Extremal Cumulants in $GI/GI/1$ Queues Given Two Moments: Exploiting Tchebycheff Systems

Yan Chen
Industrial Engineering and Operations Research, Columbia University, yc3107@columbia.edu

Ward Whitt
Industrial Engineering and Operations Research, Columbia University, ww2040@columbia.edu

History: March 14, 2022

1. Introduction

This paper focuses on extremal distributions for the $GI/GI/1$ queue given a partial specification of the interarrival-time and service distributions. Extremal distributions attain tight upper and/or lower bounds for performance measures of interest given the partial model information. In particular, this paper supplements Chen and Whitt (2021a, 2022b), which investigate extremal distributions for the mean and higher moments of the waiting time in the $GI/GI/1$ queue given the first two moments of the interarrival time and service time; see Daley et al. (1992), Wolff and Wang (2003), Chen and Whitt (2020a) for additional background.

Here we apply the theory of Tchebycheff (T) systems from Karlin and Studden (1966) briefly reviewed in Chen and Whitt (2020b) to establish extremal distributions for the k^{th} cumulant for $k \geq 2$ of the steady-state waiting time in the $GI/GI/1$ queue given the first two moments of the interarrival time and service time. See (4) below for the formula. The second cumulant is of special interest because it is the variance.

For this purpose we apply Lemma 3 here, which coincides with Lemma 3 in Chen and Whitt (2021a) as corrected in Chen and Whitt (2022b) with revised proof here. Given those errors, the new results here provide the first valid application of Lemma 3. These new results here enable us to provide corrected proofs of Theorems 7 and 8 of Chen and Whitt (2021a). The continuity results for the mean steady-state waiting time in Theorem 7 in Chen and Whitt (2021a) depend on the bound in Theorem 8, which in turn depends on Theorem 1 in Chen and Whitt (2021a). Thus, these become unproved as well because Theorem 1 in Chen and Whitt (2021a) is now unproved. However, the new extremal results for the higher cumulants of the steady-state waiting time here provide corrected proofs of Theorems 7 and 8 in Chen and Whitt (2021a). These bounds for higher cumulants are interesting and important because they clearly demonstrate the value of Lemma 3 in Chen and Whitt (2021a) and highlight its limitation for treating the mean. In particular, the decreasing pdf condition in 3 (i) prevents positive results for the mean that we now obtain for the higher cumulants from Lemma 3 (ii) and (iii).

2. Background

2.1. The $GI/GI/1$ Queueing Model

The $GI/GI/1$ single-server queue has unlimited waiting space and the first-come first-served service discipline. There is a sequence of independent and identically distributed (i.i.d.) service times $\{V_n : n \geq 1\}$, each distributed as V with cumulative distribution function (cdf) G, which is independent of a sequence of i.i.d. interarrival times $\{U_n : n \geq 1\}$ each distributed as U with cdf F. With the understanding that the first customer (customer 1) arrives at time 0, V_n is the service time of customer n, while U_n is the interarrival time between customers n and $n+1$.

Let $\equiv$ denote equality by definition. Let U have mean $E[U] \equiv 1$ and squared coefficient of variation (scv, variance divided by the square of the mean) c_a^2; let a service time V have mean $E[V] \equiv \tau \equiv \rho$ and scv c_s^2, where $\rho < 1$, so that the model is stable.

Let W_n be the waiting time of customer n, i.e., the time from arrival until starting service, assuming that the system starts with an initial workload W_0 having cdf H_0 with a finite mean. The sequence $\{W_n : n \geq 0\}$ is well known to satisfy the Lindley recursion

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

where $x^+ \equiv \max\{x, 0\}$. Let H_n be the cdf of W_n, which is determined by (25). Let $W \equiv W_\infty$ (both used) be the steady-state waiting time, satisfying $W_n \Rightarrow W_\infty$ as $n \to \infty$, where $\Rightarrow$ denotes convergence in distribution; see §§X.1-X.2 of Asmussen (2003). The cdf H_∞ of $W \equiv W_\infty$ is the unique cdf satisfying the stochastic fixed point equation

$$W_\infty \stackrel{\mathrm{d}}{=} (W_\infty + V - U)^+, \tag{2}$$

where $\stackrel{\mathrm{d}}{=}$ denotes equality in distribution. If $P(W_0 = 0) = 1$, then $W_n \stackrel{\mathrm{d}}{=} \max\{S_k : 0 \le k \le n\}$ for $n \le \infty$, $S_0 \equiv 0$, $S_k \equiv X_1 + \cdots + X_k$ and $X_k \equiv V_k - U_k$, $k \ge 1$. Under the specified finite moment conditions, for $1 \le n \le \infty$, W_n is a proper random variable with finite mean, given by

$$E[W_n|W_0 = 0] = \sum_{k=1}^{n} \frac{E[S_k^+]}{k} < \infty, \quad 1 \le n < \infty, \quad \text{and} \quad E[W_\infty] = \sum_{k=1}^{\infty} \frac{E[S_k^+]}{k} < \infty. \tag{3}$$

Let $C_k(W)$ be the k^{th} cumulant of $W \equiv W_\infty$, which takes the form

$$C_k(W) = \sum_{n=1}^{\infty} n^{-1} E[(S_n^+)^k]; \tag{4}$$

see Section 2 of Janssen and van Leeuwaarden (2007), which draws on Spitzer (1956), Kingman (1962c,b). For example, the first cumulant is the mean, while the second cumulant is the variance. (We remark that stochastic comparison results for the higher cumulants were obtained in Bergmann et al. (1979).)

Our proof is done by considering finite sums obtained from truncating the cumulants of W and then taking a limit. We define the truncated k^{th} cumulants of W by

$$C_{k,m}(W) \equiv \sum_{n=1}^{m} n^{-1} E[(S_n^+)^k], \quad 1 \le m \le \infty. \tag{5}$$

For $k = 1$, $C_{k,m} = E[W_m]$, the mean of the transient waiting time, but this simple relation does not extend to higher cumulants, as can be seen for $k = 2$ from Theorem 3 in Section 6 of Kingman (1962c).

2.2. Sets of Probability Distributions with Specified Moments

Let $\mathcal{P}_n$ be the set of all probability measures on a subset of the positive real line $[0, \infty)$ with specified first n moments. The set $\mathcal{P}_n$ is a convex set, because the convex combination of two probability measures is just the mixture; i.e., for all p, $0 \le p \le 1$, $pP_1 + (1-p)P_2 \in \mathcal{P}_n$ if $P_1 \in \mathcal{P}_n$

and $P_2 \in \mathcal{P}_n$, because the n^{th} moment of the mixture is the mixture of the n^{th} moments, which is just the common value of the components. let $\mathcal{P}_{n,k}$ be the subset of probability measures in $\mathcal{P}_n$ that have support on at most k points.

Let $\mathcal{P}_2(m, c^2)$ be the subset of all cdf's in $\mathcal{P}_2$ with support in the interval $[0, \infty)$ having mean m and second moment $m^2(c^2+1)$. Let $\mathcal{P}_2(m, c^2, M)$ be subset of $\mathcal{P}_2(m, c^2)$ denoting all cdf's with support in the close interval $[0, M]$, where $1 + c^2 < M < \infty$ (The last property ensures that the set is non-empty.). Let subscripts a and s denote sets for the inter-arrival and service times, respectively. Therefore, $\mathcal{P}_{a,2}(1, c_a^2, M_a)$ is the set of all interarrival-time cdf's F with mean 1, scv c_a^2 and compact support within $[0, M_a]$, while $\mathcal{P}_{s,2}(\rho, c_s^2, M_s)$ is the set of all service-time cdf's G with mean ρ, scv c_s^2 and compact support within $[0, \rho M_s]$.

A special role is played by two-point distributions, which necessarily have finite support. Let $\mathcal{P}_{2,2}(m_1, c^2, M)$ be the set of all two-point distributions with mean m_1 and second moment $m_2 = m_1^2(c^2+1)$ with support in $[0, m_1 M]$. The set $\mathcal{P}_{2,2}(m_1, c^2, M)$ is a one-dimensional parametric family. Any element is determined by specifying one mass point. Let $F_b^{(2)}$ be the cdf that has probability mass $c^2/(c^2+(b-1)^2)$ on $m_1 b$, and mass $(b-1)^2/(c^2+(b-1)^2)$ on $m_1(1-c^2/(b-1))$ for $1+c^2 \le b \le M$. The cases $b = 1+c^2$ and $b = M$ constitute the two extremal distributions.

Since we are only interested in the extremal cdf's here, we will use different notation. We let $F_0 \equiv F_{1+c^2}^{(2)}$, because it is the unique element that has lower mass point 0 and we let $F_u \equiv F_M^{(2)}$, because it is the unique element that has upper mass point $m_1 M$. We use this definition for both the cdf's we consider: F of U and G of V, but recall that our parameter specification with $\mathbb{E}[U] = 1$ makes the support of F_u be $[0, M_a]$, while the support of G_u is $[0, \rho M_s]$. Therefore, with $M_a \ge 1 + c_a^2$ for F and $M_s \ge 1 + c_s^2$ for G, we have:

- $F_0 : c_a^2/(1+c_a^2)$ on 0 and $1/(1+c_a^2)$ on $1+c_a^2$;
- F_u: $(M_a-1)^2/(c_a^2+(M_a-1)^2)$ on $1-c_a^2/(M_a-1)$ and $c_a^2/(c_a^2+(M_a-1)^2)$ on M_a;
- $G_0 : c_s^2/(1+c_s^2)$ on 0 and $1/(1+c_s^2)$ on $\rho(1+c_s^2)$;
- G_u: $(M_s-1)^2/(c_s^2+(M_s-1)^2)$ on $\rho(1-c_s^2/(M_s-1))$ and $c_s^2/(c_s^2+(M_s-1)^2)$ on ρM_s.

3. Tight Bounds for the Cumulants

We consider service times V with cdf's G having mean ρ and support in $[0, \rho M_s]$ and interarrival times U with cdf F having mean 1 and support in $[0, M_a]$, but some of the results extend to unbounded support with qualifications; see Remark 1 below.

THEOREM 1. (*higher cumulants of the steady-state waiting time*) *Consider the $GI/GI/1$ model where $F \in \mathcal{P}_{a,2}(1, c_a^2, M_a)$ and $G \in \mathcal{P}_{s,2}(\rho, c_s^2, M_s)$ with $2 \le k < \infty$ and $1 \le m \le \infty$.*

(*a*) *Let the service-time cdf G be fixed. If $M_a \le \rho M_s$, then*

$$C_{k,m}(W(F_u, G)) \le C_{k,m}(W(F, G)) \le C_{k,m}(W(F_0, G)) \tag{6}$$

for all $F \in \mathcal{P}_{a,2}(1, c_a^2)$; these extrema are unique.

(*b*) *Let the interarrival-time cdf F be fixed. If $\rho M_s \le M_a$, then*

$$C_{k,m}(W(F, G_0)) \le C_{k,m}(W(F, G)) \le C_{k,m}(W(F, G_u)) \tag{7}$$

for all $G \in \mathcal{P}_{s,2}(\rho, c_s^2, M_s)$; these extrema are unique.

(*c*) *Suppose that neither F nor G is fixed. If $M_a = \rho M_s$, then*

$$C_{k,m}(W(F_u, G_0)) \le C_{k,m}(W(F, G)) \le C_{k,m}(W(F_0, G_u)) \tag{8}$$

for all $F \in \mathcal{P}_{a,2}(1, c_a^2, M_a)$ and $G \in \mathcal{P}_{s,2}(\rho, c_s^2, M_s)$; these extrema are unique.

REMARK 1. (extensions to unbounded support) Case (a) holds for F and G with unbounded support provided that we assume that $E[V^{k+1}] < \infty$, which holds for all k if the service time satisfies (9) of Chen and Whitt (2021a). For the condition $E[V^{k+1}] < \infty$, we can apply Theorem X.2.1 of Asmussen (2003) plus (25) in §6 below. The situation is different if we try to let $M_s \to \infty$ in cases (b) and (c) given only the first two moments of V, because the higher cumulants diverge to infinity as $M_s \to \infty$, as shown in Corollary 5.3 of Chen and Whitt (2020a). For (b), we can let $M_a = \infty$ if $M_s < \infty$. For (b), If $M_a = M_s = \infty$, then

$$C_{k,m}(W(F, G_0)) \le C_{k,m}(W(F, G)) \le \lim_{M_s \to \infty} C_{k,m}(W(F, G_u)) = \infty \tag{9}$$

for all $G \in \mathcal{P}_{s,2}(\rho, c_s^2, M_s)$. Finally, for case (c), the overall lower bound over F is then attained at D, the limit of F_u as $M_a \to \infty$, while the upper bound remains infinite.

4. Background on Tcheycheff Systems

In this section we briefly review the theory of Tchebycheff (T) systems. In particular, we give the brief account from §3 of Chen and Whitt (2021a) except we make the correction indicated in §3 of Chen and Whitt (2022b). An expanded account appears in §2 of Chen and Whitt (2020b), which draws on Karlin and Studden (1966). In particular, we apply Lemma 2.1 in §2.3 Chen and Whitt (2020b), which is a consequence of the tractable Wronskian condition for a T system.

Definition 1. (T System) Consider a set of $n+1$ continuous real-valued functions $\{u_i(t): 0 \le i \le n\}$ on the closed interval $[a,b]$. This set of functions constitutes a T system if the $(n+1)^{\text{st}}$-order determinant of the $(n+1)\times(n+1)$ matrix formed by $u_i(t_j)$, $0 \le i \le n$ and $0 \le j \le n$, is strictly positive for all $a \le t_0 < t_1 < \cdots < t_n \le b$.

Equivalently, except for an appropriate choice of sign, we could instead require that every nontrivial real linear combination $\sum_{i=0}^{n} a_i u_i(t)$ of the $n+1$ functions (called a u-polynomial; see §I.4 of Karlin and Studden (1966)) possesses at most n distinct zeros in $[a,b]$. (Nontrivial means that $\sum_{i=0}^{n} a_i^2 > 0$.)

We next state a consequence of Lemma 2.1 in §2.3 of Chen and Whitt (2020b). Let $\phi^{(n)}$ denote the n^{th} derivative of the function ϕ.

Lemma 1. (*from the* $(n+1)^{\text{st}}$ *derivative to a* T *system*) *Consider the real-valued functions* $u_i(t) \equiv t^i$, $0 \le i \le n$, *and* ϕ *on the interval* $[a,b]$ *for* $0 \le a < b < \infty$. *Suppose that* ϕ *has* $n+1$ *continuous derivatives. If* $\phi^{(n+1)}(t) > 0$ *for* $a \le t \le b$, *then* $\{u_0(t), u_1(t), \ldots, u_n(t), \phi(t)\}$ *is a* T *system of functions on* $[a,b]$. *If* $(-1)^{n+1}\phi^{(n+1)}(t) > 0$ *for* $a \le t \le b$, *then* $\{u_0(t), u_1(t), \ldots, u_n(t), -\phi(t)\}$ *is a* T *system of functions on* $[a,b]$.

As reviewed in §2 of Chen and Whitt (2020b), Lemma 1 applies to our setting when $n=2$. For Theorem 1 (a), we want the UB and LB of the integral

$$\int_0^{M_a} \phi(u)\, dF(u), \tag{10}$$

so that we will be applying Lemma 1 over the interval $[0, M_a]$. In part (a) of our queueing extremal problem we work with the integral form in (10) with integrand

$$\phi(u) \equiv \int_a^b h((y-u)^+)\, d\Gamma(y) = h(0)\Gamma(u) + \int_{u+}^b h(y-u)\, d\Gamma(y), \quad 0 \le u \le M_a, \tag{11}$$

where

$$-\infty \le a \le 0 < M_a \le b \le \infty, \tag{12}$$

Γ is a cdf of a real-valued random variable Y with a continuous positive density function over the interval $[a, b]$.

The following lemma combines Lemma 1 with the known extremal distributions in a T system, as given in Theorem 2.4 of Rolski (1976).

LEMMA 2. *If the condition of Lemma 1 is satisfied with $n = 2$ and $(-1)^3\phi^{(3)}(u) > 0$ for $0 \le u \le M_a$, then*

$$\sup\{\int_0^{M_a} \phi(u)\, dF(u) : F \in \mathcal{P}_{a,2}(1, c_a^2, M_a)\} = \int_0^{M_a} \phi(u)\, dF_0(u) \tag{13}$$

and

$$\inf\{\int_0^{M_a} \phi(u)\, dF(u) : F \in \mathcal{P}_{a,2}(1, c_a^2, M_a)\} = \int_0^{M_a} \phi(u)\, dF_u(u). \tag{14}$$

If the condition of Lemma 1 is satisfied with $n = 2$ and $(-1)^3\phi^{(3)}(u) < 0$ for $0 \le u \le M_a$, then then the roles of F_0 and F_u are switched in (13) and (14).

We now give sufficient conditions on h and the cdf Γ in (11) for the system $\{1, u, u^2, -\phi(u)\}$ to be a T system on $[0, M_a]$. For a real-valued function h of a real variable that has at least k continuous derivatives, let $h^{(k)}$ denote its k^{th} derivative; let $h^{(0)} \equiv h$. Let 1_A be the indicator function of the set A, which equals 1 on A and 0 on its complement. For part of this result, we will be assuming that the cdf Γ has a smooth pdf γ, but afterwards we can relax that assumption by using a limiting argument, as shown at the end of §5.

LEMMA 3. (*condition for the third derivative to be negative*) *Consider a nonnegative real-valued random variable Y with a finite mgf and the cdf Γ with support in $[a, b]$ or $[a, b)$ such that (12) holds. For ϕ in (11), in order to have*

$$(-1)^3\phi^{(3)}(u) > 0 \quad \textit{for} \quad 0 \le u \le M_a, \tag{15}$$

so that for $\{1, u, u^2, -\phi(u)\}$ *to be a* T *system on* $[0, M_a]$*, implying that* F_0 *attains the UB in* (13)*, while* F_u *attains the LB* (14)*, each of the following is a sufficient condition:*

(i) $h(x) \equiv x$ *and* Γ *has a positive pdf* γ *that is differentiable with* $\gamma^{(1)}(x) < 0$ *for* $a \leq x \leq M_a$*,*

(ii) $h(x) \equiv x^2$ *and* Γ *has a positive pdf* γ *for* $a \leq x \leq M_a$*,*

(iii) $h(x) \equiv h(x;p) \equiv x^p$ *for* $p \geq 3$*,*

(iv) $h(x) \equiv h(x;t) \equiv e^{tx} - tx - \frac{(tx)^2}{2} - \frac{(tx)^3}{6} = 1 + \sum_{k=4}^{\infty} \frac{(tx)^k}{k!}$ *for* $t > 0$*,*

(v) $h^{(k)}(x) > 0, \quad a < x \leq M_a, \quad 0 \leq k \leq 3$ *and* $h^{(k)}(a) = 0, \quad 1 \leq k \leq 2$.

For the function $h(x) \equiv x$ *in condition* (i)*, the condition on* γ *is necessary as well as sufficient, given that* γ *has a continuous positive derivative. In condition* (i)*, if instead* $\gamma^{(1)}(x) > 0$ *for* $0 \leq x \leq M_a$*, then the roles of* F_0 *and* F_u *are switched in* (13) *and* (14)*.*

Proof. First, observe that the finite mgf condition implies that all integrals are finite. In each case we can apply Lemmas 1 and 2 with (11) and (12). To do so, we apply the Leibniz rule for differentiation of an integral with (11). Using (12), we have

$$\begin{aligned} \phi(u) &= \int_a^b h((y-u)^+)\, d\Gamma(y) = \int_u^b h(y-u)\, d\Gamma(y) + h(0)\Gamma(u) \quad \text{and} \\ \phi^{(1)}(u) &= -\int_u^b h^{(1)}(y-u)\, d\Gamma(y) - h(0)\gamma(u) + h(0)\gamma(u) = -\int_u^b h^{(1)}(y-u)\, d\Gamma(y). \end{aligned} \tag{16}$$

For $h(x) \equiv x$ in condition (i), we have $h^{(1)}(x) = 1$ for all x, so that

$$\phi^{(1)}(u) = -\int_u^b h^{(1)}(y-u)\, d\Gamma(y) = -\int_u^b d\Gamma(y) = -(1 - \Gamma(u)), \tag{17}$$

so that, by the condition on Γ,

$$\phi^{(2)}(u) = \gamma(u) > 0 \quad \text{and} \quad \phi^{(3)}(u) = \gamma^{(1)}(u) < 0 \quad \text{for} \quad 0 \leq u \leq M_a. \tag{18}$$

From the form of $\phi^{(3)}(u)$ in (18), we see that the condition on γ is necessary as well as sufficient. We also see that the UB and LB are switched if instead $\gamma^{(1)}(u) > 0$.

Turning to $h(x) = x^2$ in condition (ii), we use $h^{(1)}(0) = 0$ and $h^{(2)}(x) = 2$ for all x with the second line of (16) to get

$$\phi^{(2)}(u) = \int_u^b h^{(2)}(y-u)\, d\Gamma(y) = 2\int_u^b d\Gamma(y) = 2(1 - \Gamma(u)) > 0, \tag{19}$$

so that $\phi^{(3)}(u) = -2\gamma(u) < 0$ for $0 \le u \le M_a$.

Conditions (iii) and (iv) are both special cases of condition (v), which implies that

$$\phi^{(3)}(u) = -\int_u^b h^{(3)}(y-u)\, d\Gamma(y) < 0. \quad \blacksquare \tag{20}$$

5. Proof of Theorem 1

We do the three cases (a), (b) and (c) in turn.

(a) We do the proof for case $k=2$; the argument is essentially the same for higher k. We do the proof for finite m and use a limiting argument to treat $m=\infty$. For $k=2$ and arbitrary finite m,

$$C_{2,m}(W) \equiv \sum_{n=1}^{m} n^{-1} E[(S_n^+)^2].$$

For the extremal problem, we need to solve the optimization

$$(OPT1) \quad \{\max \sum_{n=1}^{m} n^{-1} E[(S_n^+)^2]: \ F \in \mathcal{P}_{a,2}(M_a)\}.$$

Relax the above program by allowing different F for different $U_1, U_2, \ldots U_m$, i.e., consider the optimization

$$(OPT2) \quad \{\max \sum_{n=1}^{m} n^{-1} E[(S_n^+)^2], \ F_1, \ldots F_m \in \mathcal{P}_{a,2}(M_a)\}.$$

The optimal value for OPT2 is no smaller than OPT1 due to the relaxation. But if the optimal solution for OPT2 satisfies $F_1^* = F_2^* = \ldots = F_m^*$, then the solution will also be the optimal solution for OPT1. This idea has been applied in the Section 2.2 of van Eekelen et al. (2022).

Therefore, it suffices to consider the univariate case, which is a classical moment problem; see Smith (1995a). For the univariate case in OPT2 with $m=2$, we obtain

$$(OPT2, m=2) \ \{\max_{F_1}: \int_0^{M_a} \left(E[((V-u_1)^+)^2] + \frac{1}{2} E[((V_1+V_2-U_2-u_1)^+)^2] \right) dF_1(u_1)\} \tag{21}$$

such that $F_1 \in \mathcal{P}_{a,2}(M_a)$. We now apply Lemma 3 (ii) of Chen and Whitt (2021a) with $h(x) = x^2$, but with (25) of Chen and Whitt (2021a) replaced by (12) above. We initially assume that V has a positive continuous pdf over its support $[0, \rho M_s]$. Afterwards, we can get the desired result by taking a limit, because any cdf can be expressed as the limit (convergence in distribution) of cdf's with positive pdf's.

Since $E[U]=1$, the smallest closed interval containing the support of U is some $[c,d]$, where $0 \le c \le 1 \le d \le M_a$. Consider the two terms in the integrand of (21). In the first term, V has a positive pdf over $[0,\rho M_s]$. The conditions that $\rho M_s \ge M_a$ ensures that V has a positive pdf over $[0,M_a]$. Then consider the second term in (21). The random variable $V_1+V_2-U_2$ thus has a positive pdf over the interval $[-d, 2\rho M_s - c]$. It would suffice to have $2\rho M_s - 1 \ge M_a$ to guarantee that $V_1+V_2-U_2$ has a positive pdf over $[0,M_a]$, but that is implied by the support condition needed for the first term. The support requirement relaxes as k increases, so the given condition suffices. Overall, $V_1+V_2-U_2$ might not have a continuous pdf; e.g., if U were a two-point distribution. Nevertheless, the pdf is clearly continuous over $[0,M_a]$. Hence, we can apply Lemma 3 (ii) of Chen and Whitt (2021a) with $h(x)=x^2$ to each term of (21). For the first (second) term, the third derivative is $\phi^{(3)}(u) = -2\gamma_1(u) < 0$ ($\phi^{(3)}(u) = -\gamma_2(u) < 0$), $0 \le u \le M_a$, where γ_1 is the pdf of V and γ_2 is the pdf of $V_1+V_2-U_2$. We recursively apply the univariate result. Suppose we first apply this to u_1, the extremal distribution is $F^* = F_0$ according to Lemma 2 of Chen and Whitt (2021a), which is independent with $u_2, \dots u_l$. Hence, we can apply to the univariate case to u_2, etc. The proof is analogous for $k>2$.

For treating the more general case in which V has a cdf G without a density, we let $G \equiv \lim_{j\to\infty} G_j$, where G_j has a density satisfying the conditions above. Then the inequalities in (6) hold for each j. We then get convergence of the cumulants as $j \to \infty$ from convergence in distribution plus appropriate uniform integrability, as in §6 below. This step is elementary with bounded interarrival times and service times. Then convergence in distribution implies convergence of all moments and thus all cumulants.

Finally, for $m=\infty$, since $C_{k,m}(W(F,G)) \le C_{k,m}(W(F_0,G))$ for all $m \ge 1$ and $\sup_{m>0} C_{k,m}(W(F,G)) < \infty$ for all $F \in \mathcal{P}_{a,2}(M_a)$ and the $C_k(W(F,G)) = \lim_{m\to\infty} C_{k,m}(W(F,G)) < \infty$. Thus

$$C_k(W(F,G)) = \lim_{m\to\infty} C_{k,m}(W(F,G)) \le \lim_{m\to\infty} C_{k,m}(W(F_0,G)) = C_k(W(F_0,G)).$$

(b) As in Chen and Whitt (2021a), we reduce case (b) to case (a) by doing a time reversal. Specifically, here we work with $\tilde{V} \equiv \rho M_s - V$ and $\tilde{U} \equiv M_a - U$, assuming that the moment constraints are adjusted consistently. The analog of (21) for case (b) is

$$(OPT2b, m=2)\ \{\max_{\tilde{G}_1} : \int_0^{\rho M_s} \left(E[((\tilde{U} - \tilde{v}_1)^+)^2] + \frac{1}{2} E[((\tilde{U}_1 + \tilde{U}_2 - \tilde{V}_2 - \tilde{v}_1)^+)^2] \right) d\tilde{G}_1(\tilde{v}_1)\} \tag{22}$$

such that $\tilde{G}_1 \in \mathcal{P}_{s,2}(M_s)$, where $\tilde{G}$ is the cdf of $\tilde{V} \equiv \rho M_s - V$ with consistently adjusted moments. The remaining reasoning follows the proof of case (a) above.

We initially assume that U has a positive continuous pdf over $[0, M_a]$ and again treat the general case by a limiting argument. Then $\tilde{U}$ has a positive continuous pdf over $[0, M_a]$ too. We use the condition $\rho M_s \leq M_a$ to ensure that $\tilde{U}$ has a positive continuous pdf over $[0, \rho M_s]$. That covers the first term of the integrand in (22). The possible values of $\tilde{V} \equiv \rho M_s - V$ lie in $[0, \rho M_s]$, just like V. We see that $\tilde{U}_1 + \tilde{U}_2 - \tilde{V}_2$ has positive pdf over an interval $[c, d]$, where $c < 0 < d$, where $d \geq 2M_a - \rho \geq 2\rho M_s - \rho \geq \rho M_s$ Finally, the time reversal causes the bounds to switch; the upper bound in (b) involves G_u instead of F_0 in (a).

(c) Combine (a) and (b).∎

We now relax the pdf condition.

LEMMA 4. (*preservation of optimality*) *Suppose that* $\{Y_n : n \geq 1\}$ *is a sequence of real-valued random variables such that the conditions of Lemma 3 are satisfied for each* $n \geq 1$ *and* $Y_n \Rightarrow Y$ *as* $n \to \infty$. *If* F_0 (F_u) *yields the UB for* (13) *and* F_u (F_0) *yields the LB in* (14) *for all* $n \geq 1$, *then the same is true for the limit* Y.

Proof. We directly compare F_0 to any alternative cdf F for the UB. First, by the continuous mapping theorem, we obtain

$$\phi_n(u) \to \phi(u) \quad \text{as} \quad n \to \infty \tag{23}$$

for each u from (11). Then, by the dominated convergence theorem,

$$\int_0^{M_a} \phi(u)\, dF(u) = \lim_{n\to\infty} \int_0^{M_a} \phi_n(u)\, dF(u) \leq \lim_{n\to\infty} \int_0^{M_a} \phi_n(u)\, dF_0(u) = \int_0^{M_a} \phi(u)\, dF_0(u). \tag{24}$$

Hence. F_0 remains optimal for the limit. Essentially the same argument applies to the lower bound. ∎

6. Application to Bound Higher Moments

We now apply Theorem 1 above for the higher cumulants to obtain bounds for the higher moments. As shown in equation (5) of Smith (1995b), the k^{th} moment can be expressed as a function of the first $k-1$ moments and the first k cumulants via the recursive relation (with $E[W^0] \equiv 1$)

$$E[W^k(F,G)] = \sum_{i=0}^{k-1} \binom{k-1}{i} C_{k-i}(W(F,G))E[W^i(F,G)]. \tag{25}$$

We cannot apply (25) with Theorem 1 to immediately obtain tight bounds on all the moments, because we do not in general have a tight bound for the first cumulant, i.e., the mean. Of course, for any case in which Theorem 1 does hold for $k=1$, then (25) implies that the same extremal distribution works for all the moments as well. For example, case (a) holds for the mean when G is exponential by virtue of Eckberg (1977), Whitt (1984). However, we do know that there are significant differences for the mean. First, the lower bound (c) is known, and it does not involve the pair (F_u, G_0); see Section 2.4.1 of Chen and Whitt (2020a) and Section 5 of Chen and Whitt (2021b), Second, counterexamples have been established for the upper bound in cases (a) and (b); see Chen and Whitt (2021b).

Nevertheless, we can use (25) with Theorem 1 together with the classical Kingman (1962a) bound for the mean $E[W]$ to obtain upper bounds for the higher moments that serves to replace Theorem 8 used in the proof of Theorem 7 in Chen and Whitt (2021a). As indicated in Section 6 of Chen and Whitt (2021a), we are here using the T theory for probability distributions on the unbounded interval $[0,\infty)$ given the first three moments. Then the extremal cdf does not depend on M_s for M_s sufficiently large, as in case (a) given only two moments.

7. Concluding Remrks

7.1. On Extremal Distributions for the Mean

The proof of Theorem 1 does not extend directly to the mean. To apply Lemma 3 to the mean, we would need to have $V_1 + \cdots + V_n$ to have a decreasing pdf. Even if G is exponential, $V_1 + V_2$ fails to have a decreasing pdf. To establish Theorem 2 (a) of Chen and Whitt (2021a), it would appear that we should exploit the known representation of the steady-state waiting-time and sojourn-time cdf's. By Theorem 6 of Chen and Whitt (2021a), the steady-state sojourn-time cdf is completely monotone if the service-time cdf is completely monotone. Perhaps there is a way to exploit §II.5.9 or §II.5.10 of Cohen (1982).

7.2. Remaining Open Problems

We conclude by stating two important conjectures.

CONJECTURE 1. (*the tight upper bound for* $1 \le n \le \infty$)

(*a*) *Given any parameter vector* $(1, c_a^2, \rho, c_s^2)$ *and a bounded interval* $[0, \rho M_s]$ *for the service-time cdf* G, *where* $M_s \ge c_s^2 + 1$, *the pair* (F_0, G_u) *attains the tight upper bound of the steady-state mean* $E[W]$, *i.e.,*

$$E[W(F,G)] \le E[W(F_0, G_u)] \quad \textit{for all} \quad F \in \mathcal{P}_{a,2} \quad \textit{and} \quad G \in \mathcal{P}_{s,2}(M_s),$$

while a pair $(F_0, G_{u,n})$ *attains the tight upper bound of the transient mean* $E[W_n]$, *i.e.,*

$$E[W_n(F,G)] \le E[W_n(F_0, G_{u,n}))] \quad \textit{for all} \quad F \in \mathcal{P}_{a,2} \quad \textit{and} \quad G \in \mathcal{P}_{s,2}(M_s),$$

where $G_{u,n}$ *is a two-point distribution with* $G_{u,n} \Rightarrow G_u$ *as* $n \to \infty$.

(*b*) *When both* F *and* G *have unbounded support* $[0, \infty)$, *the tight upper bound of* $E[W(F,G)]$ *is obtained asymptotically in the limit as* $M_s \to \infty$ *in part (a), i.e.,*

$$E[W(F,G)] \le \lim_{M_s \to \infty} E[W(F_0, G_u)] \equiv E[W(F_0, G_{u^*})] \quad \textit{for all} \quad F \in \mathcal{P}_{a,2} \quad \textit{and} \quad G \in \mathcal{P}_{s,2}.$$

CONJECTURE 2. (*three-point extremal distributions*) *All the tight upper bounds and the corresponding tight lower bounds are attained by three-point distributions (allowing for limits as* $M_s \to \infty$*).*

We have studied $E[W(F_0, G_{u^*})]$ in Chen and Whitt (2020a). We have established and applied numerical algorithms to gain insight in Chen and Whitt (2022a, 2021b). As shown in Chen and Whitt (2020a), the gap between the upper bounds and the tight lower bound is quite wide, indicating that extra information should be used in order to get accurate approximations. A heuristic approach to refined bounds for the mean based on tight bounds for the asymptotic decay rate of the steady-state waiting time W is described in Chen and Whitt (2022c).

References

Asmussen, S. 2003. *Applied Probability and Queues*. 2nd ed. Springer, New York.

Bergmann, R., D. J. Daley, T. Rolski, D. Stoyan. 1979. Bounds for cumulants of waiting times in $GI/GI/1$ queues. *Mathematische Operationsforschung und Statistik. Series Optimization* **10**(2) 257–263.

Chen, Y., W. Whitt. 2020a. Algorithms for the upper bound mean waiting time in the $GI/GI/1$ queue. *Queueing Systems* **94** 327–356.

Chen, Y., W. Whitt. 2020b. Extremal models for the $GI/GI/K$ waiting-time tail-probability decay rate. *Operations Research Letters* **48** 770–776.

Chen, Y., W. Whitt. 2021a. Extremal $GI/GI/1$ queues given two moments: Exploiting Tchebycheff systems. *Queueing Systems* **97** 101–124.

Chen, Y., W. Whitt. 2021b. Extremal $GI/GI/1$ queues given two moments: Three-point and two-point distributions. Columbia University, http://www.columbia.edu/∼ww2040/allpapers.html.

Chen, Y., W. Whitt. 2022a. Applying optimization to study extremal $GI/GI/1$ transient mean waiting times. *Queueing Systems* **100** 1–20.

Chen, Y., W. Whitt. 2022b. Errata to 'extremal $GI/GI/1$ queues given two moments: Exploiting Tchebycheff systems'. Working paper, Columbia University.

Chen, Y., W. Whitt. 2022c. Set-valued performance approximations for the $GI/GI/K$ queue given partial information. *Probability in the Engineering and Informational Sciences* **x** 1–23.

Cohen, J. W. 1982. *The Single Server Queue*. 2nd ed. North-Holland, Amsterdam.

Daley, D. J., A. Ya. Kreinin, C.D. Trengove. 1992. Inequalities concerning the waiting-time in single-server queues: a survey. U. N. Bhat, I. V. Basawa, eds., *Queueing and Related Models*. Clarendon Press, 177–223.

Eckberg, A. E. 1977. Sharp bounds on Laplace-Stieltjes transforms, with applications to various queueing problems. *Mathematics of Operations Research* **2**(2) 135–142.

Janssen, A. J. E. M., J. S. H. van Leeuwaarden. 2007. Cumulants of the maximum of the Gaussian random walk. *Stoch. Proc. Appl.* **117** 1928–1959.

Karlin, S., W. J. Studden. 1966. *Tchebycheff Systems; With Applications in Analysis and Statistics*, vol. 137. Wiley, New York.

Kingman, J. F. C. 1962a. Inequalities for the queue $GI/G/1$. *Biometrika* **49**(3/4) 315–324.

Kingman, J. F. C. 1962b. Spitzer's identity and its use in probability theory. *J. London Math. Soc.* **s1-37**(1) 309–316.

Kingman, J. F. C. 1962c. The use of Spitzer's identity in the investigation of the busy period and other quantities in the queue $GI/G/1$. *J. Australian Math. Soc.* **2**(3) 345–356.

Rolski, T. 1976. Order relations in the set of probability distribution functions and their applications in queueing theory. *Dissertationes Mathematicae, Polish Academy of Sciences* **132**(1) 3–47.

Smith, J. 1995a. Generalized Chebychev inequalities: Theory and application in decision analysis. *Operations Research* **43** 807–825.

Smith, P. J. 1995b. A recursive formuation of the old problem of obtaining moments from cumulants and vice versa. *The American Staistican* **49**(2) 217–218.

Spitzer, F. L. 1956. A combinatorial lemma and its application to probability theory. *Trans. Amer. Math. Soc.* **82** 323–339.

van Eekelen, W., D. den Hartog, J. S. H. van Leeuwaarden. 2022. MAD dispersion measure makes extremal queue analysis simple. *INFORMS J. Computing* **34** 1–12.

Whitt, W. 1984. On approximations for queues, I: Extremal distributions. *AT&T Bell Laboratories Technical Journal* **63**(1) 115–137.

Wolff, R. W., C. Wang. 2003. Idle period approximations and bounds for the $GI/G/1$ queue. *Advances in Applied Probability* **35**(3) 773–792.