Extremal Queueing Theory

Yan Chen

Advisor: Ward Whitt
Ph.D. Qualification Research Presentation
Department of Industrial Engineering and Operations Research

September 28, 2018

Overview

- Motivation: Queueing Network Analyzer in Whitt (1983).
- Objective: Develop a series of theoretical and numerical methodologies to explore

  "Extremal Queues Given Partial Information".
- Important Idea: Focus on *Queues* instead of *Formulas*.

Overview

- Motivation: Queueing Network Analyzer in Whitt (1983).
- Objective: Develop a series of theoretical and numerical methodologies to explore
  "Extremal Queues Given Partial Information".
- Important Idea: Focus on *Queues* instead of *Formulas*.

Overview

- Motivation: Queueing Network Analyzer in Whitt (1983).
- Objective: Develop a series of theoretical and numerical methodologies to explore
  "Extremal Queues Given Partial Information".
- Important Idea: Focus on *Queues* instead of *Formulas*.

Overview

- Motivation: Queueing Network Analyzer in Whitt (1983).
- Objective: Develop a series of theoretical and numerical methodologies to explore
  "Extremal Queues Given Partial Information".
- Important Idea: Focus on *Queues* instead of *Formulas*.

The $GI/GI/1$ Model

We consider $GI/GI/1$ queue,

- a) unlimited waiting room and FCFS discipline;
- b) given mean and variance of F and G;
- c) F over bounded support $[0, M_a]$ and G over $[0, M_s]$;
- d) traffic level < 1.

Let W_n be n-th transient waiting time, assuming that the system starts empty with $W_0 = 0$.The sequence $\{W_n : n \geq 0\}$ is well known to satisfy the *Lindley* recursion

$$W_{n+1} = [W_n + V_n - U_n]^+, \quad n \geq 0. \tag{1}$$

Let $W = W_\infty$ denote the steady-state waiting time.

The $GI/GI/1$ Model

We consider $GI/GI/1$ queue,

- a) unlimited waiting room and FCFS discipline;
- b) given mean and variance of F and G;
- c) F over bounded support $[0, M_a]$ and G over $[0, M_s]$;
- d) traffic level < 1.

Let W_n be n-th transient waiting time, assuming that the system starts empty with $W_0 = 0$.The sequence $\{W_n : n \geq 0\}$ is well known to satisfy the *Lindley* recursion

$$W_{n+1} = [W_n + V_n - U_n]^+, \quad n \geq 0. \tag{1}$$

Let $W = W_\infty$ denote the steady-state waiting time.

The $GI/GI/1$ Model

We consider $GI/GI/1$ queue,

- a) unlimited waiting room and FCFS discipline;
- b) given mean and variance of F and G;
- c) F over bounded support $[0, M_a]$ and G over $[0, M_s]$;
- d) traffic level< 1.

Let W_n be n-th transient waiting time, assuming that the system starts empty with $W_0 = 0$.The sequence $\{W_n : n \geq 0\}$ is well known to satisfy the *Lindley* recursion

$$W_{n+1} = [W_n + V_n - U_n]^+, \quad n \geq 0. \tag{1}$$

Let $W = W_\infty$ denote the steady-state waiting time.

The $GI/GI/1$ Model

We consider $GI/GI/1$ queue,

- a) unlimited waiting room and FCFS discipline;
- b) given mean and variance of F and G;
- c) F over bounded support $[0, M_a]$ and G over $[0, M_s]$;
- d) traffic level< 1.

Let W_n be n-th transient waiting time, assuming that the system starts empty with $W_0 = 0$.The sequence $\{W_n : n \geq 0\}$ is well known to satisfy the *Lindley* recursion

$$W_{n+1} = [W_n + V_n - U_n]^+, \quad n \geq 0. \tag{1}$$

Let $W = W_\infty$ denote the steady-state waiting time.

The $GI/GI/1$ Model

We consider $GI/GI/1$ queue,

- a) unlimited waiting room and FCFS discipline;
- b) given mean and variance of F and G;
- c) F over bounded support $[0, M_a]$ and G over $[0, M_s]$;
- d) traffic level< 1.

Let W_n be n-th transient waiting time, assuming that the system starts empty with $W_0 = 0$.The sequence $\{W_n : n \geq 0\}$ is well known to satisfy the *Lindley* recursion

$$W_{n+1} = [W_n + V_n - U_n]^+, \quad n \geq 0. \tag{1}$$

Let $W = W_\infty$ denote the steady-state waiting time.

Classical $GI/GI/1$ Problem

We want to answer a Long Standing Open Problem in Queueing Theory.

- Given any G, what is the extremal inter-arrival time dist F^* attaining the UB and LB of $\mathbb{E}[W(F/G/1)]$?
- Given any F, what is the extremal service time dist G^* attaining the UB and LB of $\mathbb{E}[W(F/G/1)]$?
- What are the extremal F^* and G^* leading to overall UB and LB of $\mathbb{E}[W(F/G/1)]$?

Classical $GI/GI/1$ Problem

We want to answer a Long Standing Open Problem in Queueing Theory.

- Given any G, what is the extremal inter-arrival time dist F^* attaining the UB and LB of $\mathbb{E}[W(F/G/1)]$?
- Given any F, what is the extremal service time dist G^* attaining the UB and LB of $\mathbb{E}[W(F/G/1)]$?
- What are the extremal F^* and G^* leading to overall UB and LB of $\mathbb{E}[W(F/G/1)]$?

Classical $GI/GI/1$ Problem

We want to answer a Long Standing Open Problem in Queueing Theory.

- Given any G, what is the extremal inter-arrival time dist F^* attaining the UB and LB of $\mathbb{E}[W(F/G/1)]$?
- Given any F, what is the extremal service time dist G^* attaining the UB and LB of $\mathbb{E}[W(F/G/1)]$?
- What are the extremal F^* and G^* leading to overall UB and LB of $\mathbb{E}[W(F/G/1)]$?

Classical $GI/GI/1$ Problem

We want to answer a Long Standing Open Problem in Queueing Theory.

- Given any G, what is the extremal inter-arrival time dist F^* attaining the UB and LB of $\mathbb{E}[W(F/G/1)]$?
- Given any F, what is the extremal service time dist G^* attaining the UB and LB of $\mathbb{E}[W(F/G/1)]$?
- What are the extremal F^* and G^* leading to overall UB and LB of $\mathbb{E}[W(F/G/1)]$?

Main Theoretical Results

Under the classical settings of $GI/GI/1$ model,

- Given any G service time dist, $F^*(G)$ is a three-point dist.
- Given any F inter-arrival time dist, $G^*(F)$ is a three-point dist.
- Under some regularity conditions, $F^*(G)$, $G^*(F)$ are two-point dists.

Remark

We also study the transient mean waiting time $\mathbb{E}[W_n]$ for $n \geq 1$ and unbounded support.

Main Theoretical Results

Under the classical settings of $GI/GI/1$ model,

- Given any G service time dist, $F^*(G)$ is a three-point dist.
- Given any F inter-arrival time dist, $G^*(F)$ is a three-point dist.
- Under some regularity conditions, $F^*(G)$, $G^*(F)$ are two-point dists.

Remark

We also study the transient mean waiting time $\mathbb{E}[W_n]$ for $n \geq 1$ and unbounded support.

Main Theoretical Results

Under the classical settings of $GI/GI/1$ model,

- Given any G service time dist, $F^*(G)$ is a three-point dist.
- Given any F inter-arrival time dist, $G^*(F)$ is a three-point dist.
- Under some regularity conditions, $F^*(G)$, $G^*(F)$ are two-point dists.

Remark

We also study the transient mean waiting time $\mathbb{E}[W_n]$ for $n \geq 1$ and unbounded support.

Main Theoretical Results

Under the classical settings of $GI/GI/1$ model,

- Given any G service time dist, $F^*(G)$ is a three-point dist.
- Given any F inter-arrival time dist, $G^*(F)$ is a three-point dist.
- Under some regularity conditions, $F^*(G), G^*(F)$ are two-point dists.

Remark

We also study the transient mean waiting time $\mathbb{E}[W_n]$ for $n \geq 1$ and unbounded support.

Main Theoretical Results

Under the classical settings of $GI/GI/1$ model,

- Given any G service time dist, $F^*(G)$ is a three-point dist.
- Given any F inter-arrival time dist, $G^*(F)$ is a three-point dist.
- Under some regularity conditions, $F^*(G)$, $G^*(F)$ are two-point dists.

Remark

We also study the transient mean waiting time $\mathbb{E}[W_n]$ for $n \geq 1$ and unbounded support.

Summary for Extremal $GI/GI/1$ Queues I

Overall Upper Bound Inequalities:

$$
\begin{aligned}
\mathbb{E}[W(F/G/1)] &\leq \mathbb{E}[W(F_0/G_{u^*}/1)] && (2)\\
&\leq \frac{2(1-\rho)\rho/(1-\delta)c_a^2 + \rho^2 c_s^2}{2(1-\rho)} && (3)\\
&< \text{Daley's Bound} < \text{Kingman's Bound.} && (4)
\end{aligned}
$$

$\delta \in (0,1)$ and $\delta = \exp(-(1-\delta)/\rho)$.

Overall Lower Bound Inequalities:

$$
\begin{aligned}
\mathbb{E}[W(F/G/1)] &\geq \mathbb{E}[W(F_u/A_3/1)] && (5)\\
&\geq \mathbb{E}[W(D/A_3/1)] = \text{Ott's Bound.} && (6)
\end{aligned}
$$

Summary for Extremal $GI/GI/1$ Queues I

Overall Upper Bound Inequalities:

$$\begin{aligned}
\mathbb{E}[W(F/G/1)] &\leq \mathbb{E}[W(F_0/G_{u^*}/1)] && (2)\\
&\leq \frac{2(1-\rho)\rho/(1-\delta)c_a^2 + \rho^2 c_s^2}{2(1-\rho)} && (3)\\
&< \text{Daley's Bound} < \text{Kingman's Bound.} && (4)
\end{aligned}$$

$\delta \in (0,1)$ and $\delta = \exp(-(1-\delta)/\rho)$.

Overall Lower Bound Inequalities:

$$\begin{aligned}
\mathbb{E}[W(F/G/1)] &\geq \mathbb{E}[W(F_u/A_3/1)] && (5)\\
&\geq \mathbb{E}[W(D/A_3/1)] = \text{Ott's Bound.} && (6)
\end{aligned}$$

Summary for Extremal $GI/GI/1$ Queues I

Overall Upper Bound Inequalities:

$$\begin{aligned}
\mathbb{E}[W(F/G/1)] &\leq \mathbb{E}[W(F_0/G_{u^*}/1)] && (2)\\
&\leq \frac{2(1-\rho)\rho/(1-\delta)c_a^2+\rho^2 c_s^2}{2(1-\rho)} && (3)\\
&< \text{Daley's Bound} < \text{Kingman's Bound}. && (4)
\end{aligned}$$

$\delta \in (0,1)$ and $\delta = \exp(-(1-\delta)/\rho)$.

Overall Lower Bound Inequalities:

$$\begin{aligned}
\mathbb{E}[W(F/G/1)] &\geq \mathbb{E}[W(F_u/A_3/1)] && (5)\\
&\geq \mathbb{E}[W(D/A_3/1)] = \text{Ott's Bound}. && (6)
\end{aligned}$$

Summary for Extremal $GI/GI/1$ Queues I

Overall Upper Bound Inequalities:

$$\begin{aligned}
\mathbb{E}[W(F/G/1)] &\leq \mathbb{E}[W(F_0/G_{u^*}/1)] && (2)\\
&\leq \frac{2(1-\rho)\rho/(1-\delta)c_a^2 + \rho^2 c_s^2}{2(1-\rho)} && (3)\\
&< \text{Daley's Bound} < \text{Kingman's Bound.} && (4)
\end{aligned}$$

$\delta \in (0,1)$ and $\delta = \exp(-(1-\delta)/\rho)$.

Overall Lower Bound Inequalities:

$$\begin{aligned}
\mathbb{E}[W(F/G/1)] &\geq \mathbb{E}[W(F_u/A_3/1)] && (5)\\
&\geq \mathbb{E}[W(D/A_3/1)] = \text{Ott's Bound.} && (6)
\end{aligned}$$

Summary for Extremal $GI/GI/1$ Queues II

Table: A comparison of the bounds and approximations for the scaled steady-state mean $(1-\rho)E[W]/\rho^2$ in the $GI/GI/1$ model as a function of ρ for the case $c_a^2 = c_s^2 = 4.0$.

ρ	Tight LB	HTA	Tight UB (2)	New UB (3)	δ	MRE	Daley	Kingman
0.10	0.000	4.000	37.989	38.002	0.000	0.0%	40.000	202.000
0.20	0.000	4.000	18.080	18.112	0.007	0.2%	20.000	52.000
0.30	0.000	4.000	11.661	11.731	0.041	0.6%	13.333	24.222
0.40	0.000	4.000	8.641	8.722	0.107	0.9%	10.000	14.500
0.50	0.500	4.000	6.941	7.020	0.203	1.1%	8.000	10.000
0.60	1.111	4.000	5.884	5.946	0.324	1.1%	6.667	7.556
0.70	1.480	4.000	5.168	5.216	0.467	0.9%	5.714	6.082
0.80	1.719	4.000	4.662	4.693	0.629	0.7%	5.000	5.125
0.90	1.883	4.000	4.287	4.302	0.807	0.4%	4.444	4.469
0.95	1.946	4.000	4.134	4.142	0.902	0.2%	4.211	4.216
0.98	1.979	4.000	4.052	4.055	0.960	0.1%	4.082	4.082
0.99	1.990	4.000	4.025	4.027	0.980	0.0%	4.040	4.041

Summary for Effective Algorithms III

We develop effective Numerical and Simulation algorithms in Extremal Queues.

- Queueing Reduction:
  1) Deterministic Arrival Batches: $F_0/G_{u^*}/1$ to $D/RS(V,p)/1$.
  2) Daley's Decomposition for Service: $F_0/G_{u^*}/1$ to $F_0/D/1$.
- Numerical Alg: Negative Binomial (NB) & Discrete Time Markov Chain (DTMC).
- Simulation Alg: idle-time simulation algorithm.

Summary for Effective Algorithms III

We develop effective Numerical and Simulation algorithms in Extremal Queues.

- Queueing Reduction:
  1) Deterministic Arrival Batches: $F_0/G_{u^*}/1$ to $D/RS(V,p)/1$.
  2) Daley's Decomposition for Service: $F_0/G_{u^*}/1$ to $F_0/D/1$.
- Numerical Alg: Negative Binomial (NB) & Discrete Time Markov Chain (DTMC).
- Simulation Alg: idle-time simulation algorithm.

Summary for Effective Algorithms III

We develop effective Numerical and Simulation algorithms in Extremal Queues.

- Queueing Reduction:
  1) Deterministic Arrival Batches: $F_0/G_{u^*}/1$ to $D/RS(V,p)/1$.
  2) Daley's Decomposition for Service: $F_0/G_{u^*}/1$ to $F_0/D/1$.
- Numerical Alg: Negative Binomial (NB) & Discrete Time Markov Chain (DTMC).
- Simulation Alg: idle-time simulation algorithm.

Summary for Effective Algorithms III

We develop effective Numerical and Simulation algorithms in Extremal Queues.

- Queueing Reduction:
  1) Deterministic Arrival Batches: $F_0/G_{u^*}/1$ to $D/RS(V,p)/1$.
  2) Daley's Decomposition for Service: $F_0/G_{u^*}/1$ to $F_0/D/1$.
- Numerical Alg: Negative Binomial (NB) & Discrete Time Markov Chain (DTMC).
- Simulation Alg: idle-time simulation algorithm.

The Possible Best Results

Theorem

(*Counterexamples*) *Fix any service time dist* G, $F^*(G) = F_0$; *Fix any inter-arrival dist* F, $G^*(F)$ *is* G_0 *or* G_u. *The both arguments are invalid.*

Conjecture

(*Chen and Whitt I*) *Fix any* G, *the extremal* $F^*(G)$ *is a two-point distribution.*

Conjecture

(*Chen and Whitt II*) *Fix any two-point* F, *the extremal* $G^*(F)$ *is a two-point distribution.*

The Possible Best Results

Theorem

(*Counterexamples*) *Fix any service time dist* G, $F^*(G) = F_0$; *Fix any inter-arrival dist* F, $G^*(F)$ *is* G_0 *or* G_u. *The both arguments are invalid.*

Conjecture

(*Chen and Whitt I*) *Fix any* G, *the extremal* $F^*(G)$ *is a two-point distribution.*

Conjecture

(*Chen and Whitt II*) *Fix any two-point* F, *the extremal* $G^*(F)$ *is a two-point distribution.*

The Possible Best Results

Theorem

(Counterexamples) Fix any service time dist G, $F^*(G) = F_0$; Fix any inter-arrival dist F, $G^*(F)$ is G_0 or G_u. The both arguments are invalid.

Conjecture

(Chen and Whitt I) Fix any G, the extremal $F^*(G)$ is a two-point distribution.

Conjecture

(Chen and Whitt II) Fix any two-point F, the extremal $G^*(F)$ is a two-point distribution.

Current Work

- Complete proof for conjectures and solve the whole problem.
- Study shape properties of W in phase-type queues.

Current Work

- Complete proof for conjectures and solve the whole problem.
- Study shape properties of W in phase-type queues.

Roadmap

- Establish Extremal Multi-server Queueing Theory
  long-standing open problem for $GI/GI/K$ *for* $K \geq 2$.
- Extension to Queues under Service Operations
  scheduling, staffing, readmission.
- Extension to Open and Closed Queueing Networks
  dependence of arrival and departure process information.

Roadmap

- Establish Extremal Multi-server Queueing Theory
  long-standing open problem for $GI/GI/K$ *for* $K \geq 2$.
- Extension to Queues under Service Operations
  scheduling, staffing, readmission.
- Extension to Open and Closed Queueing Networks
  dependence of arrival and departure process information.

Roadmap

- Establish Extremal Multi-server Queueing Theory
  long-standing open problem for $GI/GI/K$ *for* $K \geq 2$.
- Extension to Queues under Service Operations
  scheduling, staffing, readmission.
- Extension to Open and Closed Queueing Networks
  dependence of arrival and departure process information.

References

Kingman (1962)
Inequalities for the queue $GI/G/1$
Biometrika.

Ward Whitt (1983)
Queueing Network Analyzer
Bell System Technical Journal.

Ward Whitt (1984)
On Approximation for Queues, I: Extremal Distributions
AT&T Bell Technical Journal.

Daley, Kreinin and Trengove (1992)
Inequalities concerning the waiting time in single-server queues: a survey
Queueing and Related Model, Clarendon Press.

Chen and Whitt (2018)
Extremal $GI/GI/1$ Queues Given First Two Moments
Under Second Review of Operations Research.

Chen and Whitt (2018)
Algorithms for the Upper Bound Mean Waiting Time in the $GI/GI/1$ Queue
Submitted into INFORMS Journal of Computing.