MAD dispersion measure makes extremal queue analysis simple

Dick den Hertog
Department of Econometrics and Operations Research, Tilburg University, d.denhertog@tilburguniversity.edu

Johan S.H. van Leeuwaarden
Department of Econometrics and Operations Research, Tilburg University, and Department of Mathematics and Computer Science, Eindhoven University of Technology, j.s.h.vanleeuwaarden@tilburguniversity.edu

A notorious problem in queueing theory is to identify the worst possible performance of the GI/G/1 queue under mean-dispersion constraints for the interarrival and service time distributions. We address this extremal queue problem by measuring dispersion in terms of Mean Absolute Deviation (MAD) instead of variance, making available recently developed techniques from Distributionally Robust Optimization (DRO). Combined with classical random walk theory, we obtain the extremal interarrival time and service time distributions, and hence the best possible upper bounds, on all moments of the waiting time. We also employ DRO techniques to obtain lower bounds and to solve queueing related optimization problems. We leverage the novel DRO-MAD perspective to sketch several extensions and describe now-opened research directions in queueing theory, scheduling and inventory theory.

Key words: extremal queue problem, GI/G/1 queue, random walk theory, tight bounds, distributionally robust optimization

History: This paper was first submitted on July 30, 2019

1. Introduction

In practice the analysis of queueing systems is hindered by two factors. Firstly, the arrival and service probability distributions are not known exactly. Secondly, even if we do know these distributions, the analysis is often computationally intractable. In this paper we advocate to use a simple Distributionally Robust Optimization (DRO) approach to overcome these two hurdles. Essential in our approach is the use of Mean Absolute Deviation (MAD) instead of the variance as measure of dispersion. This enables us not only to derive simple expressions for both the worst-case and best-case distributions, but also to efficiently solve multi-stage queueing related optimization problems.

Queueing theory exists more than a century with throughout a central role for the GI/G/1 queue with i.i.d. interarrival times $\{U_n\}$ distributed as $U$ and i.i.d. service times $\{V_n\}$ distributed as $V$. The waiting times in the GI/G/1 queue can be expressed as the maxima of a random walk with step size $X = V - U$, the subject of an enormous literature: [Chung (2001), Feller (1971)]. For all moments of the maxima (i.e., waiting time), general expressions are available that involve
convolutions of the distribution of $X$. To use these general expressions, one thus needs to specify the precise distribution of $X$, and in the case of the GI/G/1 queue the distributions of both $U$ and $V$.

Special cases of the GI/G/1 queue can be studied with dedicated techniques for Markov chains. For instance, the M/G/1 queue with Poisson arrivals and the GI/M/1 queue with exponential services have explicit solutions that are more insightful than the general random walk results; Asmussen (2003), Cohen (1982). Another large, somewhat opposite branch of queueing theory concerns finding approximations and bounds. For the steady-state waiting time $W$ in the GI/G/1 queue, arguably the most famous upper bound for $\mathbb{E}[W]$ was obtained by Kingman (1962) in terms of the first two moments of both $U$ and $V$. While Kingman’s bound is sharp in situations of heavy traffic, when $\mathbb{E}[U]/\mathbb{E}[V]$ approaches 1, it leaves room for improvement for all other values of $\mathbb{E}[U]/\mathbb{E}[V]$.

In search for that sharpest possible (tight) upper bound under the first two moments constraints, foundational work was done by Rolski (1972), Eckberg Jr (1977), Whitt (1984) in the context of the GI/M/1 queue. Whitt (1984) considered the GI/M/1 queue with given mean and variance of $U$, and showed that $\mathbb{E}[W]$ is maximized when the interarrival follow a specific two-point distribution. Similar findings were made for other special cases of $U$ and $V$. It also led to the conjecture that the overall worst case behavior (in terms of $\mathbb{E}[W]$) would be caused by two-point distributions, for both $U$ and $V$. That conjecture was proved invalid by counterexamples in Whitt (1984) when fixing either $U$ or $V$, but the conjecture remained standing for the case when both $U$ and $V$ are unspecified, except for their first two moments. After that it remained silent for a while, until Chen and Whitt (2019) showed recently, for distributions with finite support, that the extremal distributions of $U$ and $V$ both have supports on at most three points. While existence is thus proved, the exact form of the extremal three-or-fewer-points distributions can only be determined numerically, as the solution of a hard non-convex nonlinear optimization problem. Extensive numerical experiments led Chen and Whitt to conjecture that the worst case is formed by two-point distributions for both $U$ and $V$, in line with the conjecture postulated several decades ago. Finding the extremal queue for given mean-variance information is therefore one of the longest standing problems in the field. That problem remains open, also after publication of the present paper.

We do consider the same problem of finding the sharpest possible bounds on GI/G/1 queue performance, but take a radical turn by quantifying dispersion in terms of mean absolute deviation (MAD) instead of variance. That may appear a bold decision, because MAD is hardly used in queueing theory, or random walk theory for that matter. We can only speculate about the historical reasons for variance preference, but the random walk and GI/G/1 queue are intrinsically linked with i.i.d. sums of random variables, and variance then enters naturally (e.g., variance of the
sum, central limit theorem). The variance and MAD, however, are equally adequate descriptors of
dispersion, and are both easily calibrated on data using basic statistical estimators.

The MAD perspective offered in this paper departs from the variance-based formulations of
the past (see [Rolski (1972), Eckberg Jr (1977), Whitt (1984), Chen and Whitt (2019) and the
references therein]), and brings to bear the rich theory of robust optimization, in particular the
rapidly expanding theory of distributionally robust optimization (DRO). The exact expressions for
the random walk maxima form a crucial ingredient for our proof methodology. These expressions
are convex functions of the driving random variables, a prerequisite for the mean-MAD approach.
Indeed, recent advances in DRO, see [Postek et al. (2018)], show that knowledge on the support,
mean and MAD can lead to closed-form expressions for stochastic quantities such as the minimum
and maximum expectation of a convex function. As will become apparent, the effectiveness and
mathematical elegance that comes naturally with the mean-MAD DRO perspective is startling, not
only for solving the extremal GI/G/1 queue problem, but for many related or extended problems
in service operations management.

The contributions of this paper can be summarized as follows:

1. We suggest to use MAD instead of variance, and obtain by concise mathematical proof the
   worst-case three-point distribution for a rich class of extremal problems. This proof for MAD
gives insight into why the traditional moment constraints, although a popular choice, may not
necessarily yield tractable counterparts.

2. We leverage this result to obtain tight upper and lower bounds for performance measures,
   including transient and steady-state queue length moments. Under mean-MAD constraints,
   these bounds are the sharpest possible (and thus cannot be improved).

3. We demonstrate how the MAD approach can effectively solve a large class of queueing opti-
nization problems, with a wide range of possible applications. This taps into a new quantitative
method for queueing and related stochastic systems. The generic approach described in this
paper is a novel computationally tractable way to analyze and optimize such systems.

The remainder of the paper is organized as follows. Section 2 presents the MAD perspective.
Section 3 discusses methods to obtain upper and lower bounds for both best and worst-case per-
formance. Section 4 presents a full solution of the extremal queue problem with mean-MAD con-
straints, and draws a comparison with the traditional mean-variance setting. Section 5 introduces
DRO techniques that can be applied for optimization of queueing systems. We conclude in Section
6 with an outlook of the many possibilities for follow-up research.

Notation. Boldfaced characters represent vectors, and $x_i$ denotes the $i$-th element of vector $x$.
For a random variable $X$, we use $X \sim P \in \mathcal{P}$ to say that $X$ is a random variable with probability
distribution $P$ from the set of probability distributions $\mathcal{P}$. We denote $E_P[\cdot]$ as the expectation over
the probability distribution \( P \). When we consider \( E_P[f(X)] \) with \( X = (X_1, \ldots, X_n) \), it is tacitly assumed that \( f(\cdot) \) is a measurable function from \( \mathbb{R}^n \) to \( \mathbb{R} \), and such that \( E_P[f(X)] \) exists.

2. Extremal random walk

Consider the partial sums \( S_n := X_1 + \cdots + X_n \) (\( S_0 := 0 \)) of i.i.d. random variables \( X_1, X_2, \ldots \) distributed as \( X \). The random walk \( (S_n, n \geq 0) \) arises in many application domains, including queueing theory, inventory management and risk theory. If \( (S_n, n \geq 0) \) indeed models congestion, shortfall or capital position, large values of \( S_n \) are of particular interest, and it is natural to consider the maxima sequence \( M_n := \max \{S_0, S_1, \ldots, S_n\} \). The random walk and its maxima can be studied with mathematical techniques for sums or random variables, covered in many standard texts on probability theory, e.g., Asmussen (2003), Chung (2001), Cohen (1982), Feller (1971). For the distribution and moments of \( M_n \) there exist general formulas in terms of finitely many convolutions. However, applying these exact formula requires full specification of the distribution of \( X \). This paper searches for the sharpest possible bounds on \( E[M_n] \) and related quantities, when only information is available on the mean and dispersion of \( X \). We now present such bounds when the partial information consists of the mean, range and MAD of \( X \).

2.1. Extremal distribution

Notice that \( M_n \) can be expressed as \( h_n(X_1, \ldots, X_n) \), with

\[
h_n(x_1, \ldots, x_n) = \max\{0, x_1, \ldots, x_1 + \cdots + x_n\},
\]

and the expected maximum can be expressed as \( E[M_n] = E[h_n(X)] \) with \( X = (X_1, \ldots, X_n) \). For now assume that \( X_1, \ldots, X_n \) are independent, but that each \( X_i \) can have a different distribution. Assuming we only have partial information consisting of means and dispersion measures of the random variables \( X_1, \ldots, X_n \), the first question we ask and answer in this paper is: What extremal distributions of \( X_i \) result in the worst-case expected maxima? Extremal distributions have been studied in many contexts, and in the literature variance is predominantly used as the dispersion measure. Here we shall use the MAD. To describe all considered distributions we define an ambiguity set that consists of all distributions of componentwise independent \( X \) with known supports, means, and MADs. The partial information for \( (X_1, \ldots, X_n) \) consists of (i) \( X_i \) has support \( \text{supp}(X_i) = [a_i, b_i] \) with \(-\infty < a_i \leq b_i < \infty, i = 1, \ldots, n\), (ii) \( E_P(X_i) = \mu_i \) and (iii) \( E_P|X_i - \mu_i| = d_i \). This defines the \((\mu, d)\) ambiguity set, consisting of the distributions with known (i), (ii), and (iii) for each \( X_i \):

\[
P_{(\mu, d)} = \{P : \text{supp}(X_i) \subseteq [a_i, b_i], E_P(X_i) = \mu_i, E_P|X_i - \mu_i| = d_i, \forall i, X_i \perp \perp X_j, \forall i \neq j\},
\]

where \( X_i \perp \perp X_j \) denotes the stochastic independence of \( X_i \) and \( X_j \). In what follows, \( X \) is a vector of random variables whose distribution \( P \) belongs to the set \( P_{(\mu, d)} \).
As the title says, with MAD as dispersion measure, the extremal problem becomes simple. Observe that the function \( h_n \) is convex in the vector \((x_1, \ldots, x_n)\). We can thus apply the general upper bound in Ben-Tal and Hochman (1972) on the expectation of a convex function of independent random variables with mean-MAD ambiguity, which gives the following result:

**Theorem 1.** The extremal distribution that solves

\[
\max_{P \in \mathcal{P}(\mu, d)} \mathbb{E}_P[h_n(X)]
\]

consists for each \( X_i \) of a three-point distribution with values \( \tau_1^{(i)} = a_i, \tau_2^{(i)} = \mu, \tau_3^{(i)} = b_i \) and probabilities

\[
p_1^{(i)} = \frac{d_i}{2(\mu - a_i)}, \quad p_2^{(i)} = 1 - \frac{d_i}{2(\mu - a_i)} - \frac{d_i}{2(b_i - \mu_i)}, \quad p_3^{(i)} = \frac{d_i}{2(b_i - \mu_i)}.
\]

The proof of this theorem thus directly follows from Ben-Tal and Hochman (1972). In the next section we give another proof of Theorem 1 under more general conditions, that also gives insight into why using as dispersion measure MAD instead of variance makes the analysis so simple.

### 2.2. Why MAD simplifies analysis

Let us first show why the worst-case distribution a three-point distribution. We consider some univariate measurable function \( f(x) \) (with the univariate function \( h_1(x_1) \) as an example) that has finite values on \([a, b]\), the support of the distribution \( p(x) \). Under mean-MAD ambiguity of one random variable \( X \) we thus need to solve

\[
\max_{p(x) \in \mathcal{P}(\mu, d)} \int_x f(x)p(x)dx
\]

s.t.

\[
\int_x |x - \mu|p(x)dx = d, \int_x xp(x)dx = \mu, \int_x p(x)dx = 1, \quad p(x) \geq 0,
\]

a semi-infinite linear optimization problem in standard form with three equality constraints. From basic LP theory we then know that if \((5)\) has a feasible solution, then there exists an optimal solution for \((5)\) for which \(p(x) > 0\) in at most three points. Observe that the same proof argument works when \( \sigma^2 \) is given instead of \( d \), i.e., when \(|x - \mu|\) in \((5)\) is replaced by \((x - \mu)^2\). Hence, irrespective of whether MAD or variance is used as dispersion measure, for determining the tight upper bound of \( f(x) \), it suffices to consider distributions with support on at most three points.

How can we then find this worst-case distribution? Consider the dual of \((5)\),

\[
\min_{\lambda_1, \lambda_2, \lambda_3} \lambda_1 d + \lambda_2 \mu + \lambda_3
\]

s.t.

\[
f(x) - \lambda_1|x - \mu| - \lambda_2 x - \lambda_3 \leq 0, \quad \forall x \in [a, b].
\]

Define \( F(x) = \lambda_1|x - \mu| + \lambda_2 x + \lambda_3 \). Then the inequality in \((6)\) can be written as \( f(x) \leq F(x), \forall x \), i.e. \( F(x) \) majorizes \( f(x) \). Note that \( F(x) \) has a ‘kink’ at \( x = \mu \). Since the dual problem \((6)\) has three
variables, the tightest majorant $F(x)$ touches $f(x)$ at three points: $x = a$, $\mu$ and $b$, as illustrated in Figure 1. The optimal probabilities of (5) can now easily be obtained by solving the linear system resulting from the equations of (5). This is a linear system of three unknown probabilities and three equations, with solution stated in Theorem 1. A crucial observation is that this result holds independent of $f(x)$! This is because $F(x)$ is piecewise linear and convex, and hence each convex function $f(x)$ for which $f(x) = F(x)$ at the end points $a, b$ and at the kink point $x = \mu$, is majorized by $F(x)$.

Why is $d$ then computationally easier than $\sigma^2$? It can easily be verified that when $\sigma^2$ is used as dispersion measure, the end points and kink point do not necessarily form the extremal distribution. That is, upon replacing $|x - \mu|$ with $(x - \mu)^2$, the tightest majorant $F(x)$ does not necessarily touch $f(x)$ in $a$, $b$ and $\mu$ anymore. Hence, if the variance is used as dispersion measure, then the worst-case distribution depends on the function $f(x)$. This has extremely important consequences for the multivariate case, i.e., when we consider $h_n(x_1, \ldots, x_n)$.

If the MAD is used as dispersion measure, we can recursively apply the above result for univariate functions. Suppose we first apply this result to $x_1$, then the worst-case distribution is as in Theorem 1 independent of the values for $x_2, \ldots, x_n$. Moreover, the worst-case expectation becomes a convex function in $x_2, \ldots, x_n$, since the worst-case probabilities for $x_1$ are nonnegative. Hence, we can apply the result above for the univariate case to $x_2$, etc.

However, in case the variance is used as dispersion measure, the worst-case distribution depends on the values of $x_2, \ldots, x_n$. Calculating the worst-case distribution as a function of $x_2, \ldots, x_n$ seems to be impossible. Moreover, even if we would be able to derive such a worst-case distribution, substituting this distribution in the worst-case expectation would result in an extremely difficult function in $x_2, \ldots, x_n$ that most probably is not convex, and hence applying the univariate result to $x_2$ is not possible anymore.
3. Sharpest possible bounds

A direct consequence of Theorem 1 is that the worst-case expectation of $h_n(X)$ is obtained by enumerating over all $3^n$ permutations of outcomes $a_i, \mu_i, b_i$ of components $X_i$.

**Corollary 1.**

$$\max_{P \in \mathcal{P}(\mu, d)} \mathbb{E}_P[h_n(X)] = \sum_{\alpha \in \{1, 2, 3\}^n} h_n(\tau_{\alpha_1}^{(1)}, \ldots, \tau_{\alpha_n}^{(n)}) \prod_{i=1}^{n} P_{\alpha_i}^{(i)}. \quad (7)$$

Thus, under the partial information contained in $\mathcal{P}(\mu, d)$, (7) is an upper bound on $\mathbb{E}[M_n]$ that cannot be improved. We next specialize to the random walk setting with $X_1, X_2, \ldots$ independent and distributed as $X$, and obtain representations for the tight upper bound that are computationally less cumbersome than (7), and extend to all moments of the all-time maximum (when $n \to \infty$).

3.1. Random walk upper bounds

We recall that Spitzer (1956) used combinatorial arguments to establish for $\mathbb{E}[M_n]$ the alternative expression (which strictly requires i.i.d. increments)

$$\mathbb{E}[M_n] = \sum_{k=1}^{n} \frac{1}{k} \mathbb{E}[S_k^{+}], \quad (8)$$

with $x^{+} = \max\{0, x\}$. This can be written as $\mathbb{E}[M_n] = \mathbb{E}[f_n(X)]$ with

$$f_n(x_1, \ldots, x_n) = \sum_{k=1}^{n} \frac{1}{k} \max\{0, x_1 + \ldots + x_k\}. \quad (9)$$

A first usage of Spitzer’s formula (8) is a considerable improvement, in terms of computational complexity, of the tight bound for $\mathbb{E}[M_n]$ in (7). To state the result and for later reference, let $\Omega(\mu, d, a, b)$ denote a three-point distribution on the values $\{a, \mu, b\}$ with probabilities

$$p_1 = \frac{d}{2(\mu - a)}, \quad p_2 = 1 - \frac{d}{2(\mu - a)} - \frac{d}{2(b - \mu)}, \quad p_3 = \frac{d}{2(b - \mu)} \quad (10)$$

Let $X_{(3)}$ denote the random variable with the extremal three-point distribution, identified in Theorem 1 for the special case when $X_1, X_2, \ldots$ are i.i.d., hence $X_{(3)} \sim \Omega(\mu, d, a, b)$.

**Corollary 2.**

$$\max_{P \in \mathcal{P}(\mu, d)} \mathbb{E}_P[f_n(X)] = \sum_{k=1}^{n} \frac{1}{k} \sum_{\sum_{i=1}^{k} a_i + k_2 \mu + k_3 b} \frac{k!}{k_1! k_2! k_3!} p_1^{k_1} p_2^{k_2} p_3^{k_3}. \quad (11)$$

Note that for each fixed $k$, (11) contains a multinomial distribution with support set $\{(k_1, k_2, k_3) \in \mathbb{N}^3 : k_1 + k_2 + k_3 = k\}$ with cardinality $\binom{k+2}{2}$. This implies that the sum over $k$ in (11) is over roughly $n^3$ terms, which is way better than the $3^n$ terms in (7).
For $\mathbb{E}[X] < 0$ the all-time maximum $M := \lim_{n \to \infty} M_n$ is a proper random variable ($M_n$ converges in distribution to $M$, which will be finite with probability one if $\mathbb{E}[X] < 0$). Let $c_m(M)$ denote the $m$-th cumulant of $M$. Recall that $c_1(M)$ is the mean, $c_2(M)$ is the variance, and $c_3(M)$ is the central moment $\mathbb{E}[(M - \mathbb{E}[M])^3]$. From general random walk theory we know that (see e.g., Abate et al. (1993))

$$c_m(M) = \sum_{k=1}^{\infty} \frac{1}{k} \mathbb{E}[(S_k^+)^m]. \quad (12)$$

We can now prove results similar as for $\mathbb{E}[M_n]$, regarding the extremal distribution and tight upper bound.

**Theorem 2.** Consider the random walk with generic step size $X$ contained in the ambiguity set $\mathcal{P}_{(\mu,d)}$. The tight upper bounds for all cumulants $c_m(M)$ of the all-time maximum $M$ are the cumulants of the random walk with extremal step size $X^{(3)}$.

**Proof.** Consider the function

$$f_n^m(x_1, \ldots, x_n) = \sum_{k=1}^{n} \frac{1}{k} (\max\{0, x_1 + \ldots + x_k\})^m, \quad (13)$$

which is convex in the vector $(x_1, \ldots, x_n)$. Hence, for i.i.d. increments with generic $X$,

$$\max_{P \in \mathcal{P}_{(\mu,d)}} \mathbb{E}_P[f_n^m(X)] \quad (14)$$

is solved by the extremal random variable $X^{(3)}$. This gives the bound, with $X_1^*, X_2^*, \ldots$ i.i.d. as $X^{(3)}$,

$$a_n := \sum_{k=1}^{n} \frac{1}{k} \mathbb{E}[(S_k^+)^m] \leq \mathbb{E}f_n^m(X_1^*, \ldots, X_n^*) =: b_n. \quad (15)$$

The result follows by observing that the sequences $\{a_n\}$ and $\{b_n\}$ are both monotone, and converging to well-defined limits. □

We conclude that the extremal three-point distribution for $\mathbb{E}[M_n]$ in Theorem 1 is also the extremal distribution for all cumulants of $M$. When calculating the associate tight upper bounds for $c_m(M)$, we are confronted with an infinite summation of increasingly complex summands. Here, another line of classical random walk theory can help, which transforms such infinite sums into complex contour integrals. Let $\phi_{X^{(3)}}(s) := \mathbb{E}[e^{sX^{(3)}}] = p_1e^{sa} + p_2e^{sb} + p_3e^{sa}$.

**Corollary 3.** The tight upper bounds on $c_m(M)$ identified in Theorem 2 are given by

$$\frac{(-1)^m}{2\pi i} \int_{\mathcal{C}} \frac{\log(1 - \phi_{X^{(3)}}(-u))}{u^{m+1}} du, \quad m = 1, 2, \ldots, \quad (16)$$

where $\mathcal{C}$ is a contour to the left of, and parallel to, the imaginary axis, and to the right of any singularities of $\log(1 - \phi_{X^{(3)}}(-u))$ in the left half plane.

Observe that (16) bypasses the cumbersome calculations with convolutions. In Section 2 we present more details and confirm that this is a numerically efficient way of calculating the tights bounds.
3.2. Random walk lower bounds

The tight upper bounds correspond to worst-case scenarios. We next show how the same MAD approach can identify best-case scenarios and hence tight lower bounds. For each $X_i$, define a second ambiguity set, which is a subset of $\mathcal{P}(\mu,d)$:

$$\mathcal{P}_{(\mu,d,\beta)} = \left\{ \mathbb{P} \mid \mathbb{P} \in \mathcal{P}(\mu,d), \mathbb{P}(X_i \geq \mu_i) = \beta_i, \forall i \right\}.$$  \hfill (17)

**Theorem 3.**

$$\min_{\mathbb{P} \in \mathcal{P}_{(\mu,d,\beta)}} \mathbb{E}_{\mathbb{P}}[h_n(X)] = \sum_{\alpha \in \{1,2\}^n} h_n(v_{\alpha_1}^{(1)}, \ldots, v_{\alpha_n}^{(n)}) \prod_{i=1}^{n} q_i^{(i)},$$

where

$$q_1^{(i)} = \beta_i, \quad q_2^{(i)} = 1 - \beta_i, \quad v_1^{(i)} = \mu_i + d_i/2\beta_i, \quad v_2^{(i)} = \mu_i - d_i/2(1 - \beta_i).$$ \hfill (19)

**Proof.** The proof of Theorem 3 follows from the general lower bound in Ben-Tal and Hochman (1972) on the expectation of a convex function of independent random variables with ambiguity. \hfill □

Again specialize to the i.i.d. setting, and denote by $Y$ the random variable with two-point distribution on values $v_1 = \mu + \frac{d}{2\eta}$, $v_2 = \mu - \frac{d}{2(1 - \eta)}$, with probabilities $\beta$ and $1 - \beta$, respectively. Using similar reasonings as for the upper bound, we obtain for the tight lower bound on $\mathbb{E}[M_n]$ an expression that sums over $O(n^2)$ terms:

$$\sum_{k=1}^{n} \frac{1}{k} \sum_{k_1+k_2=k} \frac{k!}{k_1!k_2!} \beta^{k_1}(1 - \beta)^{k_2} \max\{0, k_1v_1 + k_2v_2\}.$$ \hfill (20)

The tight lower bound on $c_m(M)$ can be expressed in terms of the integral

$$\frac{(-1)^m}{2\pi i} \int_C \frac{\log(1 - \phi_Y(-u))}{u^{m+1}} \text{d}u,$$ \hfill (21)

where $\phi_Y(s) = \beta e^{sv_1} + (1 - \beta)e^{sv_2}$, $C$ is a contour to the left of, and parallel to, the imaginary axis, and to the right of any singularities of $\log(1 - \phi_Y(-u))$ in the left half plane.

We illustrate the lower bound (18) (calculated using (20)) in Figure 2 for the random walk with step size $X$ having a uniform distribution on $[a,b]$. The MAD of $X$ can be shown to be $(b-a)/4$. In Figure 2 we choose $b = -a = 2$ so that $\mu = 0$ and $d = 1$. Observe that upper and lower bound together provide a tight interval for all possible distribution in the ambiguity set $\mathcal{P}_{(0,1,1/2)}$.

Figure 3 shows the tight upper bound (16) and the lower bound (21) for $\mathbb{E}[W]$ with ambiguity set with $\mu = -1$, $d = b/2$ and range $[-b-2,b]$. Observe that the bounds increase with the range and the MAD (which can be shown to hold in general). For a point of reference, we also plot the exact results for one member of the ambiguity set, when generic increment having a uniform distribution on $[-b-2,b]$. 


4. Extremal GI/G/1 queue

Let us now turn to the extremal GI/G/1 queue problem, as described in the introduction. Let $W_n$ be the waiting time of customer $n$. The sequence $(W_n, n \geq 0)$ with $W_0 = 0$ satisfies the Lindley recursion

$$W_{n+1} = (W_n + V_n - U_n)^+, \quad n \geq 0.$$ 

Let $W$ be the steady-state waiting time. Since $W_n \overset{d}{=} M_n$ and $W \overset{d}{=} M$ the results for the random walk maxima likely carry over to the waiting times. The main difference is that the step size $X$ is now interpreted as the difference $V - U$ between the generic service time and generic interarrival time.
If one has mean-MAD information about both $V$ and $U$ this is more informative than mean-MAD information about $V - U$, and this additional information should lead to even sharper bounds.

### 4.1. A complete picture

The GI/G/1 queue assumes that interarrival times and service times are independent, so it is natural to assume that $V$ has ambiguity set $\mathcal{P}(\mu_V, d_V)$ and $U$ has ambiguity set $\mathcal{P}(\mu_U, d_U)$. The extremal queue problem with mean-MAD dispersion information can then be phrased as

$$\max_{P \in \mathcal{P}(\mu_V, d_V) \times \mathcal{P}(\mu_U, d_U)} \mathbb{E}[f(X)]$$

where $f$ describes $\mathbb{E}[W_n]$ or $c_m(W)$ and $X$ is the random vector with elements $U_1, V_1, U_2, V_2, \ldots$.

This is the classical setting of the extremal GI/G/1 queue treated in Rolski (1972), Eckberg Jr (1977), Whitt (1984), Chen and Whitt (2019), but with MADs instead of variances describing the ambiguity set. Let the random variables $V(3)$ and $U(3)$ follow the extremal three-point distributions $\Omega(\mu_V, d_V, a_V, b_V)$ and $\Omega(\mu_U, d_U, a_U, b_U)$, respectively.

**Theorem 4.** Consider the GI/G/1 queue with generic interarrival time $U$ with ambiguity set $\mathcal{P}(\mu_U, d_U)$ and generic service times $V$ with ambiguity set $\mathcal{P}(\mu_V, d_V)$. Consider the tight upper bounds for the transient mean waiting time $\mathbb{E}[W_n]$ and all cumulants of the steady-state waiting time $W$.

- For given interarrival time $U$, the tight upper bounds follow from the service time $V(3)$.
- For given service time $V$, the tight upper bounds follow from the interarrival time $U(3)$.
- The overall tight upper bounds follow from interarrival time $U(3)$ and service time $V(3)$.

**Proof.** Like Theorem 1, the tight bounds for $\mathbb{E}[W_n]$ follow from the general upper bound in Ben-Tal and Hochman (1972) on the expectation of a convex function of the random vector $(X_1, \ldots, X_n)$ with mean-MAD ambiguity, but now with $X_i$ replaced by $V_i - U_i$. The function describing $\mathbb{E}[W_n]$ (see Theorem 1) is indeed convex in both $V_i$ and $U_i$, and hence the result follows. Similarly, Spitzer’s formula for $c_m(W)$ (see Theorem 2) is also convex in both $V_i$ and $U_i$, and hence the tight bounds for $c_m(W)$ follow from our proof of Theorem 2.

Using the earlier results for the random walk, we present in EC.2 expressions that are helpful in evaluating the tight bounds. Table 1 shows an example of the tight bound on $\mathbb{E}[W]$ associated with $(U(3), V(3))$, also compared with other known bounds that require variance information (see EC.3). The variance of the extremal three-point distribution $\Omega(\mu, d, a, b)$ is $\frac{1}{2}(b - a)$, the maximal variance for distributions in the ambiguity set $\mathcal{P}(\mu, d)$. We thus know the variances of $U(3)$ and $V(3)$, and can calculate the other three bounds. The tight bound for $\mathbb{E}[W]$ considerably improves the other bounds for $\rho$ away from 1. In heavy traffic, Kingman’s bound is known to be asymptotically correct, and hence the other three (sharper) bounds also converge to the heavy-traffic limit as $\rho \uparrow 1$. See EC.5 for more numerical results.
Table 1  Bounds for $(1-\rho)\mathbb{E}[W]/\rho$ for $(\mu_V, d_V, a_V, b_V) = (1, 1, 0, 10)$ and $(\mu_U, d_U, a_U, b_U) = (\rho, 1, 0, 10)$.

| $\rho$ | Tight   | C & W   | Daley  | Kingman |
|--------|---------|---------|--------|---------|
|        | (Thm. 4) | (EC.15) | (EC.14) | (EC.13) |
| 0.1    | 4.06613 | 29.50020 | 29.7500 | 50.00000 |
| 0.2    | 4.12557 | 16.52810 | 17.0000 | 25.00000 |
| 0.5    | 4.41664 | 8.13750  | 8.7500  | 10.00000 |
| 0.7    | 4.67281 | 6.38567  | 6.8214  | 7.14286  |
| 0.8    | 4.79164 | 5.81773  | 6.1250  | 6.25000  |
| 0.9    | 4.90140 | 5.36711  | 5.5277  | 5.55556  |
| 0.95   | 4.95197 | 5.17469  | 5.2565  | 5.26316  |
| 0.99   | 4.99054 | 5.03364  | 5.0500  | 5.05051  |

4.2. Further comparison between MAD and variance

For the variance counterpart, Chen and Whitt (2019) also formulate a semi-infinite linear optimization problem. The crucial difference is that they cannot use the univariate function extension (as explained in Section 2), and hence should work directly with the multivariate function. This in turn implies that the dual problem cannot be solved explicitly (like in the univariate case), let alone that there is a zero duality gap. Another complication is that the multivariate function based on Spitzer’s formulas (8) and (11) cannot be expressed directly in $V$ and $U$, but rather in terms of convolutions of the distributions of $V$ and $U$. Chen and Whitt (2019) resolve these considerable challenges by several ingenious arguments, a.o. exploiting the description of $W$ as a fixed point in the stochastic equation $W \overset{d}{=} (W + V - U)^+$, and by imposing additional regularity conditions on $V$. In this way, Chen and Whitt (2019) prove a similar but weaker result than Theorem 4 for the exact same setting, but with variance as dispersion measure. They show that the extremal distributions of $U$ and $V$ both have supports on at most three points.

An important message of this paper is that with MAD the extremal distribution remains unaltered going from the univariate to the multivariate setting, and that with variance this reasoning fails. In fact, one intuitively expects formidable challenges when seeking for extremal distributions under variance constraints. This intuition is confirmed by Chen and Whitt’s formulation of the extremal distribution as the solution of a non-convex nonlinear optimization problem. While this optimization problem can be solved numerically, a closed-form solution and hence identification of the extremal distribution remains out of reach.

Under variance constraints, it is conjectured that the tight bound comes from specific two-point distributions for both $U$ and $V$. In fact, the bound (EC.15) in Table 1 holds under assumption that this conjecture is true, and was shown by Chen and Whitt (2019) to be very close to the tight upper bound. Theorem 4 rules out a similar conjecture in the MAD setting. The tight bounds in Theorem 4 always involve three-point distributions. Compared with the variance-based two-point bounds, the MAD-based three-point bounds contain more information and likely serve as better...
approximations with a smaller range of possible values. This is confirmed in Table 1 and many more numerical experiments.

5. Optimization of queueing systems

In the previous sections we have seen that using MAD instead of variance leads to simple explicit expressions for the worst-case distribution. In this section we show that we can use these expressions not only for performance analysis of queueing systems, but also to optimize them.

Let us demonstrate that for the generic dimensioning problem

\[
\min_{y \in Y} \max_{P(\mu, d, \beta)} \mathbb{E}_P[f(y, x)]
\]

with

\[
f(y, x) = c(y) + g(h_n(y, x))
\]

and

\[
h_n(y, x) = \max \{0, X_1 - y_1, \ldots, X_1 - y_1 + \cdots + X_n - y_n\},
\]

with $Y$ the feasible region for $y$, $c(\cdot)$ and $g(\cdot)$ positive convex functions, and $X_1, X_2, \ldots$ i.i.d. random variables with $P(\mu, d, \beta)$ ambiguity set. Note that $h_n(y, x)$ describes the expected maximum/waiting time after $n$ steps. One possible interpretation of the underlying random walk is that during slot $k$ the queue receives $X_k$ (new) demand and $y_k$ capacity, so that the net effect on the queue length is $X_k - y_k$. The minmax capacity vector $y$ finds the best trade-off between capacity costs $c(y)$ and quality-of-service level $g(h_n(y, x))$.

Recently developed DRO techniques in Postek et al. (2018) show that the extremal distribution in an optimization problem like (23) is independent of $y$; see Proposition EC.1. Hence, we can substitute the $3^n$ terms, which leads to a convex function in $y$, and the minimization problem over $y$ becomes tractable. As an example, select $y$ according to the linear cost function

\[
\min_{y \in Y} \max_{P(\mu, d, \beta)} \mathbb{E}_P[c^T y + c_Q \max \{0, X_1 - y_1, \ldots, X_1 - y_1 + \cdots + X_n - y_n\}],
\]

with $c$ a cost vector for capacity and $c_Q$ the cost for queueing. We know the worst-case distribution, hence (25) can be equivalently written as

\[
\min_{y \in Y} \left\{ c^T y + c_Q \sum_{\alpha \in \{1, 2, 3\}^n} \prod_{i=1}^{n} p^{(i)}_{\alpha_i} \max \{0, \tau^{(1)}_{\alpha_1} - y_1, \ldots, \tau^{(1)}_{\alpha_1} - y_1 + \cdots + \tau^{(n)}_{\alpha_n} - y_n\} \right\}.
\]

By reformulating the max-terms by linear terms, (26) can be rewritten as a linear optimization problem:

\[
\min_{y \in Y, z} c^T y + c_Q \sum_{\alpha \in \{1, 2, 3\}^n} \left( \prod_{i=1}^{n} p^{(i)}_{\alpha_i} \right) z_\alpha \\
\text{s.t.} \quad z_\alpha \geq 0, \quad \alpha \in \{1, 2, 3\}^n \\
\quad \quad z_\alpha \geq \sum_{k=1}^{K} (\tau^{(k)}_{\alpha_k} - y_k), \quad K = 1, \ldots, n; \quad \alpha \in \{1, 2, 3\}^n.
\]
Figure 4 Function $3y + \mathbb{E}[M_n]$ for $[a, b] = [-4, 4]$, $d = 2$ and $\beta = 1/2$ for the worst-case three-distribution (red curve) and the best-case two-point distribution (green curve). The dotted lines are the underlying values of $\mathbb{E}[M_n]$ and the blue curve is $3y$.

The best-case distribution gives rise to a similar LP, see Proposition [EC.1] and the two LPs together provide tight bounds for the optimum. For the sake of example, assume fixed capacity $y$ in each slot, and search for the optimal $y$ that solves

$$
\min_{y > 0} \max_{P(\mu, d, \beta)} \mathbb{E}[c \cdot y + c_Q \sum_{k=1}^n \max \{0, (X_1 - y) + \ldots + (X_k - y)\}],
$$

with $c, c_Q$ positive constants, and $M_n$ the maximum after $n$ steps of a random walk with generic increment $X - y$. For this easy example we can find the optima by visual inspection or binary search, rather than by solving the LPs. Figure 4 shows an example for $f(y, X_1, \ldots, X_n) = 3y + \mathbb{E}[M_n]$. Denote the coordinates of the optima on the lower and upper curves by $(y_{LB}, f_{LB})$ and $(y_{UB}, f_{UB})$. General DRO theory then guarantees that the optimal values are contained in the interval $(f_{LB}, f_{UB})$ for all distributions in the ambiguity set; see Corollary [EC.1].

We next discuss an adjustable DRO approach. Suppose that at time $t$ we have observed the realizations $x_1, \ldots, x_{t-1}$ of the random variables $X_1, \ldots, X_t$. Hence, for the decision on $y_t$ we can make use of that knowledge, i.e. $y_t = y_t(x_1, \ldots, x_{t-1})$. Since this leads to an NP-hard problem, often Linear Decision Rules are used in Robust Optimization, i.e.

$$
y_t = a_t + \sum_{i=1}^{t-1} b_{ti} x_i, \quad (28)
$$

where $a_t$ and $b_{ti}$ become optimization variables in the new optimization problem. Substituting (28) into (25), and following the same steps as above, results in the linear optimization problem
\[
\min_{y \in \mathcal{Y}, a, b, z} \quad c^T y + c_Q \sum_{\alpha \in \{1, 2, 3\}^n} \left( \prod_{i=1}^n p_{\alpha_i}^{(i)} \right) z_\alpha \\
s.t. \quad z_\alpha \geq 0, \quad \alpha \in \{1, 2, 3\}^n
\]

\[
z_\alpha \geq \sum_{k=1}^K \left( \tau_{\alpha_k}^{(k)} - a_k - \sum_{i=1}^k b_{ki} \tau_{\alpha_k}^{(k)} \right), \quad K = 1, \ldots, n; \quad \alpha \in \{1, 2, 3\}^n.
\]

6. Outlook
This paper offers a deep understanding of why MAD simplifies comparable variance-based optimization problems, in a way that is almost unreasonably effective, resulting in a full solution to the extremal queue problem with mean-MAD constraints. When partial information is available in the form of mean, range and MAD, we have obtained the sharpest possible bounds. Through basic statistical estimation of this partial information, the GI/G/1 queue becomes a data-driven model that adjusts to available training data, for which this paper presents tight performance guarantees.

The key idea of using MAD instead of variance as dispersion measure, is likely applicable to many other queueing system. Examples are queues with dependency and correlation structures in the series \(\{U_n\}\) and \(\{V_n\}\), multi-server GI/G/c queue and networks of queues. Indeed, most of the key performance measures for such systems are expectations of functions that are convex in the random variables (see e.g., Shaked and Shanthikumar (1988)), and therefore the DRO-MAD approach can be used.

The MAD perspective is of interest beyond queueing theory, because the search for extremal distributions of convex functions is relevant in many other settings. For instance, the GI/G/1 queue also models shortfall in a production-inventory system, see e.g., Glasserman (1997), Bradley and Glynn (2002), and delays in appointment scheduling, Kong et al. (2013), Mak et al. (2014), Qi (2016), and hence our approach can be immediately applied to these settings. More generally, whenever a performance measure can be viewed as a convex function of i.i.d. random variables with mean-MAD ambiguity (e.g., nested max-operators in production systems), our approach will identify the extremal distribution and tight bounds.

It is well-known that the use of probability distributions in stochastic systems often leads to computationally intractability (e.g., calculation of high dimensional convolutions). Therefore, Bandi and Bertsimas (2012), Bandi et al. (2015), Whitt and You (2017) suggest to use uncertainty sets instead of probability distributions. The MAD approach described in this paper can serve in many situations as an alternative (not per se better), bringing new opportunities: (i) The uncertainty set approach yields a worst-case scenario. Our approach yields both worst-case and best-case distributions, i.e., both upper and lower bounds. (ii) In stochastic systems one often studies convex
functions in the stochastic variables. In the uncertainty set approach it is in general hard (in fact, NP-hard) to find worst-case scenarios for such convex functions. Our approach can easily find worst-case distributions as shown in this paper. (iii) Contrary to the uncertainty set approach, our approach can easily be extended to optimization problems, and even to adjustable robust optimization problems.

The MAD approach stays close to the common practice in the stochastic field, namely to use probability distributions to model uncertainty. The nucleus of the MAD approach consists of the explicitly solvable dual LP described in Section 2. A simple reasoning then showed that this solution is independent of the precise objective function (in this paper describing waiting time moments of the GI/G/1 queue). Hence, the MAD approach is a generic, computationally tractable way to analyze stochastic systems, with a host of potential applications.

Acknowledgments
The authors would like to thank Daniel Kuhn and Krzysztof Postek for pointing out the primal-dual reasoning that gives the intuitive proof of Theorem 1 and Marko Boon for helping with the experiments in Section EC.2
E-Companion to “MAD dispersion measure makes extremal queue analysis simple”

EC.1. Properties of MAD

We recall some well known properties of the MAD, see e.g. [Ben-Tal and Hochman 1985]. Denote by $\sigma^2$ the variance of the random variable $X$, whose distribution is known to belong to the set $\mathcal{P}(\mu,d)$. Then

$$\frac{d^2}{4\beta(1-\beta)} \leq \sigma^2 \leq \frac{d(b-a)}{2}.$$ 

In particular, since

$$d^2 \leq 4\beta(1-\beta)\sigma^2 \leq \sigma^2,$$

it holds that $d \leq \sigma$. For a proof, we refer the reader to [Ben-Tal and Hochman 1985]. For some distributions, an explicit formula for $d$ is available:

- Uniform distribution on $[a,b]$:
  $$d = \frac{1}{4}(b-a)$$

- Normal distribution $N(\mu,\sigma^2)$:
  $$d = \sqrt{\frac{2}{\pi}\sigma}$$

- Gamma distribution with parameters $\lambda$ and $k$ (for which $\mu = k/\lambda$):
  $$d = \frac{2k^k}{\Gamma(k) \exp(k)} \frac{1}{\lambda}.$$ 

The MAD is known to satisfy the bound

$$0 \leq d \leq \frac{2(b - \mu)(\mu - a)}{b - a}. \quad (\text{EC.1})$$

Let $\beta = \mathbb{P}(X \geq \mu)$. For example, in the case of continuous symmetric distribution of $X$ we know that $\beta = 0.5$. This quantity is known to satisfy the bounds:

$$\frac{d}{2(b - \mu)} \leq \beta \leq 1 - \frac{d}{2(\mu - a)}. \quad (\text{EC.2})$$

EC.2. Representations for the tight bounds

We now present some efficient ways of calculating the tight bounds identified in this paper. We first discuss the contour integral representations.
EC.2.1. Contour integral representations

Consider the random walk with generic step size \( X \). Formal solutions of the distribution of \( M_n \) and \( M \) can be expressed in terms of complex contour integrals (see Abate et al. (1993), Janssen et al. (2015) for the algorithmic aspects of these contour integrals). Assume that \( \phi_X(s) = \mathbb{E}[e^{sX}] \) is analytic for complex \( s \) in the strip \( |\text{Re}(s)| < \delta \) for some \( \delta > 0 \). A sufficient condition is that the moment generating function \( \phi_X(s) \) is finite in a neighborhood of the origin, and hence all moments of \( X \) exist. Then

\[
\mathbb{E}[e^{-sM}] = \exp \left\{ -\frac{1}{2\pi i} \int_{\mathcal{C}} \frac{s}{u(s-u)} \log(1-\phi_X(-u)) du \right\},
\]

(EC.3)

where \( s \) is a complex number with \( \text{Re}(s) \geq 0 \), \( \mathcal{C} \) is a contour to the left of, and parallel to, the imaginary axis, and to the right of any singularities of \( \log(1-\phi_X(-u)) \) the left half plane. From (EC.3) contour integral expressions for the cumulants follow by differentiation:

\[
c_m(M) = \frac{(-1)^m}{2\pi i} \int_{\mathcal{C}} \frac{\log(1-\phi_X(-u))}{u^{m+1}} du.
\]

(EC.4)

Consider \( X = X_{(3)} \) with a three point distribution on values \( \{a, b, c\} \) with probabilities \( p_a, p_b, p_c \) and moment generating function

\[
\phi_{X_{(3)}}(s) = p_a e^{sa} + p_b e^{sb} + p_c e^{sc}.
\]

(EC.5)

Notice that all moments of \( X_{(3)} \) exist, and hence \( \phi_{X_{(3)}}(s) \) satisfies the assumption required for representation (EC.3) to hold. Since \( X_{(3)} \) follows the extremal three-point distribution associated with the tight upper bounds for \( c_m(M) \), we obtain the representation:

\[
c_m(M) = \frac{(-1)^m}{2\pi i} \int_{\mathcal{C}} \frac{\log(1-\phi_{X_{(3)}}(-u))}{u^{m+1}} du,
\]

(EC.6)

which serves as expression for the tight upper bound (16).

EC.2.2. Numerical experiments with contour integrals

Numerical aspects of integrals of the above type have been discussed in e.g., Abate et al. (1993), Janssen et al. (2015), Chen and Whitt (2018). For distributions with support on a finite set of points, potential numerical problems can arise, because \( |\text{Re}(\phi_X(u))| \) does not converge to zero as \( |u| \to \infty \); see Abate and Whitt (1992), Chen and Whitt (2018). For the three-point distributions required in this paper we have performed extensive numerical experiments with (EC.6). These experiments confirmed that the integrals can be calculated up to high accuracy with standard integration routines in Mathematica (our code is available upon request).

For many parameter values \( a, b, \mu, d \) such that (EC.1) holds, we have calculated \( \mathbb{E}[M] \) for generic increment \( X_{(3)} \) using (EC.6), and compared this with results from extensive stochastic simulations.
We also compared the results with a third numerical procedure, known to be extremely stable and accurate. Let us explain the third procedure, which might be of independent interest.

Choose the boundaries of the support as multiples of $\beta = |\mu|$ by writing that $a = -s\beta$ and $b = m\beta$ with $s, m$ positive integers. Denote by $M_\beta = M/\beta$ the normalized steady-state waiting time. We then get

$$M_\beta \overset{d}{=} (M_\beta + X_\beta)^+, \tag{EC.7}$$

with $X_\beta = X/\beta$ a discrete random variable with support $\{-s, -1, m\}$ and MAD

$$d_\beta := \mathbb{E}[|X_\beta - \mathbb{E}[X_\beta]|] = \frac{1}{\beta} \mathbb{E}[|X - \mathbb{E}[X]|] = d.$$  

Define $X_\beta = A_\beta - s$, so that

$$M_\beta \overset{d}{=} (M_\beta + A_\beta - s)^+$$

for a discrete random variable $A_\beta$ with support $\{0, s-1, s+m\}$ and probability generating function

$$\mathbb{E}[z^{A_\beta}] = p_a + p_\mu z^{s-1} + p_b z^{m+s},$$

with

$$p_a = \frac{d_\beta}{2(s-1)}, \quad p_\mu = 1 - \frac{d_\beta}{2(s-1)} - \frac{d_\beta}{2(m+1)}, \quad p_b = \frac{d_\beta}{2(m+1)}.$$  

Notice that $\mathbb{E}[A_\beta] = s-1$. The resulting discrete queueing system is sometimes referred to as a bulk service queue. Let $r_0$ be the unique zero of $z^s - \mathbb{E}[z^{A_\beta}]$ with real $z > 1$. For any $\varepsilon > 0$ with $1 + \varepsilon < r_0$,

$$\mathbb{E}[w^{M_\beta}] = \exp\left(\frac{1}{2\pi i} \int_{|z| = 1+\varepsilon} \ln\left(\frac{w-z}{1-z}\right) \frac{(z^s - \mathbb{E}[z^{A_\beta}])'}{z^s - \mathbb{E}[z^{A_\beta}]} \, dz\right) \tag{EC.8}$$

holds when $|w| < 1 + \varepsilon$. Alternatively,

$$\mathbb{E}[w^{M_\beta}] = \frac{(s - \mathbb{E}[A_\beta])(w-1)}{w^s - A(w)} \prod_{k=1}^{s-1} \frac{w-z_k}{1-z_k}$$  

that holds for all $w$, $|w| < r_0$, in which $z_1, \ldots, z_{s-1}$ are the $s-1$ zeros of $z^s - \mathbb{E}[z^{A_\beta}]$ in $|z| < 1$. Upon differentiation, \eqref{EC.7} and \eqref{EC.8} provide expressions for all cumulants of $M_\beta$ that are known to allow for accurate numerical evaluation, see Janssen et al. (2015). We have then performed for a wide range of parameters, the following experiment:

1. Fix $\beta$, and then choose integers $s$ and $m$. In this way we create a standard bulk service queue with discrete-valued generic increment $A_\beta$.
2. For ranging $d_\beta$, calculate $\mathbb{E}[M_\beta]$ using root-finding procedures and \eqref{EC.8} or using the contour integral \eqref{EC.7}.
3. Calculate

$$\mathbb{E}[M] = \frac{-1}{2\pi i} \int_C \log(1 - (p_a e^{-ua} + p_b e^{-ub} + p_c e^{-uc})) \, du.$$  

4. Check whether $\mathbb{E}[M] = \beta \mathbb{E}[M_\beta]$.  

ec3
Calculations for $\mathbb{E}[W_n]$ and $c_n(W)$ in the GI/G/1 queue can be performed using similar expressions as for the random walk. Let the random variable $V_{(3)}$ follow a three-point distribution on values \{s_1, s_2, s_3\} with probabilities

$$p_1 = \frac{d_V}{2(\mu_V - a_V)}, \quad p_2 = 1 - \frac{d_V}{2(\mu_V - a_V)} - \frac{d_V}{2(b_U - \mu_U)}, \quad p_3 = \frac{d_V}{2(b_U - \mu_U)}, \quad (EC.9)$$

with $0 \leq a_V < \mu_U < b_U$, so that $V_{(3)}$ has mean $\mu_U$ and MAD $d_V$. Similarly, let $U_{(3)}$ have a three-point distribution on values \{t_1, t_2, t_3\} with probabilities

$$r_1 = \frac{d_U}{2(\mu_U - a_U)}, \quad r_2 = 1 - \frac{d_U}{2(\mu_U - a_U)} - \frac{d_U}{2(b_U - \mu_U)}, \quad r_3 = \frac{d_U}{2(b_U - \mu_U)} \quad (EC.10)$$

and $0 \leq a_U < \mu_U < b_U$, so that $U_{(3)}$ has mean $\mu_U$ and MAD $d_U$.

We then have the representation, see also Chen and Whitt (2019),

$$\mathbb{E}[W_n] = \sum_{k=1}^{n} \frac{1}{k} \sum_{\sum k_it_i = k} \max\{0, \sum_{i=1}^{3} k_is_i - \sum_{j=1}^{3} l_it_i\} \cdot P(k_1, k_2, k_3) \cdot R(l_1, l_2, l_3) \quad (EC.11)$$

with

$$P(k_1, k_2, k_3) = \frac{k!}{k_1!k_2!k_3!} p_1^{k_1} p_2^{k_2} p_3^{k_3}, \quad R(l_1, l_2, l_3) = \frac{k!}{l_1!l_2!l_3!} r_1^{l_1} r_2^{l_2} r_3^{l_3},$$

which requires summing $O(n^5)$ terms.

Let $\phi_{V_{(3)}}(s)$ and $\phi_{U_{(3)}}(s)$ denote the moment generating functions of $V_{(3)}$ and $U_{(3)}$. The tight upper bounds on $c_m(W)$ are given by

$$c_m(W) \leq \frac{(-1)^m}{2\pi i} \int_{\mathcal{C}} \frac{\log(1 - \phi_{V_{(3)}}(-u)\phi_{U_{(3)}}(u))}{u^{m+1}} du, \quad (EC.12)$$

where $\mathcal{C}$ is a contour to the left of, and parallel to, the imaginary axis, and to the right of any singularities of $\log(1 - \phi_{V_{(3)}}(-u)\phi_{U_{(3)}}(u))$ in the left half plane. Again comparing with extensive simulation, we have found the expression (EC.12) accurate and hence suitable for calculating the tight bounds.

### EC.3. Distribution-free upper bounds for the GI/G/1 queue

Consider the steady-state queue length $W$ in the GI/G/1 queue, which satisfies $W \overset{d}{=} (W + V - U)^+$. Denote by $\sigma_V^2$ and $\sigma_U^2$ the variances of $U$ and $V$, respectively. Let $\rho = \mathbb{E}[V]/\mathbb{E}[U] < 1$. The following bounds on $\mathbb{E}[W]$ only require information about the first two moments of $U$ and $V$:

- **Kingman’s upper bound:**
  $$\mathbb{E}[W] = \frac{\sigma_V^2 + \sigma_U^2}{2(\mathbb{E}[U] - \mathbb{E}[V])}. \quad (EC.13)$$

- **Daley’s upper bound:**
  $$\mathbb{E}[W] = \frac{\sigma_V^2 + \rho(2 - \rho)\sigma_U^2}{2(\mathbb{E}[U] - \mathbb{E}[V])}. \quad (EC.14)$$
• Upper bound of Chen and Whitt (2019) based on the two-point conjecture:

$$\mathbb{E}[W] = \frac{\sigma^2_V + \kappa(\rho)\sigma^2_U}{2(\mathbb{E}[U] - \mathbb{E}[V])},$$

(EC.15)

with $\kappa(\rho) = 2\rho(1-\rho)/(1-\delta)$ and $\delta \in (0,1)$ the solution of $\delta = \exp(-(1-\delta)/\rho)$.

EC.4. Recent DRO results

In Postek et al. (2018), the following result was proved (for a much larger class of functions $f(y, X)$ than in Section 5):

**Proposition EC.1.** If $f(y, \cdot)$ is convex,

$$\sup_{P \in \mathcal{P}(\mu,d)} \mathbb{E}_P[f(y, X)] = g_U(y) = \sum_{\alpha \in \{1,2,3\}^n} \prod_{i=1}^n p^{(i)}_{\alpha_i} f(y, \tau^{(1)}_{\alpha_i}, \ldots, \tau^{(n)}_{\alpha_i}),$$

(EC.16)

with $p^{(i)}_{\alpha_i}, \tau^{(i)}_{\alpha_i}$ defined as in Theorem 1. If $f(y, \cdot)$ is concave,

$$\sup_{P \in \mathcal{P}(\mu,d,\beta)} \mathbb{E}_P[f(y, X)] = g_L(y) = \sum_{\alpha \in \{1,2\}^n} \prod_{i=1}^n q^{(i)}_{\alpha_i} f(y, \upsilon^{(1)}_{\alpha_i}, \ldots, \upsilon^{(n)}_{\alpha_i}),$$

(EC.17)

with $q^{(i)}_{\alpha_i}, \upsilon^{(i)}_{\alpha_i}$ defined in (19).

Hence, $g_U(\cdot)$ in (EC.16) inherits the convexity in $y$ from $f(\cdot, X)$ and its functional form depends only on the form of $f(\cdot, X)$ (and similarly for $g_L(\cdot)$). The upper and lower bound give a closed interval for

$$\text{Val}_P(y) = \mathbb{E}_P[f(y, X)] \quad \forall P \in \mathcal{P}(\mu,d,\beta).$$

(EC.18)

**Corollary EC.1.** If $f(y, \cdot)$ is convex for all $y$ then $\text{Val}_P(y) \in [g_L(y), g_U(y)] \quad \forall P \in \mathcal{P}(\mu,d,\beta)$. If $f(y, \cdot)$ is concave for all $y$ then $\text{Val}_P(y) \in [g_U(y), g_L(y)] \quad \forall P \in \mathcal{P}(\mu,d,\beta)$.

From Proposition EC.1 we see that the extremal distribution is independent of $y$. Hence, we can substitute the $3^n$ terms. This leads to a convex function in $y$, and hence the minimization problem over $y$ is tractable.

EC.5. Further numerical results for the bounds

We now complement Table 1 with some more numerical values for the bounds on $\mathbb{E}[W]$. Table EC.1 gives the unscaled values of $\mathbb{E}[W]$ for the same parameter values as in Table 1.

The variance bounds are often reported in terms of the squared coefficient of variation (variance divided by the square of the mean), see Chen and Whitt (2019). For the extremal distributions with $(\mu_V, d_V, a_V, b_V) = (\rho, d_V, 0, b_V)$ and $(\mu_U, d_U, a_U, b_U) = (1, d_U, 0, b_U)$ this gives

$$c^2_{\psi_V} = \frac{\sigma^2_V}{\mu^2_V} = \frac{d_Vb_V}{2\rho^2}, \quad c^2_{\psi_U} = \frac{\sigma^2_U}{\mu^2_U} = \frac{d_Vb_U}{2}.$$
Table EC.1  Bounds for $E[W]$ for $(\mu_U, d_U, a_U, b_U) = (1, 1, 0, 10)$ and $(\mu_V, d_V, a_V, b_V) = (\rho, 1, 0, 10)$.

| $\rho$ | Tight (Thm. 4) | C & W (EC.15) | Daley (EC.14) | Kingman (EC.13) |
|-------|----------------|----------------|---------------|-----------------|
| 0.1   | 0.45179        | 3.27780        | 3.30556       | 5.55556         |
| 0.2   | 1.03139        | 4.13203        | 4.25000       | 6.25000         |
| 0.5   | 4.11664        | 8.13750        | 8.75000       | 10.0000         |
| 0.7   | 10.90322       | 14.89990       | 15.91670      | 16.66670        |
| 0.8   | 19.16657       | 23.27090       | 24.50000      | 25.00000        |
| 0.9   | 44.11263       | 48.30400       | 49.75000      | 50.00000        |
| 0.95  | 94.08734       | 98.31910       | 99.87500      | 100.00000       |
| 0.99  | 494.06334      | 498.33100      | 499.97500     | 500.00000       |

Fixing the squared coefficient of variations $c^2_U$ and $c^2_V$ is equivalent with choosing the MADs as

$$d_V = \frac{2\rho^2 c^2_V}{b_V}, \quad d_U = \frac{2c^2_U}{b_U}. \quad \text{(EC.19)}$$

We next present in Tables EC.2-EC.5 some further numerical results, for $c^2_U = c^2_V = 0.5$ and $c^2_U = c^2_V = 4$.

Table EC.2  Bounds for $E[W]$ for $(\mu_V, d_V, a_V, b_V) = (\rho, d_V, 0, 10)$ and $(\mu_U, d_U, a_U, b_U) = (1, d_U, 0, 10)$ with $d_V, d_U$ as in (EC.19) and $c^2_U = c^2_V = 0.5$.

| $\rho$ | Tight (Thm. 4) | C & W (EC.15) | Daley (EC.14) | Kingman (EC.13) |
|-------|----------------|----------------|---------------|-----------------|
| 0.1   | 0.00785        | 0.05278        | 0.05555       | 0.28055         |
| 0.2   | 0.02230        | 0.11320        | 0.12500       | 0.32500         |
| 0.5   | 0.14921        | 0.43875        | 0.50000       | 0.62500         |
| 0.7   | 0.48818        | 1.06499        | 1.16667       | 1.24167         |
| 0.8   | 0.99509        | 1.87709        | 2.00000       | 2.05000         |
| 0.9   | 2.85149        | 4.35540        | 4.50000       | 4.52500         |
| 0.95  | 7.29378        | 9.34441        | 9.50000       | 9.51250         |
| 0.99  | 46.78335       | 49.33560       | 49.50000      | 49.50250        |

Table EC.3  Bounds for $E[W]$ for $(\mu_V, d_V, a_V, b_V) = (\rho, d_V, 0, 10)$ and $(\mu_U, d_U, a_U, b_U) = (1, d_U, 0, 10)$ with $d_V, d_U$ as in (EC.19) and $c^2_U = c^2_V = 4$.

| $\rho$ | Tight (Thm. 4) | C & W (EC.15) | Daley (EC.14) | Kingman (EC.13) |
|-------|----------------|----------------|---------------|-----------------|
| 0.1   | 0.09358        | 0.42224        | 0.44444       | 2.24444         |
| 0.2   | 0.26429        | 0.90562        | 1.00000       | 2.60000         |
| 0.5   | 2.05142        | 3.51000        | 4.00000       | 5.00000         |
| 0.7   | 6.76335        | 8.51991        | 9.33333       | 9.93333         |
| 0.8   | 13.18168       | 15.01670       | 16.00000      | 16.40000        |
| 0.9   | 32.95685       | 34.84320       | 36.00000      | 36.20000        |
| 0.95  | 72.84232       | 74.75520       | 76.00000      | 76.10000        |
| 0.99  | 392.74278      | 394.68400      | 396.00000     | 396.02000       |
Table EC.4  Bounds for \((1 - \rho)\mathbb{E}[W]/\rho\) for \((\mu_V, d_V, a_V, b_V) = (\rho, d_V, 0, 10)\) and \((\mu_V, d_U, a_U, b_U) = (1, d_U, 0, 10)\) with \(d_V, d_U\) as in (EC.19) and \(c_{U}^2 = c_{V}^2 = 0.5\).

| \(\rho\)  | Tight (Thm. 4) | C & W (EC.15) | Daley (EC.14) | Kingman (EC.13) |
|----------|----------------|---------------|--------------|-----------------|
| 0.1      | 0.07070        | 0.47502       | 0.50000      | 2.52500         |
| 0.2      | 0.08922        | 0.45281       | 0.50000      | 1.30000         |
| 0.5      | 0.14921        | 0.43875       | 0.50000      | 0.62500         |
| 0.7      | 0.20922        | 0.45642       | 0.50000      | 0.53214         |
| 0.8      | 0.24877        | 0.46927       | 0.50000      | 0.51250         |
| 0.9      | 0.31683        | 0.48393       | 0.50000      | 0.50277         |
| 0.95     | 0.38388        | 0.49181       | 0.50000      | 0.50065         |
| 0.99     | 0.47255        | 0.49833       | 0.50000      | 0.50002         |

Table EC.5  Bounds for \((1 - \rho)\mathbb{E}[W]/\rho\) for \((\mu_V, d_V, a_V, b_V) = (\rho, d_V, 0, 10)\) and \((\mu_V, d_U, a_U, b_U) = (1, d_U, 0, 10)\) with \(d_V, d_U\) as in (EC.19) and \(c_{U}^2 = c_{V}^2 = 4\).

| \(\rho\)  | Tight (Thm. 4) | C & W (EC.15) | Daley (EC.14) | Kingman (EC.13) |
|----------|----------------|---------------|--------------|-----------------|
| 0.1      | 0.84228        | 3.80016       | 4.00000      | 20.20000        |
| 0.2      | 1.05719        | 3.62248       | 4.00000      | 10.40000        |
| 0.5      | 2.05142        | 3.51000       | 4.00000      | 5.00000         |
| 0.7      | 2.89858        | 3.65139       | 4.00000      | 4.25714         |
| 0.8      | 3.29542        | 3.75418       | 4.00000      | 4.10000         |
| 0.9      | 3.66187        | 3.87146       | 4.00000      | 4.02222         |
| 0.95     | 3.83381        | 3.93449       | 4.00000      | 4.0526          |
| 0.99     | 3.96710        | 3.98671       | 4.00000      | 4.00020         |

References

Abate J, Choudhury G, Whitt W (1993) Calculation of the GI/G/1 waiting-time distribution and its cumulants from Pollaczek’s formulas. *Archiv fur Elektronik und Ubertragungstechnik (International Journal of Electronics and Communication)* 47(5/6):311–321.

Abate J, Whitt W (1992) The Fourier-series method for inverting transforms of probability distributions. *Queueing Systems* 10(1-2):5–87.

Asmussen S (2003) *Applied Probability and Queues* (New York: Springer-Verlag), second edition.

Bandi C, Bertsimas D (2012) Tractable stochastic analysis in high dimensions via robust optimization. *Mathematical Programming* 134(1):23–70.

Bandi C, Bertsimas D, Youssef N (2015) Robust queueing theory. *Operations Research* 63(3):676–700.

Ben-Tal A, Hochman E (1972) More bounds on the expectation of a convex function of a random variable. *Journal of Applied Probability* 9:803–812.

Ben-Tal A, Hochman E (1985) Approximation of expected returns and optimal decisions under uncertainty using mean and mean absolute deviation. *Zeitschrift für Operations Research* 29(7):285–300.
Bradley J, Glynn P (2002) Managing capacity and inventory jointly in manufacturing systems. *Management Science* 48(2):273–288.

Chen Y, Whitt W (2018) Algorithms for the upper bound mean waiting time in the GI/GI/1 queue. *Submitted to INFORMS Journal in Computing*, Preprint.

Chen Y, Whitt W (2019) Extremal GI/GI/1 queues given two moments. *Submitted to Operations Research*, Preprint.

Chung K (2001) *A Course in Probability Theory* (Academic Press).

Cohen J (1982) *The Single Server Queue* (Amsterdam: North-Holland Publishing Co.), second edition.

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

Feller W (1971) *An Introduction to Probability Theory and its Applications. Vol. II.* Second edition (New York: John Wiley & Sons Inc.).

Glasserman P (1997) Bounds and asymptotics for planning critical safety stocks. *Operations Research* 45(2):244–257.

Janssen A, van Leeuwaarden J, Mathijsen B (2015) Novel heavy-traffic regimes for large-scale service systems. *SIAM Journal of Applied Mathematics* 75(2):787–812.

Kingman JF (1962) Some inequalities for the queue GI/G/1. *Biometrika* 49(3/4):315–324.

Kong Q, Lee CY, Teo CP, Zheng Z (2013) Scheduling arrivals to a stochastic service delivery system using copositive cones. *Operations Research* 61(3):711–726.

Mak HY, Rong Y, Zhang J (2014) Appointment scheduling with limited distributional information. *Management Science* 61(2):316–334.

Postek K, Ben-Tal A, Den Hertog D, Melenberg B (2018) Robust optimization with ambiguous stochastic constraints under mean and dispersion information. *Operations Research* 66(3):814–833.

Qi J (2016) Mitigating delays and unfairness in appointment systems. *Management Science* 63(2):566–583.

Rolski T (1972) Some inequalities for GI/M/n queues. *Applicationes Mathematicae* 1(13):42–47.

Shaked M, Shanthikumar JG (1988) Stochastic convexity and its applications. *Advances in Applied Probability* 20(2):427–446.

Spitzer F (1956) A combinatorial lemma and its application to probability theory. *Transactions of the American Mathematical Society* 82:323–339, ISSN 0002-9947.

Whitt W (1984) On approximations for queues, I: Extremal distributions. *AT&T Bell Laboratories Technical Journal* 63(1):115–138.

Whitt W, You W (2017) Using robust queueing to expose the impact of dependence in single-server queues. *Operations Research* 66(1):184–199.