1/01/2016 5:14:00 AM	1/01/2016 5:14:00 AM	1/01/2016 6:15:00 AM		1/01/2016 5:52:00 AM	36000	SYMPTOMS	R45.
1/01/2016 5:42:00 AM	1/01/2016 5:42:00 AM	1/01/2016 5:50:00 AM			9000	INJURY	T43.
1/01/2016 6:02:00 AM	1/01/2016 6:02:00 AM	1/01/2016 6:23:00 AM	1/01/2016 8:00:00 AM		25000	PAIN (NOT A	R07.
1/01/2016 6:36:00 AM	1/01/2016 6:36:00 AM	1/01/2016 6:43:00 AM	1/01/2016 1:06:00 PM		36000	SYMPTOMS	J18.
1/01/2016 7:17:00 AM	1/01/2016 7:17:00 AM	1/01/2016 7:32:00 AM	1/01/2016 7:24:00 AM		25006	CHEST PAIN	I48

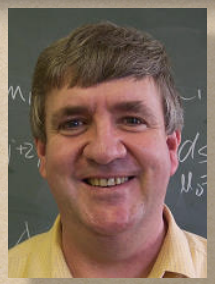
Parameter and State Estimation in Queues and Related Stochastic Models: A Bibliography

Azam Asanjarani, Yoni Nazarathy and Philip K. Pollett

The School of Mathematics and Physics, The University of Queensland,

Brisbane Australia

January 31, 2017



Some Problems of Statistical Analysis Connected with Congestion

D. R. COX, *Birkbeck College, University of London*

SUMMARY

Problems of statistical analysis connected with queues are discussed, previous work is reviewed, and solutions are outlined for a few of the simpler new p
ing models is
information a
and b) service
count of probl

THE STATISTICAL ANALYSIS OF CONGESTION

By D. R. Cox

Statistical Laboratory, University of Cambridge

1. *Introduction*

ONE common type of investigation undertaken by operational research workers aims at attaining an efficient use of resources by eliminating or reducing congestion. To provide a theoretical background to these p
predicting theoretically
as models of the comp
some of this theoretical

Congestion is usuall

(i) a flow of
(ii) some rest
(iii) irregular

There are two forms of

(*a*) a restricti
(*b*) a restricti

In many practical cases

MAXIMUM LIKELIHOOD ESTIMATES IN A SIMPLE QUEUE

By A. Bruce Clarke[1]

University of Michigan

0. Summary. The problem of obtaining maximum likelihood estimates for the parameters involved in a stationary single-channel, Markovian queuing process is considered. A method of taking observations is presented which simplifies this problem to that of determining a root of a certain quadratic equation. A useful and even simpler rational approximation is also studied.

1. Introduction. By a simple queue is meant a queue having a Poisson input and a negative exponential service time (type $M/M/1$ in the notation of Kendall

Received November 19, 1956.

[1] Research under contract with the Office of Naval Research carried out at Cornell University.

Chapter 13

Statistical analysis of queueing systems[1]

U. Narayan Bhat, Gregory K. Miller, and S. Subba Rao

ABSTRACT This paper provides an overview of the literature on the statistical analysis of queueing systems. Topics discussed include: model identification, parameter estimation using the maximum likelihood, method of moments and Bayesian frameworks, a discussion of covariance structure and autocorrelation in queueing systems, estimation from simulation experiments, hypothesis testing, and other related aspects. The bibliography, fairly exhaustive, should provide the reader with a source of articles that comprise the core of work done up to the present time.

CONTENTS

13.1	Introduction	351
13.2	Identification of models	352
13.3	Parameter estimation: The maximum likelihood method	354
13.4	Parameter estimation: The method of moments	360
13.5	Bayes estimation	361
13.6	Covariance structure, autocorrelation and mean value estimation	364
13.7	Estimation from special types of data	368
13.8	Estimation related to performance measures	372
13.9	Estimation from simulation of queueing processes	
13.10	Hypothesis testing	
13.11	Other related topics and future prospects	
	Bibliography	

Invited paper

STATISTICAL ANALYSIS OF QUEUEING SYSTEMS

U. NARAYAN BHAT

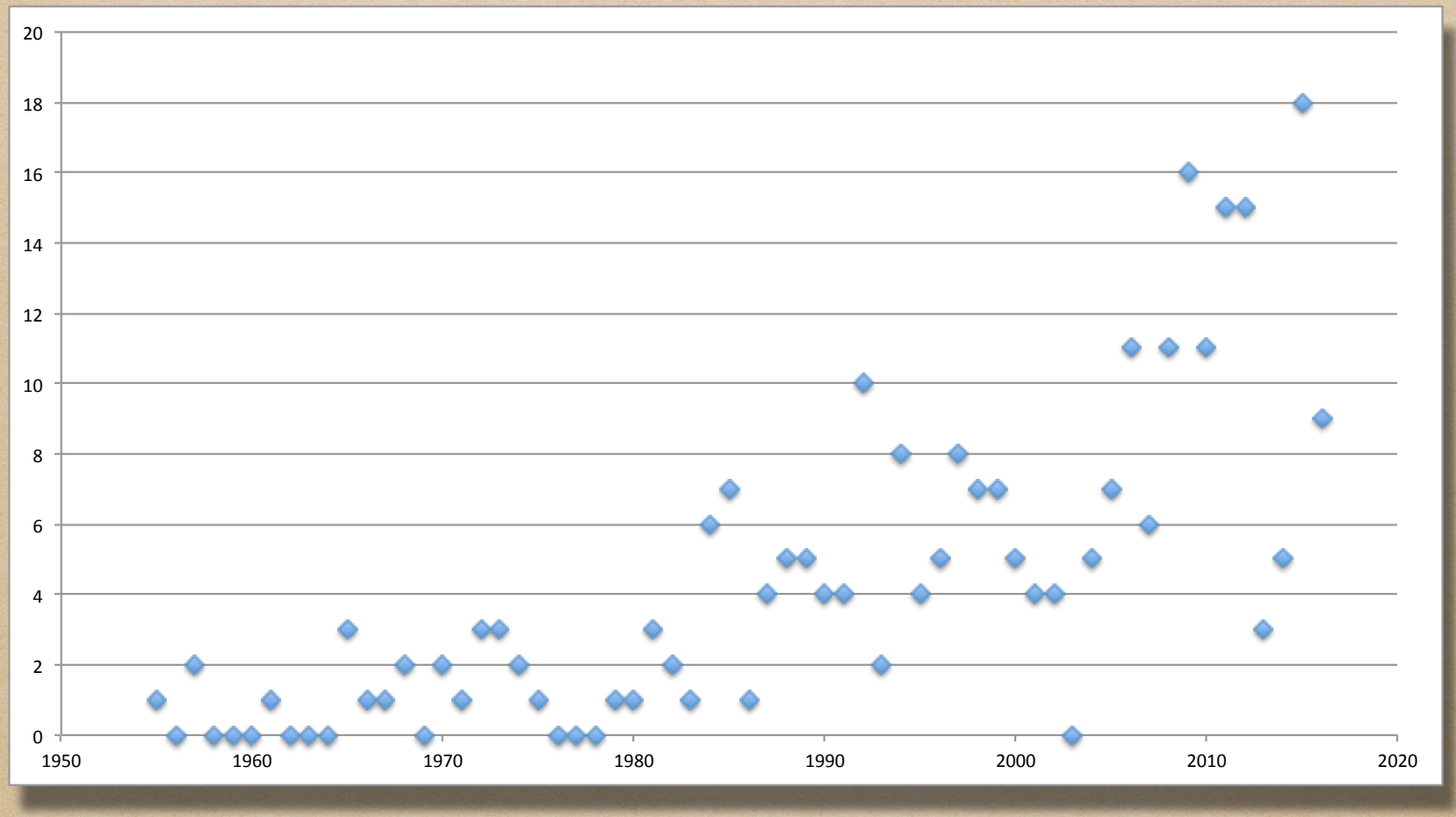
Department of Statistics, Southern Methodist University, Dallas, Texas 75275, USA

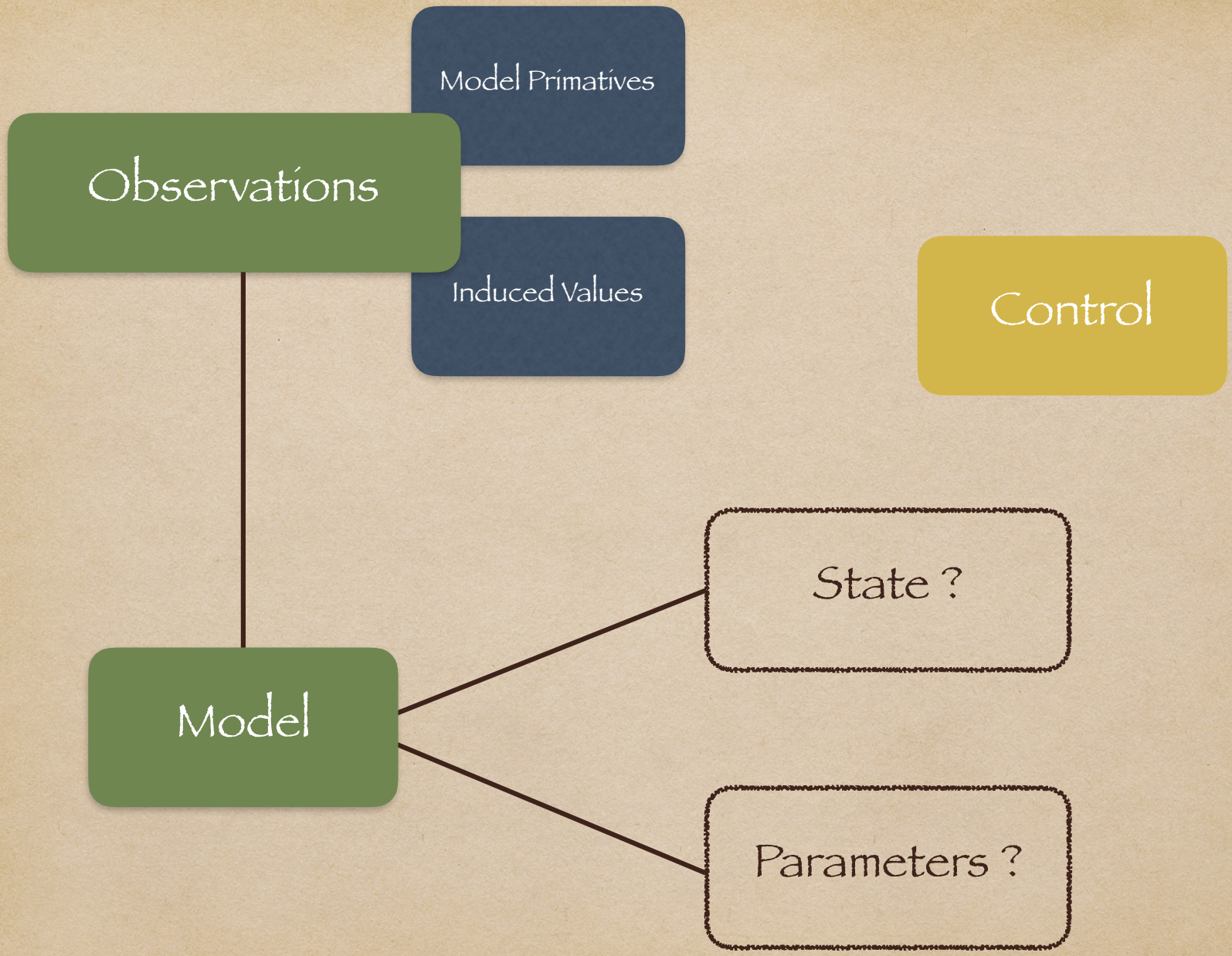
and

S. SUBBA RAO*

Department of Management and Systems, Washington State University, Pullman, Washington 99164, USA

Number of Papers per year 1955 - 2016




Model Primatives
Observations
Induced Values
Control
State ?
Model
Parameters ?

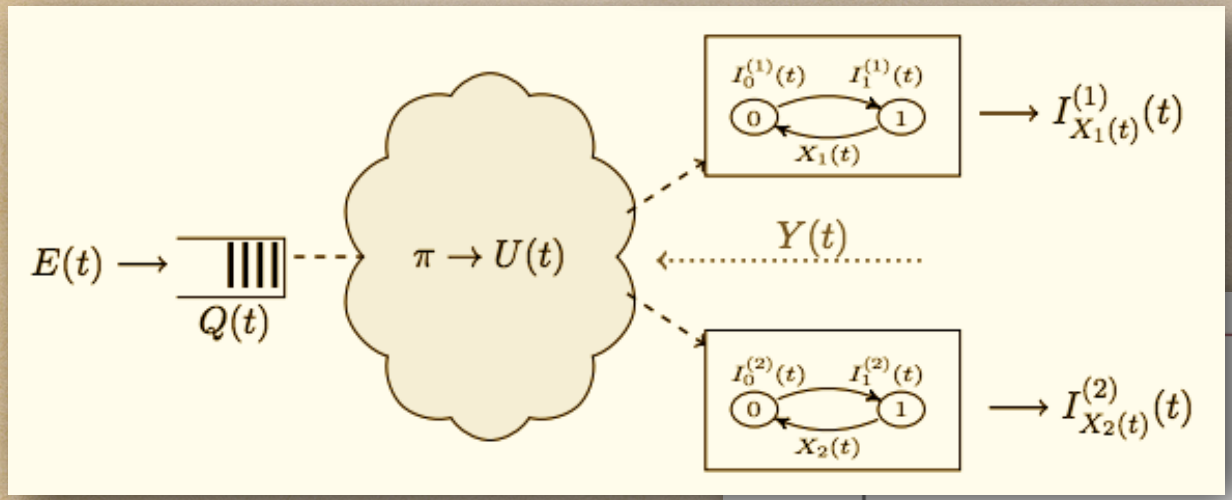
What do you want to do?

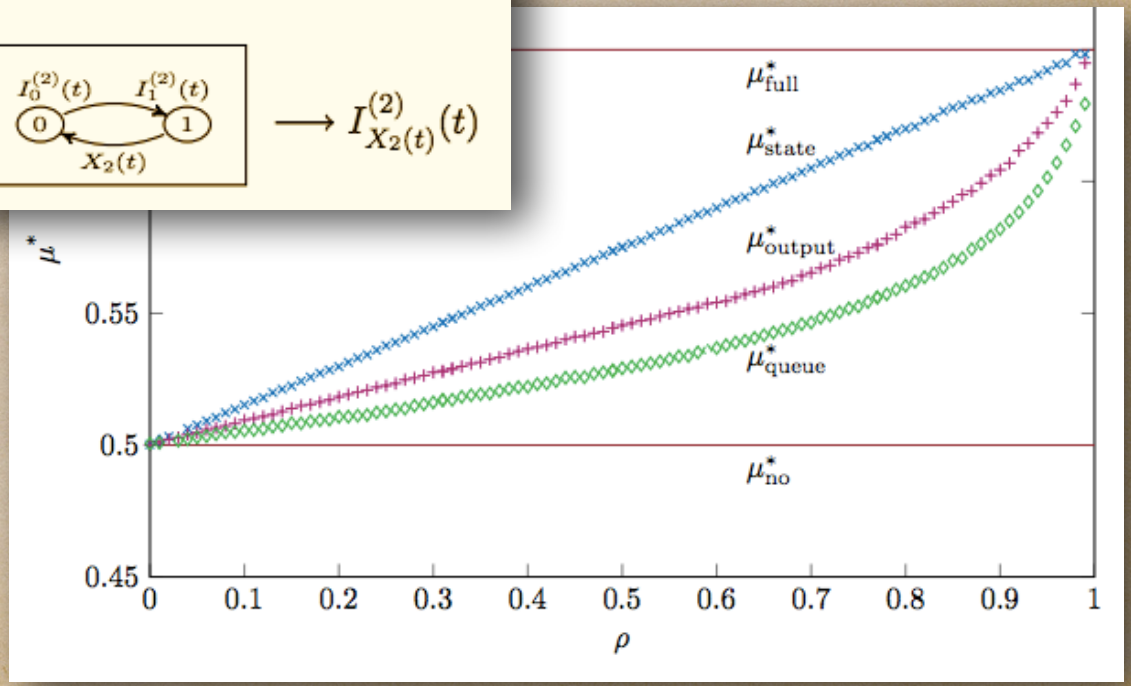
- Control
- Design
- Predict
- Probe
- Assume Customers Don't Interact
- Use Queue Inference Engine
- Consider Likelihood
- Be Bayesian
- Solve an Inverse Problem
- Handle Discrete Sampling
- Use Queueing Fundamentals
- Mine

Control

The Role of Information in System Stability with Partially Observable Servers

A. Asanjarani*, Y. Nazarathy†





Design

Architecture and robustness tradeoffs in speed-scaled queues with application to energy management

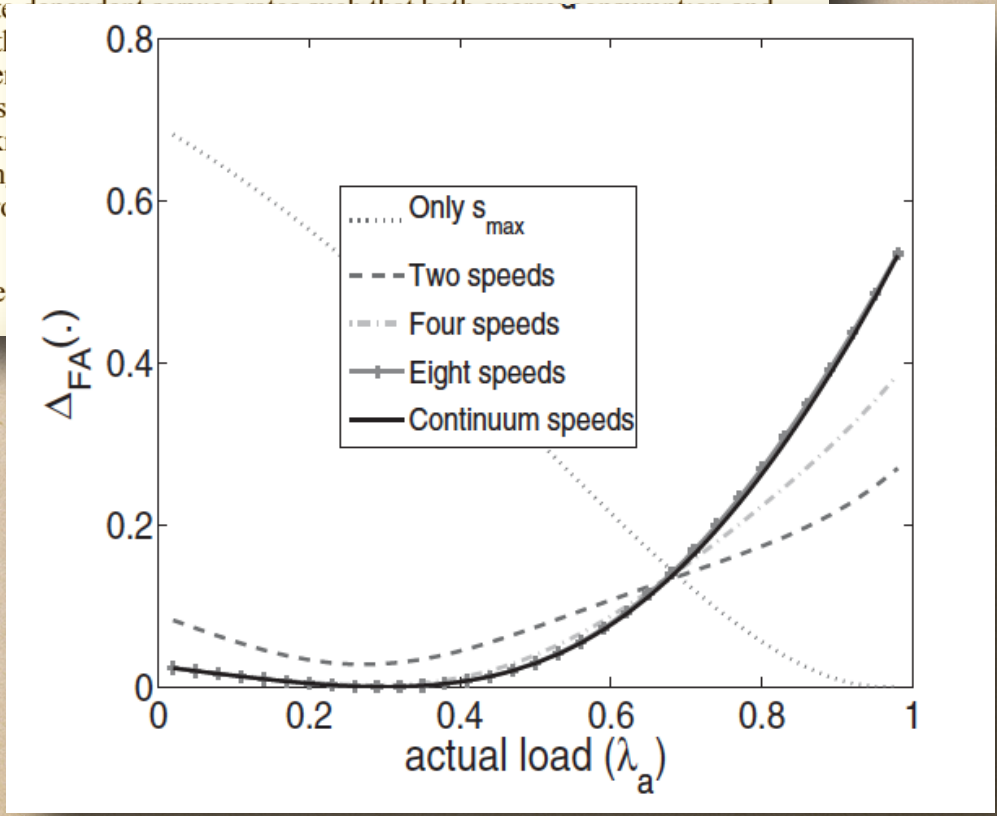
Tuan V. Dinh[a,*], Lachlan L. H. Andrew[a] and Yoni Nazarathy[b]

[a]*Swinburne University of Technology, Victoria, Australia;* [b]*The University of Queensland, Queensland, Australia*

(*Received 16 May 2012; final version received 18 October 2012*)

We consider single-pass, lossless, queueing systems at steady-state subject to Poisson job arrivals at an unknown rate. Service rates are allowed to depend on the number of jobs in the system, up to a fixed maximum, and power consumption is an increasing function of speed. The goal is to control the state-dependent service rates such that both energy consumption and delay are kept low. We consider a linear combination of th measure. We examine both the 'architecture' of the syste that the system can choose from, and the 'design' of the s work has illustrated that when the arrival rate is precisely k architectures, yet in view of parameter uncertainty, allowin the tradeoffs of architecture specification with respect to r measure which we call local robustness.

Keywords: parameter uncertainty; robust design; controlle



Predict

OPERATIONS RESEARCH
Vol. 59, No. 5, September–October 2011, pp. 1106–1118
ISSN 0030-364X | EISSN 1526-5463 | 11 | 5905 | 1106

http://dx.doi.org/10.1287/opre.1110.0974

Wait-Time Predictors for Customer Service Systems with Time-Varying Demand and Capacity

Rouba Ibrahim
Desautels Faculty of Management, McGill University, Montreal, Quebec H3A 1G5, Canada,
rouba.ibrahim@mail.mcgill.ca

Ward Whitt
Department of Industrial Engineering and Operations Research, Columbia University, New York, New York 10027,
ww2040@columbia.edu

Optimal Prediction of Queue Lengths and Delays in *GI/M/m* Multiserver Queues

C. MURRAY WOODSIDE
Carleton University, Ottawa, Canada

DAVID A. STANFORD
Bell-Northern Research Ltd., Ottawa, Canada

BERNARD PAGUREK
Carleton University, Ottawa, Canada

(Received December 1981; accepted December 1983)

We present optimal mean square predictors for queue lengths and delays in the stationary *GI/M/m* queue, based on a queue length measurement. The development specifies interrelationships among these predictors and numerical examples demonstrate basic properties of the predictors.

Probe

Estimation of available bandwidth for an $M/G/1$ queueing system

Seung Yeob Nam [a,*], Sunggon Kim [b], Dan Keun Sung [c]

[a] Department of Information and Communication Engineering, Yeungnam University, 214-1, Dae-dong, Gyeongsan 712-749, Republic of Korea
[b] Department of Information Statistics and RINS, Gyeongsang National University, Jinju 660-701, Republic of Korea
[c] School of Electrical Engineering and Computer Science, KAIST, Daejeon 305-701, Republic of Korea

ARTICLE INFO

Article history:
Received 19 July 2007
Received in revised form 15 October 2008
Accepted 22 October 2008
Available online 5 November 2008

Keywords:
Minimal-backlogging
Probing
Available bandwidth
$M/G/1$ queue

ABSTRACT

This paper is concerned with estimating the available bandwidth, the residual processing capacity, of a single-server queueing system whose service rate and the input traffic load are not known in advance. We analyze the minimal-backlogging method which has been proposed to estimate the available bandwidth by sending probing packets. We show that an $M/G/1$ queueing system probed by the minimal-backlogging method is stable. A consistent estimator of the inverse of available bandwidth is proposed based on the sojourn times of the probing packets. We also estimate the available bandwidth of a local server which is connected to the probing node with non-zero delay as an application of the theory developed for a single-server queue. We evaluate the accuracy of the proposed available bandwidth estimation scheme by simulation.

Queueing Syst (2009) 63: 59–107
DOI 10.1007/s11134-009-9150-9

INVITED PAPER

Inverse problems in queueing theory and Internet probing

F. Baccelli · B. Kauffmann · D. Veitch

Inferring Network Characteristics via Moment-Based Estimators

Sara Alouf, Philippe Nain

INRIA Sophia Antipolis, B.P. 93
06902, Sophia Antipolis Cedex, France
{salouf, nain}@sophia.inria.fr

Don Towsley

Dept. of Computer Science
University of Massachusetts
Amherst, MA 01003, USA
towsley@cs.umass.edu

Reconstructing arrival processes to $G/D/1$ queueing systems and tandem networks

Stephan Heckmüller, Bernd E. Wolfinger
Dept. of Computer Science, TKRN, University of Hamburg
Vogt-Kölln-Str. 30, D-22527 Hamburg
Tel: +49-40-42883-2350, Fax: +49-40-42883-2345
E-mail: {heckmueller|wolfinger}@informatik.uni-hamburg.de

Queueing Syst (2009) 61: 255–272
DOI 10.1007/s11134-009-9107-z

Determining an adequate probe separation for estimating the arrival rate in an M/D/1 queue using single-packet probing

A. Novak · R. Watson

Parameter Estimation for Partially Observed Queues

Thomas M. Chen, *Member, IEEE,* Jean Walrand, *Fellow, IEEE,* and David G. Messerschmitt, *Fellow, IEEE*

***Abstract*—In this paper, we consider parameter estimation for a FIFO queue with deterministic service times and two independent arrival streams of "observed" and "unobserved" packets. The arrivals of unobserved packets are Poisson with an unknown rate λ while the arrivals of observed packets are arbitrary. Maximum likelihood estimation of λ is formulated based on the arrival times and waiting times of k observed packets. The likelihood function is derived in terms of the transition probabilities of the unfinished work process which are calculated recursively. Sufficient conditions for consistency, asymptotic normality, and asymptotic efficiency are given. The mean and variance of the MLE are measured in simulation experiments. Numerical results indicate that the MLE is consistent and asymptotically normal.**

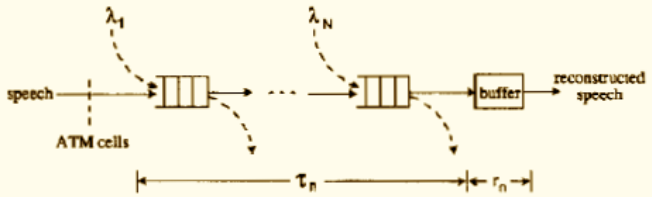


Fig. 1. ATM switched speech.

Maximum likelihood estimation of λ is formulated based on the data for k observed packets.

In Section II, the likelihood function is derived in terms of the transition probabilities of the unfinished work process;

Estimation of the traffic intensity in a piecewise-stationary $M_t/G_t/1$ queue with probing

Nelson Antunes[*] Gonçalo Jacinto[†] António Pacheco[‡] Cornelia Wichelhaus[§]

Assume Customers Don't Interact

Sojourn Time Estimation in an M/G/∞ Queue with Partial Information

Nafna Nelgabats * Yuval Nov † Gideon Weiss ‡

Adv. Appl. Prob. **48,** 1117–1138 (2016)
*doi:*10.1017/apr.2016.67

NONPARAMETRIC ESTIMATION OF THE SERVICE TIME DISTRIBUTION IN THE M/G/∞ QUEUE

ALEXANDER GOLDENSHLUGER,* *University of Haifa*

The $M/G/\infty$ estimation problem revisited

ALEXANDER GOLDENSHLUGER[1,*]

[1] *Department of Statistics, University of Haifa, Haifa 31905, Israel E-mail:* *goldensh@stat.haifa.ac.il

The subject of this paper is *the $M/G/\infty$ estimation problem*: the goal is to estimate the service time distribution G of the $M/G/\infty$ queue from the arrival–departure observations without identification of customers. We develop estimators of G and derive exact non–asymptotic expressions for their mean squared errors. The problem of estimating the service time expectation is addressed as well. We present some numerical results on comparison of different estimators of the service time distribution.

Ann. Inst. Statist. Math.
Vol. 51, No. 1, 71–97 (1999)

NON-PARAMETRIC ESTIMATION FOR THE $M/G/\infty$ QUEUE

N. H. BINGHAM[1] AND SUSAN M. PITTS[2]

[1] *Department of Mathematics and Statistics, Birkbeck College, University of London, Malet Street, London, WC1E 7HX, U.K.*
[2] *Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, 16 Mill Lane, Cambridge, CB2 1SB, U.K.*

(Received November 7, 1996; revised October 17, 1997)

J. R. Statist. Soc. B (2004)
66, *Part* 4, *pp.* 861–875

Nonparametric inference about service time distribution from indirect measurements

Peter Hall

Australian National University, Canberra, Australia

and Juhyun Park

University of North Carolina, Chapel Hill, USA

Use Queue Inference Engine

MANAGEMENT SCIENCE
Vol. 36, No. 5, May 1990
Printed in U.S.A.

THE QUEUE INFERENCE ENGINE: DEDUCING QUEUE STATISTICS FROM TRANSACTIONAL DATA*

RICHARD C. LARSON

Operations Research Center, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

The transactional data of a queueing system are the recorded times of service commencement and service completion for each customer served. With increasing use of computers to aid or even perform service one often has machine readable transactional data, but virtually no information about the queue itself. In this paper we propose a way to deduce the queueing behavior of Poisson arrival queueing systems from only the transactional data and the Poisson assumption. For each congestion period in which queues may form (in front of a single or multiple servers), the key quantities obtained are mean wait in queue, time-dependent mean number in queue, and probability distribution of the number in queue observed by a randomly arriving customer. The methodology builds on arguments of order statistics and usually requires a computer to evaluate a recursive function. The results are exact for a homogeneous Poisson arrival process (with unknown parameter) and approximately correct for a slowly time varying Poisson process.
(QUEUES; INFERENCE; DATA ANALYSIS; POISSON)

Operations Research Letters 38 (2010) 420–426

Contents lists available at ScienceDirect

Operations Research Letters

journal homepage: www.elsevier.com/locate/orl

Queue inference from periodic reporting data

Jesse C. Frey [a,*], Edward H. Kaplan [b,c,d]

Contents lists available at ScienceDirect

Computers & Industrial Engineering

journal homepage: www.elsevier.com/locate/caie

Analysis of an unobservable queue using arrival and departure times ☆

Jinsoo Park [a], Yun B. Kim [a,*], Thomas R. Willemain [b]

[a] Department of Systems Management Engineering, Sungkyunkwan University, Suwon, South Korea
[b] Department of Decision Science and Engineering Systems, Rensselaer Polytechnic Institute, Troy, NY, USA

Contents lists available at SciVerse ScienceDirect

Operations Research Letters

journal homepage: www.elsevier.com/locate/orl

Remarks on queue inference from departure data alone and the importance of the queue inference engine

Lee K. Jones

Department of Mathematical Sciences, University of Massachusetts Lowell, 1 University Ave., Lowell MA 01854, United States


Statistics and Computing **14**: 261–266, 2004

Filtering recursions for calculating likelihoods for queues based on inter-departure time data

PAUL FEARNHEAD

Department of Mathematics and Statistics, Lancaster University, LA1 4YF
p.fearnhead@lancaster.ac.uk

INFERRING BALKING BEHAVIOR FROM TRANSACTIONAL DATA

LEE K. JONES

Department of Mathematical Sciences, University of Massachusetts, Lowell, Massachusetts 01854, lee_jones@uml.edu

(Received March 1995; revisions received June 1996, December 1997; accepted January 1999)

Balking is the act of not joining a queue because the prospective arriving customer judges the queue to be too long. We analyze queues in the presence of balking, using only the service start and stop data utilized in Larson's Queue Inference Engine (Q.I.E.). Using an extension of Larson's congestion probability calculation to include balking we present new maximum likelihood, nonparametric, and Bayesian methods for inferring the arrival rate and balking functions. The methodology is applicable to businesses that wish to estimate lost sales because of balking arising from queuing-type congestion. The techniques are applied to a small transactional data set for illustrative purposes.

Estimating characteristics of queueing networks using transactional data

Avi Mandelbaum and Sergey Zeltyn

Faculty of Industrial Engineering and Management, Technion, Haifa 32000, Israel

Received 3 February 1997; revised 22 December 1997

MOMENT ESTIMATION OF CUSTOMER LOSS RATES FROM TRANSACTIONAL DATA

D.J. DALEY[1]
Australian National University
School of Mathematical Sciences
Canberra ACT 0200 Australia

L.D. SERVI
GTE Laboratories Incorporated
40 Sylvan Road
Waltham, MA 02254 USA

Inferring most likely queue length from transactional data

Dragomir D. Dimitrijevic*

Telogy Networks, Inc., 20250 Century Boulevard, German town, MD 20874, USA

Received 1 September 1992; revised 1 April 1996

Abstract

This paper presents an efficient algorithm for inferring a most likely length of an M/G/1 queue from its transactional data (service completion times) in a busy period. Practical application of queue-inferencing data (service completion times) in a busy period. Practical application of queue-inferencing algorithms may be found in evaluating servers such as automatic teller machines (ATM) or public telephones. These systems keep track on transactional data while the exact number of people waiting in line is not available. This important performance measure can be only estimated using partial information contained in transactional data. The efficiency of a queue-inferencing procedure is important since transactional data (e.g., ATM in a downtown area) may contain hundreds of transactional records. The developed procedure recursively infers a most likely scenario of length $r + 1$ from the previously inferred most likely scenario of length r. This property, and the algorithm's efficiency ($O(\log r)$ per iteration), make the algorithm an attactive solution for real-time estimation when the process under investigation is in progress and the actual busy sequence is not known yet. The total computational complexity is $O(n \log n)$ for a busy period sequence with n customers. Due to different queue-inferencing objective (most likely vs. mean queue length) and efficient organization of data structures, the developed procedure is faster than other queue-inferencing algorithms known from the available literature. The algorithm may be modified to use an additional information (if available) on upper limits of queue lengths in each service interval. In this case, the computation complexity is $O(n^2)$.

Keywords: Queue-inferencing; Transactional data; M/G/1 queue; Invisible queue

DEDUCING QUEUEING FROM TRANSACTIONAL DATA: THE QUEUE INFERENCE ENGINE, REVISITED

DIMITRIS J. BERTSIMAS

Massachusetts Institute of Technology, Cambridge, Massachusetts

L. D. SERVI

GTE Laboratories, Incorporated, Waltham, Massachusetts

(Received April 1990; revision received February 1991; accepted July 1991)

R. Larson proposed a method to statistically infer the expected transient queue length during a busy period in $O(n^5)$ solely from the n starting and stopping times of each customer's service during the busy period and assuming the arrival distribution is Poisson. We develop a new $O(n^3)$ algorithm which uses these data to deduce transient queue lengths as well as the waiting times of each customer in the busy period. We also develop an $O(n)$ on-line algorithm to dynamically update the current estimates for queue lengths after each departure. Moreover, we generalize our algorithms for the case of a time-varying Poisson process and also for the case of i.i.d. interarrival times with an arbitrary distribution. We report computational results that exhibit the speed and accuracy of our algorithms.

Consider Likelihood

Queueing Systems 9 (1991) 301–312

Estimation for a class of simple queueing networks

D. Thiruvaiyaru and I.V. Basawa

Department of Statistics, University of Georgia, Athens, GA 30602, USA

U.N. Bhat

Department of Statistics, Southern Methodist University, Dallas, TX 75275, USA

Received 21 May 1990; revised 22 March 1991

Maximum likelihood estimators of the parameters of an open Jackson network are derived, and their joint asymptotic normality is established. The problem of estimation for tandem queues is discussed as a special case of the Jackson system. These results are valid when the system is not necessarily in equilibrium.

Technical Note: Traffic Intensity Estimation

Peter C. Kiessler, Robert Lund

Department of Mathematical Sciences, Clemson University, Clemson, South Carolina 29634-0975

Comparison on five estimation approaches of intensity for a queueing system with short run

Jau-Chuan Ke · Yunn-Kuang Chu

Available online at www.sciencedirect.com
ScienceDirect
Statistics and Probability Letters 78 (2008) 1375–1383
STATISTICS & PROBABILITY LETTERS
www.elsevier.com/locate/stapro

Parameter estimation using partial information with applications to queueing and related models

I.V. Basawa[a,*], U.N. Bhat[b], J. Zhou[c]

[a] *Department of Statistics, University of Georgia, Athens, Georgia 30602, USA*
[b] *Department of Statistical Science and Operations Research, Southern Methodist University, Dallas, Texas 75275, USA*
[c] *Mathematical Bioscience Institute, Ohio State University, Columbus, Ohio 43212, USA*

Received 26 September 2005; received in revised form 27 September 2007; accepted 11 December 2007
Available online 31 December 2007

LARGE SAMPLE INFERENCE FROM SINGLE SERVER QUEUES *

I.V. BASAWA

Department of Statistics, University of Georgia, Athens, GA 30602, U.S.A.

and

N.U. PRABHU

Mathematical Sciences Institute, Cornell University, Caldwell Hall, Ithaca, NY 14853, U.S.A.

Received: 16 April 1987
Revised: 17 March 1988

Abstract

Problems of large sample estimation and tests for the parameters in a single server queue are discussed. The service time and the interarrival time densities are assumed to belong to (positive) exponential families. The queueing system is observed over a continuous time interval $(0, T]$ where T is determined by a suitable stopping rule. The limit distributions of the estimates are obtained in a unified setting, and without imposing the ergodicity condition on the queue length process. Generalized linear models, in particular, log-linear models are considered when several independent queues are observed. The mean service times and the mean interarrival times after appropriate transformations are assumed to satisfy a linear model involving unknown parameters of interest, and known covariates. These models enhance the scope and the usefulness of the standard queueing systems.

Keywords: Single server queues, maximum likelihood, stopping times, exponential families, generalized linear models, tests of fit, asymptotic inference.

Large Sample Inference in Retrial Queues

A. Rodrigo and M. Vázquez
Departamento de Analisis Economico
Facultad de Ciencias Económicas
Universidad Complutense de Madrid
Campus de Somosaguas, 28223 Madrid, Spain
ececo03@sis.ucm.es

Abstract—We analyze retrial queueing systems from a statistical viewpoint. We consider a general $G/G/1$ retrial queue where the flow of repeated attempts can be non-Markovian. We observe the operation of the system over a random time interval $(0, T]$ focusing on the case where T is a departure epoch. We obtain some asymptotic relationship among the random variables involved, even when the system in nonergodic. This allows us to get asymptotically Gaussian consistent estimators for the unknown parameters in a parametric context.

Maximum Likelihood Estimation of Closed Queueing Network Demands from Queue Length Data

Weikun Wang, Giuliano Casale
Department of Computing
Imperial College London, UK
{weikun.wang11, giuliano.casale}@imperial.ac.uk*

Be Bayesian

The Annals of Applied Statistics
2010, Vol. 4, No. 3, 1533–1557
DOI: 10.1214/10-AOAS336

BAYESIAN INFERENCE FOR DOUBLE PARETO LOGNORMAL QUEUES

BY PEPA RAMIREZ-COBO[1], ROSA E. LILLO,
SIMON WILSON AND MICHAEL P. WIPER

CNRS France, Universidad Carlos III de Madrid, Trinity College Dublin and Universidad Carlos III de Madrid

In this article we describe a method for carrying out Bayesian estimation for the double Pareto lognormal ($dPlN$) distribution which has been proposed as a model for heavy-tailed phenomena. We apply our approach to estimate the $dPlN/M/1$ and $M/dPlN/1$ queueing systems. These systems cannot be analyzed using standard techniques due to the fact that the $dPlN$ distribution does not possess a Laplace transform in closed form. This difficulty is overcome using some recent approximations for the Laplace transform of the interarrival distribution for the $Pareto/M/1$ system. Our procedure is illustrated with applications in internet traffic analysis and risk theory.

The Annals of Applied Statistics
2011, Vol. 5, No. 1, 254–282
DOI: 10.1214/10-AOAS392

BAYESIAN INFERENCE FOR QUEUEING NETWORKS AND MODELING OF INTERNET SERVICES

BY CHARLES SUTTON AND MICHAEL I. JORDAN

University of Edinburgh and University of California

Modern Internet services, such as those at Google, Yahoo!, and Amazon, handle billions of requests per day on clusters of thousands of computers. Because these services operate under strict performance requirements, a statistical understanding of their performance is of great practical interest. Such services are modeled by networks of queues, where each queue models one of the computers in the system. A key challenge is that the data are incomplete, because recording detailed information about every request to a heavily used system can require unacceptable overhead. In this paper we develop a Bayesian perspective on queueing models in which the arrival and departure times that are not observed are treated as latent variables. Underlying this viewpoint is the observation that a queueing model defines a deterministic transformation between the data and a set of independent variables called the service times. With this viewpoint in hand, we sample from the posterior distribution over missing data and model parameters using Markov chain Monte Carlo. We evaluate our framework on data from a benchmark Web application. We also present a simple technique for selection among nested queueing models. We are unaware of any previous work that considers inference in networks of queues in the presence of missing data.

Queueing Systems 15 (1994) 419–426

Short communication

Bayesian inference in Markovian queues

Carmen Armero

Department of Statistics and Operations Research, University of Valencia, Doctor Moliner 50, 46100 Burjassot, Valencia, Spain

Received 25 January 1992; revised 19 January 1993

This paper is concerned with the Bayesian analysis of general queues with Poisson input and exponential service times. Joint posterior distribution of the arrival rate and the individual service rate is obtained from a sample consisting in n observations of the interarrival process and m complete service times. Posterior distribution of traffic intensity in $M/M/c$ is also obtained and the statistical analysis of the ergodic condition from a decision point of view is discussed.

Keywords: Bayesian analysis; ergodic condition; $M/M/1$, $M/M/c$, $M/M/\infty$, $M/M/1/k$ and $M/M/c/k$ queues; traffic intensity.

Solve an Inverse Problem

Nonparametric inference from the $M/G/1$ workload

MARTIN B. HANSEN* and SUSAN M. PITTS**

* *Department of Mathematical Sciences, Aalborg University, Fredrik Bajers Vej 7G, DK-9220 Aalborg Ø, Denmark. Email: mbh@math.aau.dk*

** *Statistical Laboratory, Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WB, UK. Email: S.Pitts@statslab.cam.ac.uk*

Summary. Consider an $M/G/1$ queue with unknown service-time distribution and unknown traffic intensity ρ. Given systematically sampled observations of the workload, we construct estimators of ρ and of the service-time distribution function, and we study asymptotic properties of these estimators.

Keywords: Asymptotic normality; Empirical processes; Functional central limit theorem; Infinite dimensional delta method; $M/G/1$-queue; Regenerative processes; Workload

Queueing Syst (2009) 63: 59–107
DOI 10.1007/s11134-009-9150-9

INVITED PAPER

Inverse problems in queueing theory and Internet probing

F. Baccelli · B. Kauffmann · D. Veitch

Inverse Problems in Bandwidth Sharing Networks

Bruno Kauffmann
Orange Labs, France
bruno.kauffmann@orange.com

Nonparametric inference from $M/G/1$ busy periods

N. H. BINGHAM
Department of Statistics,
Birkbeck College (University of London)
Malet Street, London WC1E 7HX, UK

Susan M. PITTS
Statistical Laboratory, 16 Mill Lane
Cambridge CB2 1SB, UK

Abstract

Consider an $M/G/1$ queue with unknown service-time distribution and unknown Poisson arrival rate λ. Given observations of the busy and idle periods of this queue, we construct estimators of λ and of the service-time moment generating function, and we study asymptotic properties of these estimators.

The Annals of Statistics
1994, Vol. 22, No. 3, 1428–1446

NONPARAMETRIC ESTIMATION OF THE STATIONARY WAITING TIME DISTRIBUTION FUNCTION FOR THE $GI/G/1$ QUEUE

BY SUSAN M. PITTS

University College London

The $GI/G/1$ queueing model is regarded as a functional that maps the service and interarrival time distribution functions onto the stationary waiting time distribution function. By considering the output of the functional when it is applied to nonparametric estimators of the input distribution functions, we obtain a nonparametric estimator of the stationary waiting time distribution function. Using appropriate continuity and differentiability properties of the functional, we show that statistical properties of the input estimators carry over to corresponding properties for the stationary waiting time distribution function estimator.

Handle Discrete Sampling

Statistical inference for discretely observed Markov jump processes.

Mogens Bladt
IIMAS–UNAM
A.P. 20-726
01000 Mexico, D.F.
Mexico
bladt@sigma.iimas.unam.mx

Michael Sørensen
Department of Applied Mathematics and Statistics
University of Copenhagen
Universitetsparken 5
DK-2100 Copenhagen Ø
Denmark
michael@math.ku.dk

Queueing Syst (2007) 55:131–138
DOI 10.1007/s11134-006-9009-2

Estimation for queues from queue length data

J. V. Ross · T. Taimre · P. K. Pollett

Convergence rates of Laplace-transform based estimators

Arnoud V. den Boer[1]*, Michel Mandjes[2,3,4]

Use Queueing Fundamentals

INDIRECT ESTIMATION VIA $L = \lambda W$

PETER W. GLYNN
Stanford University, Stanford, California

WARD WHITT
AT&T Bell Laboratories, Murray Hill, New Jersey

(Received September 1985; revisions received December 1986, October 1987; accepted November 1987)

For a large class of queueing systems, Little's law ($L = \lambda W$) helps provide a variety of statistical estimators for the long-run time-average queue length L and the long-run customer-average waiting time W. We apply central limit theorem versions of Little's law to investigate the asymptotic efficiency of these estimators. We show that an indirect estimator for L using the natural estimator for W plus the known arrival rate λ is more efficient than a direct estimator for L, provided that the interarrival and waiting times are negatively correlated, thus extending a variance-reduction principle for the GI/G/s model due to A. M. Law and J. S. Carson. We also introduce a general framework for indirect estimation which can be applied to other problems besides $L = \lambda W$. We show that the issue of indirect-versus-direct estimation is related to estimation using nonlinear control variables. We also show, under mild regularity conditions, that any nonlinear control-variable scheme is equivalent to a linear control-variable scheme from the point of view of asymptotic efficiency. Finally, we show that asymptotic bias is typically asymptotically negligible compared to asymptotic efficiency.

ESTIMATING WAITING TIMES WITH THE TIME-VARYING LITTLE'S LAW

Song-Hee Kim and Ward Whitt

Industrial Engineering and Operations Research
Columbia University
New York, NY, 10027
{sk3116, ww2040}@columbia.edu

June 13, 2012

Statistical Analysis with Little's Law

Song-Hee Kim and Ward Whitt
Department of Industrial Engineering and Operations Research, Columbia University, New York, NY 10027-6699,
{sk3116,ww2040}@columbia.edu, http://www.columbia.edu/~{sk3116,ww2040}

The theory supporting Little's law ($L = \lambda W$) is now well developed, applying to both limits of averages and expected values of stationary distributions, but applications of Little's law with actual system data involve measurements over a finite time interval, which are neither of these. We advocate taking a statistical approach with such measurements. We investigate how estimates of L and λ can be used to estimate W when the waiting times are not observed. We show how to estimate and remove bias due to interval edge effects when the system does not begin and end empty. We advocate estimating confidence intervals. Given a single sample path segment, we suggest estimating confidence intervals using the method of batch means, as is often done in stochastic simulation output analysis. We illustrate the methods with data from a call center and simulation experiments.

ESTIMATING CUSTOMER AND TIME AVERAGES

PETER W. GLYNN
Stanford University, Stanford, California

BENJAMIN MELAMED
NEC Research USA, Inc., Princeton, New Jersey

WARD WHITT
AT&T Bell Laboratories, Murray Hill, New Jersey

(Received September 1990; revisions received January, November 1991; accepted November 1991)

In this paper we establish a joint central limit theorem for customer and time averages by applying a martingale central limit theorem in a Markov framework. The limiting values of the two averages appear in the translation terms. This central limit theorem helps to construct confidence intervals for estimators and perform statistical tests. It thus helps determine which finite average is a more asymptotically efficient estimator of its limit. As a basis for testing for PASTA (Poisson arrivals see time averages), we determine the variance constant associated with the central limit theorem for the difference between the two averages when PASTA holds.

Contents lists available at SciVerse ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Fitting birth-and-death queueing models to data

Ward Whitt

Department of Industrial Engineering and Operations Research, Columbia University, New York, NY 10027-6699, USA

ARTICLE INFO

Article history:
Received 6 January 2012
Received in revised form 11 February 2012
Accepted 12 February 2012
Available online 17 February 2012

Keywords:
Birth-and-death processes
Empirical birth-and-death processes
Fitting birth-and-death processes to data
Conservation laws
Operational analysis

ABSTRACT

Given measurements of the number of customers in a queueing system over a finite time interval, it is natural to try to fit a stationary birth-and-death process model, because it is remarkably tractable, even when the birth and death rates depend on the state in an arbitrary way. Natural estimators of the birth (death) rate in each state are the observed number of transitions up (down) from that state divided by the total time spent in that state. It is tempting to validate the model by comparing the steady-state distribution of the model based on those estimated rates to the empirical steady-state distribution recording the proportion of time spent in each state. However, it is inappropriate to draw strong conclusions from a close fit to the same data, because these two distributions are necessarily intimately related, even if the model assumptions are not nearly satisfied. We elaborate by (i) establishing stochastic comparisons between these two fitted distributions using likelihood-ratio stochastic ordering and (ii) quantifying their difference.

Using a Birth-and-Death Process to Estimate the Steady-State Distribution of a Periodic Queue

James Dong,[1] Ward Whitt[2]

Queueing Syst (2015) 79:391–426
DOI 10.1007/s11134-014-9429-3

Stochastic grey-box modeling of queueing systems: fitting birth-and-death processes to data

James Dong · Ward Whitt

Chapter 16

Parametric estimation of tail probabilities for the single-server queue[1]

Peter W. Glynn and Marcelo Torres

ABSTRACT In this chapter, we consider the question of how long the arrival process to the single-server queue needs to be observed in order to accurately estimate the long-run fraction of time that the workload exceeds y. We assume that the arrival process can be modeled parametrically. In such a parametric context, our results suggest that one typically needs to observe the arrival process over a time horizon that is large relative to y^2. This conclusion appears to hold regardless of whether the arrival process model exhibits complex dependencies or not.

CONTENTS

16.1 Introduction 449
16.2 Parametric estimation of tail probabilities for the M/M/1 queue 450
16.3 Parametric estimation of tail probabilities for Markov modulated queues 454
16.4 Further asymptotic analysis 461
16.5 Open research problems 462
Bibliography 462

Mine

Inferring Network Characteristics via Moment-Based Estimators

Sara Alouf, Philippe Nain

INRIA Sophia Antipolis, B.P. 93
06902, Sophia Antipolis Cedex, France
{salouf, nain}@sophia.inria.fr

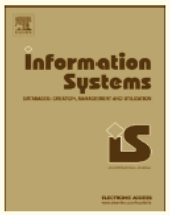
Don Towsley

Dept. of Computer Science
University of Massachusetts
Amherst, MA 01003, USA
towsley@cs.umass.edu

Contents lists available at ScienceDirect

Information Systems

journal homepage: www.elsevier.com/locate/infosys



Queue mining for delay prediction in multi-class service processes

Arik Senderovich [a,*], Matthias Weidlich [b], Avigdor Gal [a], Avishai Mandelbaum [a]

[a] Technion - Israel Institute of Technology, Haifa, Israel
[b] Imperial College London, London, United Kingdom

Workload Service Requirements Analysis: A Queueing Network Optimization Approach

Li Zhang, Cathy H. Xia, Mark S. Squillante, W. Nathaniel Mills III
IBM Thomas J. Watson Research Center
Yorktown Heights, NY 10598
{zhangli,hhcx,mss,wnm3}@us.ibm.com

Thank You

Parameter and State Estimation in Queues and Related Stochastic Models: A Bibliography

Azam Asanjarani, Yoni Nazarathy and Philip K. Pollett
The School of Mathematics and Physics, The University of Queensland,
Brisbane Australia

January 31, 2017