Proceedings of the 2024 Winter Simulation Conference
H. Lam, E. Azar, D. Batur, S. Gao, W. Xie, S. R. Hunter, and M. D. Rossetti, eds.

COMPARATIVE ANALYSIS OF DISTANCE METRICS FOR DISTRIBUTIONALLY ROBUST OPTIMIZATION IN QUEUING SYSTEMS: WASSERSTEIN VS. KINGMAN

Hyung-Khee Eun[1], Sara Shashaani[1], and Russell R. Barton[2]

[1]Fitts Dept. of Industrial and Systems Eng., North Carolina State University, Raleigh, NC, USA
[2] Smeal College of Business, Pennsylvania State University, University Park, PA, USA

ABSTRACT

This study examines the effectiveness of different metrics in constructing ambiguity sets for Distributionally Robust Optimization (DRO). Two main approaches for building ambiguity sets are the moment- and the discrepancy-based approaches. The latter is more widely adopted because it incorporates a broader range of distributional information beyond moments. Among discrepancy-based metrics, the Wasserstein distance is often preferred for its advantageous properties over ϕ-divergence. In this study, we propose a moment-based Kingman distance, an approximation of mean waiting time in G/G/1 queues, to determine the ambiguity set. We demonstrate that the Kingman distance provides a straightforward and efficient method for identifying worst-case scenarios for simple queue settings. In contrast, the Wasserstein distance requires exhaustive exploration of the entire ambiguity set to pinpoint the worst-case distributions. These findings suggest that the Kingman distance could offer a practical and effective alternative for DRO applications in some cases.

1 INTRODUCTION

The Wasserstein distance, sometimes referred to as transport distance, Mallow's distance, or earthmover distance, is a metric for the difference between two probability distributions (Panaretos and Zemel 2019). There has been recent interest in the use of the Wasserstein distance for robust analysis and DRO of stochastic models (Kuhn et al. 2019; Blanchet et al. 2022). This metric is used to bound differences from the nominal probability distributions, and results within these bounds are used as robust performance bounds for the system.

Simple queues are often key components of larger discrete-event simulation models, and in simulation optimization settings, average waiting time or, equivalently, number-in-system measures are typically the focus. These performance metrics are known to be strongly affected by the first two moments of the queuing model's probability distributions. In this case, distributions at the extremes of the Wasserstein bounds often do not correspond to extremes of system performance. This study examines which probability distribution distance metric aligns better with differences in system performance for a simple $G/G/1$ and the capacitated $G/G/1/k$ queue: a moment-based metric using the Kingman approximation (Kingman 1961) or the Wasserstein distance. It is known that the discrepancy-based approach utilizes more distributional information than just the moments (Bayraksan and Love 2015). The moment-based approach assumes that only certain moments are known exactly, while ignoring other available information about the distribution (Gao and Kleywegt 2023). Nevertheless, this study serves as a foundational exploration, suggesting that in certain applications, such as simple queuing systems, a moment-based metric may offer superior performance in DRO problems.

The paper is organized as follows. We review important concepts in distributionally in robust analysis and discrete-event simulation in Section 2. We examine use of the moment-based Kingman distance metric and its advantages over a Wasserstein distance metric for a G/G/1 queue in Section 3. Section 4 leaves the reader with a summary of insights and open questions and challenges ahead.

2 DISTRIBUTIONALLY ROBUST OPTIMIZATION

DRO seeks to optimize problems that depend on uncertain parameters often learned from data. DRO hedges against risk of selecting an inferior system due to small samples or other corruptions in the data. The main application of DRO has been in data-driven settings to avoid overfitting and failing to generalize on out-of-sample data (Bertsimas and Van Parys 2022) or for adversarial learning (Blanchet et al. 2022). The focus in this study is its use for stochastic systems.

We let the stochastic performance of the system under decision $x \in \mathcal{X} \subseteq \mathbb{R}^d$ and a random input data $\xi \in \Xi$ that follows probability distribution ν be denoted by $F(x, \xi)$ and its expectation denoted by

$$f(x, \nu) := \mathbb{E}_\nu[F(x, \xi)]. \tag{1}$$

When addressing uncertainty in the input data, it is natural to consider a set of distributions with some perturbation from the distribution fitted empirically to the data at hand, which we henceforth refer to as the *nominal* distribution, and consider the performance of the system (decision) being evaluated under all possible input distributions. The system under study is then better (say less costly) with alternative $\mathbf{x}_1$ than $\mathbf{x}_2$ if $f(x_1, \nu) < f(x_2, \nu)$ for all possible input distributions ν, which is to say under the worst possible input distribution. This is why DRO can be viewed as a way to hedge the risks associated with input uncertainty (Rahimian and Mehrotra 2019). We will further review input uncertainty in the next section.

In DRO, an ambiguity set $\mathcal{P}$ is constructed that includes distributions within a certain discrepancy from the nominal distribution. The implicit assumption is that the true input distribution resides within the ambiguity set. Decisions are then made to ensure robustness against the variations within this set that minimizes the maximum value of the objective function across all the distributions in the ambiguity set (Blanchet et al. 2022). There are two other methods we can compare DRO with. The first, stochastic optimization (SO), assumes we know the underlying distribution ν, which often is not true. Even if we observe multiple realizations of ξ, ν still may not be known exactly, and use of a distribution different from ν due to the input uncertainty may result in suboptimal decisions. The second, robust optimization (RO), on the other hand, does not rely on any specific distributional information, considering only ξ over support $\mathcal{U}$. Therefore, (DRO) acts as an intermediary framework by utilizing partial distributional information, obtained from historical data or domain-specific knowledge, to form the ambiguity set $\mathcal{P}$, providing a more balanced approach between the full distributional knowledge of SO and the distribution-agnostic perspective of RO. Specifically, if $\mathcal{P}$ contains only the true distribution of ξ, DRO reduces to SO. In addition, if $\mathcal{P}$ contains all probability distributions on the support of ξ, the DRO reduces to RO.

$$\inf_{x \in \mathcal{X}} \mathbb{E}_\nu[F(x, \xi)] \tag{SO}$$

$$\inf_{x \in \mathcal{X}} \sup_{\xi \in \mathcal{U}} F(x, \xi) \tag{RO}$$

$$\inf_{x \in \mathcal{X}} \sup_{\nu \in \mathcal{P}} \mathbb{E}_\nu[F(x, \xi)] \tag{DRO}$$

Since DRO finds a decision that minimizes the worst-case of the objective function among all probability measures in the ambiguity set, the key to DRO is how to construct the ambiguity set $\mathcal{P}$.

2.1 Input Uncertainty in Stochastic Simulation

DRO has a direct link to input uncertainty of stochastic systems. When we simulate a real-world system, we typically use distributions fitted to data assumed to be random samples from the true real-world distribution. However, generally the true distribution is not known precisely and the data we have is finite. As a result, there is statistical error in estimating the input models. This error is called input uncertainty. For instance, in queuing models, the input uncertainty often arises from the distributions of interarrival and service times that are unknown or partially known.

Let us denote by ν the input model (distribution). In this section, we show the impact of input uncertainty on the estimation of the performance measure $f(\nu)$ defined similar to (1) but dropping x for simplicity, i.e., expectation of the stochastic simulation outputs F generated from random inputs $\xi \sim \nu$. If we assume ν to be from a certain parametric family, an estimate of ν is reduced to estimating its parameters. If we have no parametric assumption on ν, its estimation is the empirical distribution (Barton et al. 2022). We denote ν_0 as the true probabilistic description, and $\hat{\nu}$ as the fitted input model. Then a natural point estimate of $f(\nu_0)$ is

$$\bar{f}(\hat{\nu}) = \frac{1}{r}\sum_{j=1}^{r} F_j(\hat{\nu}),$$

where $F_j(\hat{\nu}) := F(\xi_j \sim \hat{\nu})$ is the jth identically distributed simulation replication under input model $\hat{\nu}$. Assuming that the simulation replications $j = 1, 2, \cdots, r$ are independent, the decomposition shows

$$F_j(\hat{\nu}) - f(\nu_0) = [F_j(\hat{\nu}) - f(\hat{\nu})] + [f(\hat{\nu}) - f(\nu_0)],$$

where the first term is the simulation error and the second term is the input uncertainty error. By the law of total variance, we get

$$\mathrm{Var}(\bar{f}(\hat{\nu})) = \frac{\mathbb{E}[\mathrm{Var}(F_j(\hat{\nu})|\hat{\nu})]}{r} + \mathrm{Var}(f(\hat{\nu})), \tag{2}$$

where the variance and expectation on the right hand side is with respect to the probability distribution of the random $\hat{\nu}$ (uncertainty in the fitted input model) and the inner variance in $\mathbb{E}[\mathrm{Var}(F_j(\hat{\nu})|\hat{\nu})]$ is with respect to the simulation outputs' probability distribution. Similar to the decomposition above, we see that the estimator's variance contains the variance due to the simulation error and variance due to uncertainty in the fitted input model. In addition, the mean square error

$$\mathbb{E}[(f(\hat{\nu}) - f(\nu_0))^2] = (\mathbb{E}[f(\hat{\nu})] - f(\nu_0))^2 + \mathrm{Var}(f(\hat{\nu})) = \mathrm{Bias}(f(\nu))^2 + \mathrm{Var}(f(\hat{\nu}))$$

shows that the error of the performance measure under the input model $\hat{\nu}$ includes the bias induced by input uncertainty. Bias is harder to quantify but there are recent studies that tackle that in parametric (Morgan et al. 2019) and nonparametric (Vahdat and Shashaani 2023) settings.

As demonstrated above, the uncertainty can propagate to the reliability of decision-making processes by affecting not only the variance but also the bias of the estimated performance measures obtained from simulation (Barton et al. 2022; Lam 2016). Therefore, accurately representing the input uncertainty in the optimization models is essential for the validity of the simulation results.

Remark 1 When considering correcting for input-uncertainty bias in the outputs, a similar objectives as in the DRO may be sought, whereby one can ask how wrong could the model outputs be by "mis-thinking" or mis-specifying the model inputs; or what is the worst estimate of outputs we could have for a given decision alternative. From this standpoint, there are ties between the worst case performance estimate in DRO and the idea of debiasing model outputs by estimating and removing the input-uncertainty bias before comparing systems for optimization.

2.2 Ambiguity Set Construction Methods

The construction of an ambiguity set is a pivotal step in hedging against input uncertainty in decision making. With a parameter $\delta > 0$ that denotes the size of the distributional uncertainty, two predominant approaches to define an ambiguity set are the moment-based and the discrepancy-based approaches. In the moment-based approach, the ambiguity set incorporates distributions that satisfy specific moment conditions, for example,

$$\mathcal{P}_\delta = \left\{ \nu_1 : \left| \int_\Xi \xi \mathrm{d}\nu_0(\xi) - \int_\Xi \xi \mathrm{d}\nu_1(\xi) \right| \leq \delta \right\}.$$

On the other hand, the discrepancy-based approach includes distributions within a certain statistical distance from a nominal distribution ν_0. In this case, the ambiguity set $\mathcal{P}$ is defined as

$$\mathcal{P}_\delta = \{\nu_1 : d(\nu_0, \nu_1) \leq \delta\},$$

where $d(\nu_0, \nu_1)$ represents the distance between two distributions ν_0 and ν_1.

In this paper, we focus on a well-established metrics in DRO, namely the Wasserstein distance, due to its desirable properties over the ϕ-divergence for two main reasons (Peyré et al. 2019). First, the Wasserstein distance has been proven to effectively measure the space of probability measures with a finite p-th moment. Second, it can quantify the distance between two distributions whose supports may not overlap. ϕ-divergence, on the other hand, only considers distributions that are absolutely continuous with respect to the nominal distribution and hence fails when distributions do not overlap; see Figure 1 (adapted from (Solomon 2023)).

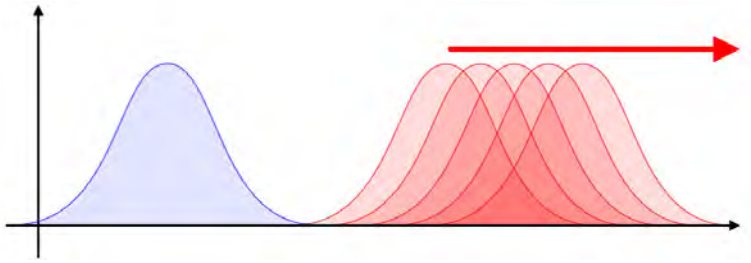

Figure 1: The ϕ-divergence between the nominal (blue) distribution and the alternative (red) distributions is zero since there is no overlap between them. In contrast, the Wasserstein distance becomes larger as the alternative distribution moves farther away from the nominal distribution.

Considering further the construction of ambiguity sets, Bertsimas and Van Parys (2022) explore the use of the bootstrap method as an alternative to other methods that are commonly used including ambiguity sets based on distance thresholds such as the Wasserstein distance (Kuhn et al. 2019). The authors use bootstrap sample in place of independent validation data, for robust optimization, to estimate the fraction of out-of-sample cost estimates (for a chosen decision) that exceed some threshold. In their setting, evaluation of the cost function can be done inexpensively. In robust simulation optimization, however, function evaluations are costly and the existence of a proxy (or metamodel) can be helpful so that not all bootstrap samples need be simulated.

The optimal transport (OT) methods address the problem of transporting a source (or nominal) probability measure $\nu_1 \in \mathcal{P}(\Xi_1)$ to a target probability measure $\nu_2 \in \mathcal{P}(\Xi_2)$, with the goal of minimizing the transportation cost c. Here, $\mathcal{P}(\Xi)$ represents the collection of all probability measures defined on the sigma-algebra of the sample space Ξ. A measure $\pi \in \Pi(\nu_1, \nu_2) \subset \mathcal{P}(\Xi_1 \times \Xi_2)$ is known as a transport plan and represents the amount of mass transferred from ν_1 to ν_2, where $\Pi(\nu_1, \nu_2)$ is the set of all transport plans between ν_1 and ν_2. By the Kantorovich formulation, OT problem can be expressed as

$$\mathrm{OT}(\nu_1, \nu_2; c) := \min_{\pi \in \Pi(\nu_1, \nu_2)} \iint_{\Xi_1 \times \Xi_2} c(\xi_1, \xi_2) \mathrm{d}\pi(\xi_2, \xi_2),$$

where $c(\xi_1, \xi_2)$ represents the cost of moving mass $\mathrm{d}\pi(\xi_1, \xi_2)$. When the cost $c(\xi_1, \xi_2)$ is defined as the p-th power of the distance metric, the OT problem is defined as the Wasserstein distance of order p, commonly referred to as the p-Wasserstein distance, that is,

$$d_{W_p}(\nu_1, \nu_2) = \min_{\pi \in \Pi(\nu_1, \nu_2)} \left(\iint_{\Xi_1 \times \Xi_2} (d(\xi_1, \xi_2))^p \, \mathrm{d}\pi(\xi_1, \xi_2) \right)^{1/p}.$$

In one dimension, the p-Wasserstein distance has a closed form given by

$$d_{W_p}(\nu_1, \nu_2) = \left(\int_0^1 \left| \mathcal{C}_{\nu 1}^{-1}(\xi) - \mathcal{C}_{\nu 2}^{-1}(\xi) \right|^p \, \mathrm{d}\xi \right)^{1/p},$$

where $\mathcal{C}_\alpha(\xi) := \int_{-\infty}^{\xi} \mathrm{d}\alpha$ is the cumulative distribution function (CDF) that maps $\mathbb{R}$ to $[0, 1]$ for a measure α. It is proven that (Santambrogio 2015) when $p = 1$, the p-Wasserstein distance simplifies to

$$d_{W_1}(\nu_1, \nu_2) = \int_0^1 \left| \mathcal{C}_{\nu 1}^{-1}(\eta) - \mathcal{C}_{\nu 2}^{-1}(\eta) \right| \, \mathrm{d}\eta = \int_{\mathbb{R}} \left| \mathcal{C}_{\nu 1}(\xi) - \mathcal{C}_{\nu 2}(\xi) \right| \, \mathrm{d}\xi. \tag{3}$$

In Figure 2 (adapted from (Rohde, Gustavo K. and Li, Shiying and Kolouri, Soheil 2022)), the 1-D case of d_{W_1} is illustrated.

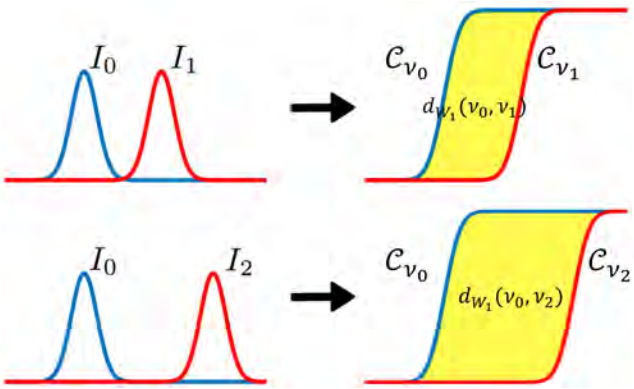

Figure 2: The 1-Wasserstein distances $d_{W_1}(\nu_0, \nu_1)$ and $d_{W_1}(\nu_0, \nu_2)$ in 1-D computes the areas between the CDFs $\mathcal{C}_{\nu_0}$ and $\mathcal{C}_{\nu_1}$, and $\mathcal{C}_{\nu_0}$ and $\mathcal{C}_{\nu_2}$, respectively. As the distance between the distributions increases, so does d_{W_1}.

In this paper, although unconventional, we utilize the difference in Kingman's formula values, henceforth referred to as the 'Kingman distance' and denote by d_K, between the distributions in an ambiguity set and the nominal distribution as a metric of distance between two probability distributions. This metric is applied to input models to simulate a queueing system that relies on moment information. In the following section, we will justify this choice by discussing the properties of Kingman's formula in relation to the performance measure, specifically focusing on the expected waiting time.

2.3 An Alternative Method: Kingman Distance

Kingman (1961) published the following approximation of waiting time for a $G/G/1$ queue with the input parametric model $\nu = (\rho, c_a^2, c_s^2, \mu)$:

$$\mathbb{E}_\nu[F(\nu)] \approx \left(\frac{\rho}{1-\rho} \right) \left(\frac{c_a^2 + c_s^2}{2} \right) \left(\frac{1}{\mu} \right), \tag{4}$$

where ν contains the interarrival times A and service times S, i.e., $\xi = (A, S)$ that follow parametric probability distributions parameterized by the utilization ρ, the average service rate μ, and the coefficients of variation for interarrival and service time, c_a and c_s, respectively. The approximation is precise when the utilization is high, and exact when interarrival times follow an exponential distribution, that is, the $M/G/1$ case. A Kingman moment-based distance metric for distributions $\nu_1 = (\rho_1, c_{a1}^2, c_{s1}^2, \mu_1)$ and

$\nu_2 = (\rho_2, c_{a2}^2, c_{s2}^2, \mu_2)$ based on (4) is

$$d_K(\nu_1, \nu_2) = \left| \left(\frac{\rho_1}{1-\rho_1} \right) \left(\frac{c_{a1}^2 + c_{s1}^2}{2} \right) \left(\frac{1}{\mu_1} \right) - \left(\frac{\rho_2}{1-\rho_2} \right) \left(\frac{c_{a2}^2 + c_{s2}^2}{2} \right) \left(\frac{1}{\mu_2} \right) \right|. \tag{5}$$

The Kingman moment-based distance metric (5) will provide approximate correspondence to performance distances for the $G/G/1$ case and exact correspondence for the $M/G/1$ case. We will see that this relationship does not hold, even approximately, for the Wasserstein distance.

3 SIMULATION ON EXPECTED WAITING TIME FOR A $G/G/1$ QUEUE

Recall from (5) that the Kingman distance d_K is positively related to the utilization ρ and the coefficient of variation c_a^2. However, Figure 3 suggests that this relationship does not apply to the Wasserstein distance. This is because the distance between the CDFs does not monotonically increase as the distance between the nominal parameters and alternatives increases. Specifically, the area between the alternative and nominal distributions first decreases until they match and then it increases. However, the waiting time consistently increases as the parameters increase.

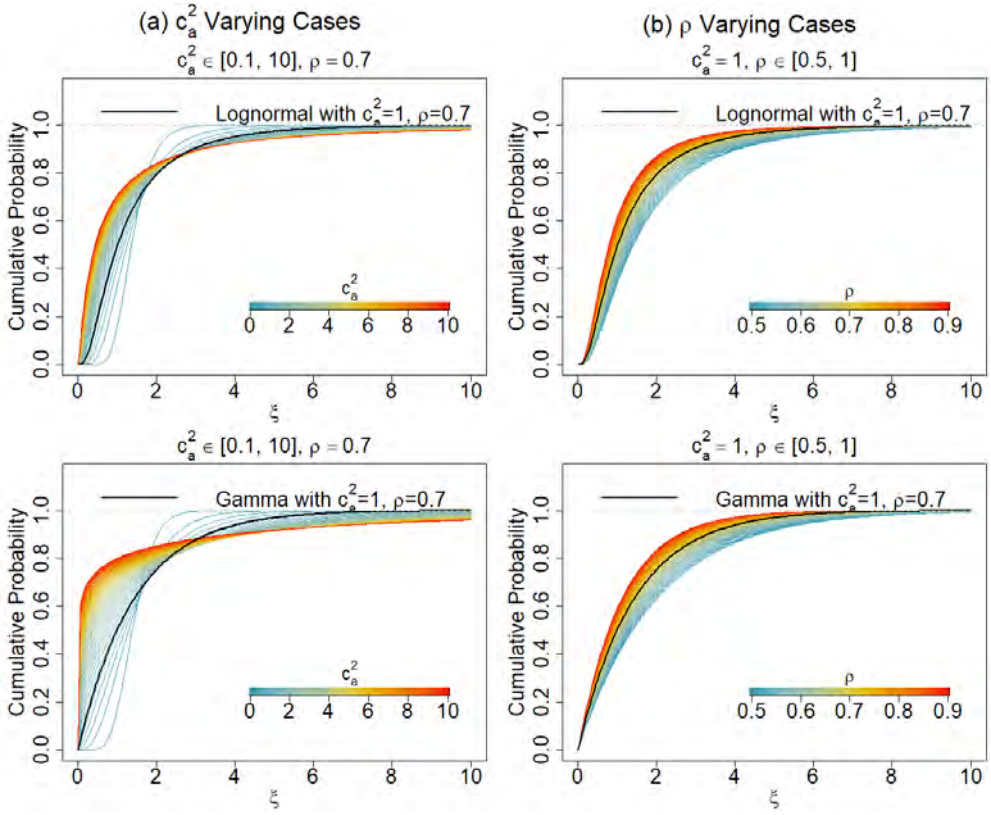

Figure 3: Alternative CDFs (colored) of Lognormal and Gamma distributions compared with the nominal CDF reveal the effect of changing each parameter on the area between CDFs. The left column shows that increasing c_a^2 initially reduces the area between CDFs, but increases it once the CDFs cross. The right column shows the same behavior as ρ increases.

In the following sections, we investigate how these different relationships can affect the identification of the worst-case scenario.

3.1 Simulation with Full-Factorial Experiment

To demonstrate how two different distance metrics perform in detecting the worst-case distribution, we performed a full-factorial experiment with a $G/G/1$ queue. In this experiment, the interarrival times follow a gamma distribution, while the service times follow a lognormal distribution. We assume that uncertainty in the system arises from the unknown distribution of interarrival times. Without loss of generality, we fix both the mean service time and the squared coefficient of variation c_s^2 at 1 with the shape parameter $\alpha = 1/c_a^2$ and the rate parameter $\lambda = \alpha\rho$ (Abate et al. 1993). We take the full factorial design of $c_a^2 \in \{0.1, 0.5, 1, 2, 5, 10\}$ and $\rho \in \{0.7, 0.8, 0.9\}$. Given the sensitivity of the queuing system to utilization and arrival time variation, we vary c_a^2 and ρ rather than shape and rate parameters for tractable insights.

As shown in Figure 4, changes in the parameters of the interarrival distribution impact the Wasserstein distance and the simulated waiting time differently. For instance, in both the Wasserstein distance of order 1 and 2, increasing the coefficient of variation c_a^2 from 10 to 20, while keeping the utilization rate ρ constant at 0.7, leads to increases in both the Wasserstein distance and the waiting time. Conversely, reducing ρ from 0.7 to 0.6, while maintaining c_a^2 at 10, also increases the Wasserstein distance but results in a decrease in waiting time. The same result is observed when the service times follow a gamma distribution.

These outcomes illustrate that similar changes in the Wasserstein distance can relate to divergent effects on waiting times. Such lack of one-to-one correspondance between the performance measure and Wasserstein distance underscores the necessity of searching the entire ambiguity set that is constructed via Wasserstein distance to identify the worst-case scenario. In contrast, it is generally sufficient to pinpoint the distribution that yields the maximum (positive) Kingman distance from the nominal distribution as that corresponding to the worst-case scenario.

3.2 Simulation with Bootstrap Ambiguity Sets

Given that input uncertainty stems from the finiteness of samples used to fit input distributions, bootstrapping can be an effective way to define ambiguity sets. Suppose we have a dataset of m interarrival times from an unknown distribution ν_A, whose empirical CDF we denote as $\hat{\nu}_A$. By applying bootstrapping to this data, we can generate b sample sets with empirical distributions $\hat{\nu}_A^{*j},\ j = 1, 2, \cdots, b$ in the neighborhood of the nominal probability distribution $\hat{\nu}_A$. This process introduces perturbations to constitute the ambiguity set as $\mathcal{P} = \bigcup_{j=1}^{b} \hat{\nu}_A^{*j}$. For each bootstrapped sample set, the Kingman distance can be computed using its fitted parameters following (4). To compute the Wasserstein distance of order 1 following (3), use the empirical CDFs for the original and bootstrapped sample sets as

$$\mathcal{C}_{\hat{\nu}_A}(a) = \sum_{i=1}^{m} \mathbf{1}\{A_i \le a\}, \quad \text{and} \quad \mathcal{C}_{\nu_A^{\star j}}(a) = \sum_{i=1}^{m} \mathbf{1}\{A_i^{\star j} \le a\},\ \forall\ j = 1, 2, \cdots, b;$$

where A_i and $A_i^\star$ represent the i-th data point in their respective sample sets..

With bootstrapped distributions forming the ambiguity set as opponsed to the full-factorial experiment in the previous section, we again compare the effectiveness of the Wasserstein and Kingman distances in identifying worst-case scenarios. The overall experimental design, depicted in Figure 5, is as follows. In each experiment replication, we generates $b = 100$ bootstrap sample sets. For each sample set, the Wasserstein and Kingman distances are calculated, and the sample sets producing the maximum distances, denoted ν_W^* and ν_K^* are selected. The waiting times for these input distributions are then simulated for a run-length of 10,000 times units to compute the estimated worst-case performance. Note, in some experiment replications the same set distribution may be identified by both metrics.

In this experiment, the interarrival and service time distributions are assumed to follow gamma and lognormal distributions, respectively. We consider multiple systems and examine whether the true most robust system can be identified with either metric. Unlike the initial experiment, the waiting time is simulated for a $G/G/1/k$ queue with capacity $k \in \{100, 500, 1000\}$. This variable capacity is introduced

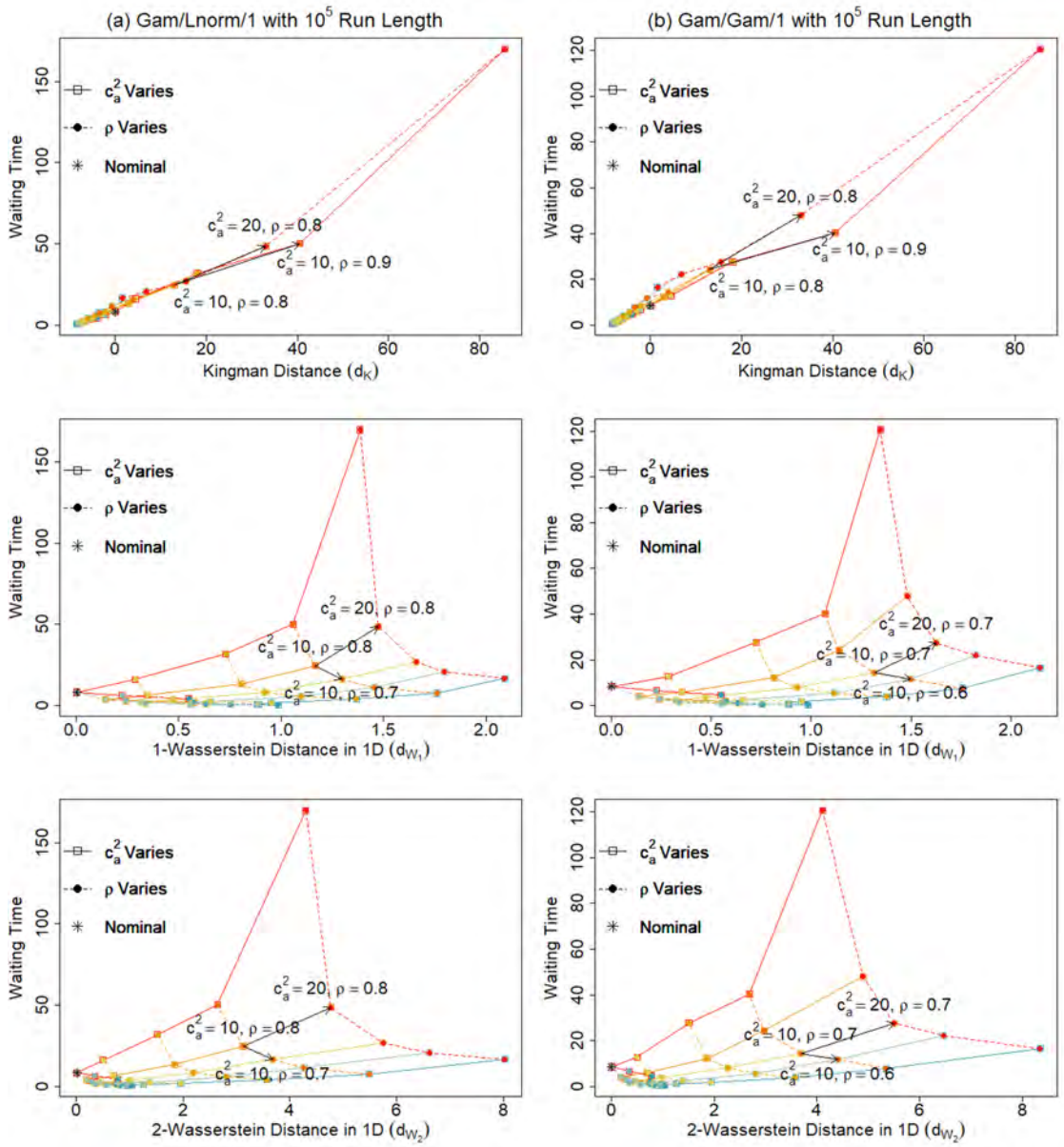

Figure 4: Simulated waiting times for a $G/G/1$ queue with Gamma distributed interarrival times and: (a) lognormal service times, and (b) Gamma service times. Kingman distance exhibits an almost monotonic relationship with the simulated waiting times, whereas the Wasserstein distance patterns are inconsistent.

to account for scenarios where the mean of the bootstrapped samples might be less than 1, which could indicate a potential for queue explosion when $\rho = \mathbb{E}[S]/\mathbb{E}[A] = 1/\mathbb{E}[A] > 1$.

The results of the experiment are depicted in Figures 6 and 7. Dotted lines connect instances that the worst case scenarios match. The 'True Mean' is the average waiting time with the true input distribution. First, the Wasserstein distance often fails to identify the worst-case scenario. In comparison, the Kingman distance more accurately identifies the true longest waiting times across all replications. We also observe that in most cases, the quartiles derived from ν_W^* are lower than those from ν_K^*. The performance difference between the two metrics is further supported by the results in Table 1 that concerns different queue capacities; from the table we note that in most cases, the worst-case distributions do not match.

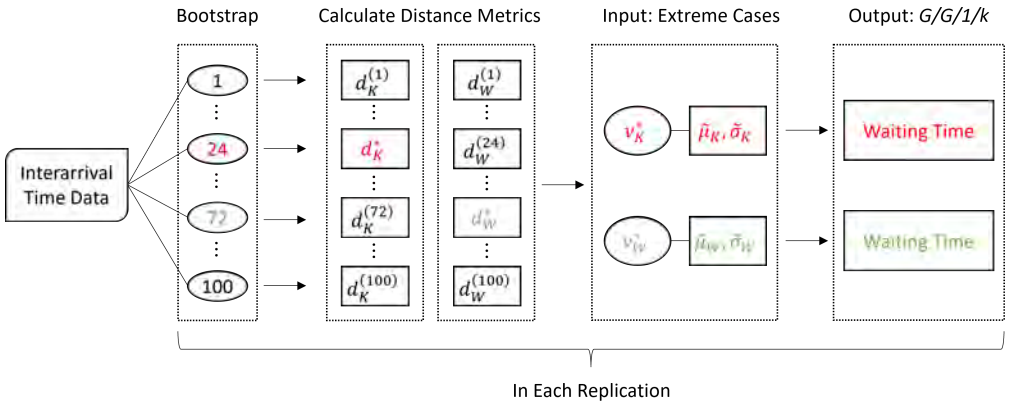

Figure 5: The bootstrapping procedure of estimating the worst average waiting time using Kingman and Wasserstein distance.

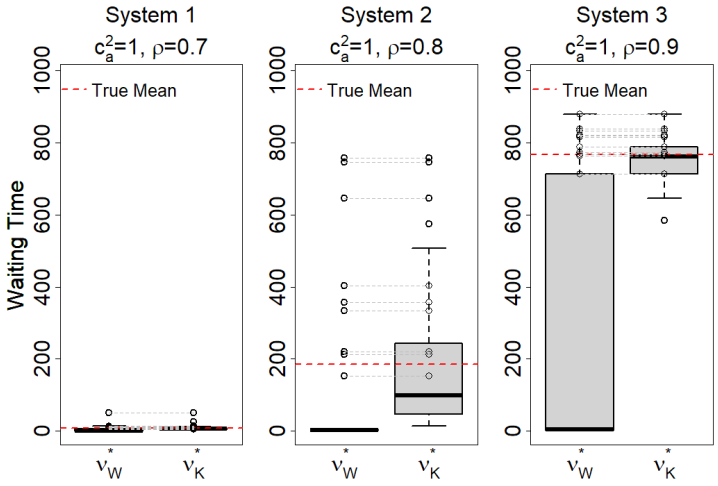

Figure 6: Simulated waiting times under worst-case distributions from 50 replications of a capacity-1000 queue with interarrival parameters $c_a^2 = 1$ and $\rho \in \{0.7, 0.8, 0.9\}$. The waiting times calculated from the distributions selected by the Kingman distance align more closely with the 'True Mean'.

Table 1: The proportion of the cases where ν_W^* matches ν_K^* out of 50 experiment replications for different queue capacities and system parameters.

Capacity k	c_a^2 (with fixed $\rho = 0.9$)					ρ (with fixed $c_a^2 = 1$)		
	0.1	0.5	1	5	10	0.7	0.8	0.9
100	0.46	0.26	0.22	0.02	0.02	0.20	0.44	0.20
500	0.44	0.34	0.16	0.04	0.00	0.20	0.18	0.18
1000	0.34	0.22	0.20	0.00	0.08	0.32	0.28	0.38

Furthermore, Figure 7 suggests that relying on the Wasserstein distance for decision-making could lead to incorrect conclusions. In Figures 6 and 7, the true mean indicates System 1 as the best system among the others. However, with the Wasserstein distance, it is hard to distinguish System 1 from System 2 in Figure 6, and from System 2, 3, and 5 in Figure 7. This underscores the Kingman distance as a potentially

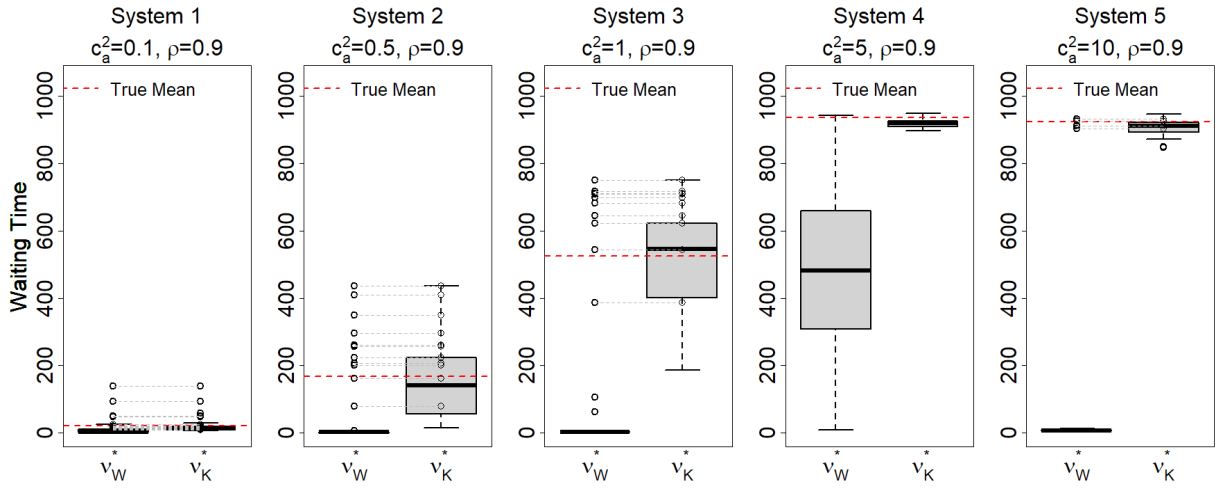

Figure 7: Simulated waiting times under the same experimental settings with interarrival parameters $c_a^2 \in \{0.1, 0.5, 1, 5, 10\}$ and a constant utilization rate of $\rho = 0.9$. The waiting times calculated from the distributions selected by the Kingman distance align more closely with the 'True Mean'.

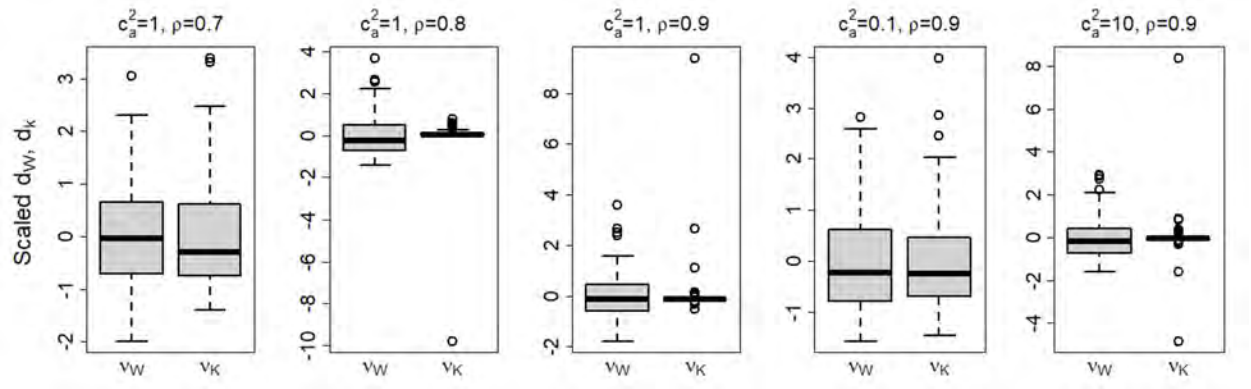

Figure 8: Box plots of the scaled distance obtained from 100 bootstrap sample sets. The queuing capacity is set to 1,000. The spreads in the Kingman distance are smaller than in the Wasserstein distance.

more reliable choice for the distance metric in robust optimization of queueing systems, where accurately identifying the worst-case scenario is crucial.

There is an additional benefit to using Kingman distance instead of the Wasserstein distance for constructing ambiguity set. Given that bootstrapping sets the size of the ambiguity set adaptively, it is advantageous for a distance metric to be more robust against the inherent variability of resampling. As shown in Figure 8, the spread of the scaled difference in the Kingman distance is generally smaller than in the Wasserstein distance, and at worst, they are almost the same. This suggests that the size of ambiguity is more sensitive if we use the Wasserstein compared to the Kingman distance.

4 CONCLUSION

This study has compared the Wasserstein and Kingman distance metrics in the context of DRO for $G/G/1$ queuing systems under input uncertainty. Our findings confirm that while the Wasserstein distance provides a comprehensive method for spanning ambiguity sets, it often requires extensive computational resources to pinpoint the worst-case scenarios. Conversely, the Kingman distance, anchored in moment-based metrics

derived from Kingman's formula, presents a more efficient alternative, closely correlating with actual system performance and facilitating quicker identification of critical scenarios. These insights underscore the potential for integrating the Kingman distance into DRO frameworks for significant computational relief, particularly for queuing systems where rapid and reliable decision-making is paramount.

Future research could explore several intriguing questions. For example, the integration of different distance metrics, specifically Wasserstein and Kingman, in the analysis of queuing systems warrants further investigation. This approach is particularly compelling given the robust mathematical foundation and beneficial properties of the Wasserstein distance. While the Wasserstein distance does not pinpoint the most extreme cases effectively, it provides a justifiable characterization of possible alternative distributions defining ambiguity sets. In certain applications, it is proven that the Wasserstein distance is superior to ϕ-divergence (Gao and Kleywegt 2023). We can think of leveraging the Kingman distance to identify the worst-case scenarios within the ambiguity set formed using the Wasserstein distance. Computational studies could validate and refine this integrated approach, potentially leading to more effective and efficient DRO strategies.

This investigation can be extended to more complex queuing systems and associated moment-based expected waiting time approximations. Additionally, for production systems, the clearing function (CF) approximates the non-linear relationship between expected output and expected workload. Its determination includes input uncertainty due to its reliance on simulation data for fitting CFs (Gopalswamy and Uzsoy 2019). Developing distributionally robust methods for fitting CFs and incorporating these functions into non-linear production planning models should be explored.

ACKNOWLEDGMENTS

This work was partially supported by National Science Foundation grant CMMI-2226347. We would like to thank the anonymous reviewers for their helpful corrections and suggestions.

REFERENCES

Abate, J., G. L. Choudhury, and W. Whitt. 1993. "Calculation of the GI/G/1 Waiting-Time Distribution and Its Cumulants from Pollaczek's Formulas". *Archiv für Elektronik und Ubertragungstechnik* 47(5/6):311–321.

Barton, R. R., H. Lam, and E. Song. 2022. "Input Uncertainty in Stochastic Simulation". In *The Palgrave Handbook of Operations Research*, 573–620. Springer.

Bayraksan, G. and D. K. Love. 2015. "Data-Driven Stochastic Programming Using Phi-Divergences". In *The Operations Research Revolution*, 1–19. INFORMS.

Bertsimas, D. and B. Van Parys. 2022. "Bootstrap Robust Prescriptive Analytics". *Mathematical Programming* 195(1):39–78.

Blanchet, J., K. Murthy, and F. Zhang. 2022. "Optimal Transport-Based Distributionally Robust Optimization: Structural Properties and Iterative Schemes". *Mathematics of Operations Research* 47(2):1500–1529.

Gao, R. and A. Kleywegt. 2023. "Distributionally Robust Stochastic Optimization with Wasserstein Distance". *Mathematics of Operations Research* 48(2):603–655.

Gopalswamy, K. and R. Uzsoy. 2019. "A Data-Driven Iterative Refinement Approach for Estimating Clearing Functions from Simulation Models of Production Systems". *International Journal of Production Research* 57(19):6013–6030.

Kingman, J. F. C. 1961, October. "The Single Server Queue in Heavy Traffic". *Mathematical Proceedings of the Cambridge Philosophical Society* 57(4):902–904.

Kuhn, D., P. M. Esfahani, V. A. Nguyen, and S. Shafieezadeh-Abadeh. 2019, October. "Wasserstein Distributionally Robust Optimization: Theory and Applications in Machine Learning". In *Operations Research & Management Science in the Age of Analytics*, edited by S. Netessine, D. Shier, and H. J. Greenberg, 130–166. INFORMS.

Lam, H. 2016. "Advanced Tutorial: Input Uncertainty and Robust Analysis in Stochastic Simulation". In *2016 Winter Simulation Conference (WSC)*, 178–192. IEEE https://doi.org/10.1109/WSC.2016.7822088.

Morgan, L. E., B. L. Nelson, A. C. Titman, and D. J. Worthington. 2019. "Detecting Bias due to Input Modelling in Computer Simulation". *European Journal of Operational Research* 279(3):869–881.

Panaretos, V. M. and Y. Zemel. 2019. "Statistical Aspects of Wasserstein Distances". *Annual Review of Statistics and Its Application* 6(1):405–431.

Peyré, G., M. Cuturi, *et al*. 2019. "Computational Optimal Transport: With Applications to Data Science". *Foundations and Trends® in Machine Learning* 11(5-6):355–607.

Rahimian, H. and S. Mehrotra. 2019. "Distributionally Robust Optimization: A Review". *arXiv preprint arXiv:1908.05659*.
Rohde, Gustavo K. and Li, Shiying and Kolouri, Soheil 2022. "Optimal Transport: A Crash Course". https://imagedatascience.com/transport/OTCrashCourse_22.pdf. Accessed 6th May 2024.
Santambrogio, F. 2015. "Optimal Transport for Applied Mathematicians". *Birkäuser, NY* 55(58-63):94.
Solomon, J. 2023. "Optimal Transport". https://groups.csail.mit.edu/gdpgroup/assets/68410_spring_2023/13_optimal_transport.pdf. Accessed 1st April 2024.
Vahdat, K. and S. Shashaani. 2023. "Robust Prediction Error Estimation with Monte-Carlo Methodology". *arXiv preprint arXiv:2207.13612*.

AUTHOR BIOGRAPHIES

HYUNG KHEE EUN is a second year Ph.D. student in the Edward P. Fitts Department of Industrial and System Engineering with research interests in simulation and stochastic optimization. His research includes feature selection, big data, and data-driven modeling. His email address is heun@ncsu.edu.

SARA SHASHAANI is an Assistant Professor and Bowman Faculty Scholar in the Edward P. Fitts Department of Industrial and System Engineering at North Carolina State University. Her research interests are simulation optimization and probabilistic data-driven models. She is a co-creator of SimOpt library. Her email address is sshasha2@ncsu.edu and her homepage is https://shashaani.wordpress.ncsu.edu/.

RUSSELL R. BARTON is Distinguished Professor of Supply Chain and Information Systems in the Smeal College of Business and Professor of Industrial Engineering at the Pennsylvania State University. His research interests include applications of statistical and simulation methods to system design and to product design, manufacturing and delivery. His email address is rbarton@psu.edu and his homepage is https://sites.psu.edu/russellbarton/.