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 IN DISTRIBUTIONALLY ROBUST OPTIMIZATION FOR 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-based approach and the discrepancy-based approach. 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 utilize the moment-based Kingman distance, an approximation of mean waiting time in $G/G/1$ queues, to determine the ambiguity set's size. Despite the lack of a strong preference for moment-based metrics, 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 distibutionally 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 $\mathbf{x} \in \mathcal{X} \subset \mathbb{R}^d$ and a random input data $\boldsymbol{\xi} \in \Xi$ that follows probability distribution ν be denoted by $F(\mathbf{x}, \boldsymbol{\xi})$ and its expectation denoted by

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

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 $f(\mathbf{x}_1, \nu) < f(\mathbf{x}_2, \nu)$ 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 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 (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 realization of $\boldsymbol{\xi}$, 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. Robust optimization (RO), on the other hand, does not rely on any specific distributional information, considering only $\boldsymbol{\xi}$ 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 $\boldsymbol{\xi}$, DRO reduces to SO. In addition, if $\mathcal{P}$ contains all probability distributions on the support of $\boldsymbol{\xi}$, the DRO reduces to RO.

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

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

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

DRO finds a decision that minimizes the worst-case of the objective function among all probability measures in the ambiguity set. Therefore, 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 distributions.

However, generally the true distributions are 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 $\mathbf{x}$ for simplicity, i.e., expectation of the stochastic simulation outputs F generated from random inputs $\boldsymbol{\xi}$. 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 $F_j(\hat{\nu}) := F(\boldsymbol{\xi}_j \sim \hat{\nu})$ is the jth identically distributed simulation replication under input model $\hat{\nu}$. Assuming that the simulation replications from one to r are independent, the decomposition shows

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

Thus, we have

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

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. In addition, the mean square error

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

shows that the error of the performance measure under the input model $f(\hat{\nu})$ can include 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 2022) 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 speaking of input uncertainty bias, we often seek same objectives as in the DRO, whereby we ask how wrong could our model outputs be by "mis-thinking" or mis-specifying our model inputs. We argue that there are ties between the idea of debiasing model outputs by estimating and removing the input uncertainty bias before comparing systems for optimization and the worst case performance estimate in DRO.

2.2 Ambiguity Set Construction Methods

The construction of an ambiguity set is a pivotal step in hedging against input uncertainty in decision making. There are two predominant approaches: the moment-based and the discrepancy-based approaches. In the moment-based approach, the ambiguity set incorporates distributions that satisfy specific moment conditions. 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, and δ is the size of the distributional uncertainty.

In this paper, we focus on the Wasserstein distance due to its desirable properties over the ϕ-divergence (Peyré et al. 2019). First, the Wasserstein distance is a well-established metric in DRO because it has been proven to effectively measure the space of probability measures with a finite pth moment. In addition, even though the supports of the comparison distributions do not have an overlapped support, it can quantify the distance between them. However, ϕ-divergence method only considers distributions that are absolutely continuous with respect to the nominal distribution. As a result, ϕ-divergence fails when distributions do not overlap.

Considering further the construction of ambiguity sets, Bertsimas and Van Parys (2022) explore the use of the bootstrap to determine a distribution ambiguity set in robust optimization. They propose this as an alternative to other methods that are commonly used, and identify many alternatives, 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 to estimate the fraction of out-of-sample cost estimates (for a chosen decision) exceed some threshold. In their setting, evaluation of the cost function can be done inexpensively. In robust simulation optimization function evaluations are costly, and a proxy can be helpful so that not all bootstrap samples need be simulated.

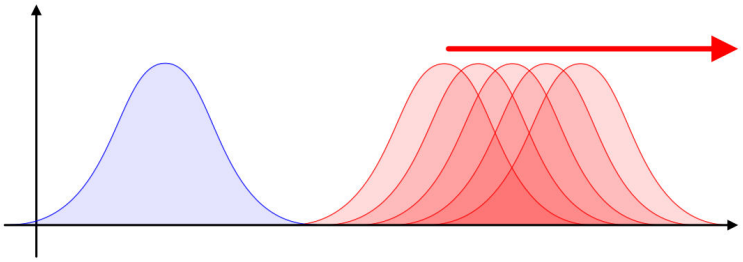

Figure 1: The ϕ-divergence between the blue distribution and all the red distributions is zero since there is no overlap between them. In contrast, the Wasserstein distance becomes larger as the red distribution move away from the blue distribution. This figure has been adopted from Solomon (2023).

The OT methods address the problem of transporting a source (or nominal) probability measure $\nu_1 \in \mathcal{P}(X)$ to a target probability measure $\nu_2 \in \mathcal{P}(Y)$, with the goal of minimizing the transportation cost c. Here, $\mathcal{P}(Z)$ represents the collection of all probability measures defined on the sigma-algebra of Z. A measure $\pi \in \Pi(\nu_1, \nu_2) \subset \mathcal{P}(X \times Y)$ 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 transport plans between ν_1 and ν_2. Now, by the Kantorovich formulation, OT problem can be expressed as follows:

$$\mathrm{OT}(\nu_1, \nu_2; c) := \min_{\pi \in \Pi(\nu_1, \nu_2)} \iint_{X \times Y} c(x, y) d\pi(x, y).$$

where $c(x, y)$ represents the cost of moving mass $d\pi(x, y)$. When the cost $c(x, y)$ is defined as the pth power of the shortest path distance $\|x - y\|^p$, OT problem is defined as the Wasserstein distance of order p (p-Wasserstein distance) such that

$$d_W^p(\nu_1, \nu_2) = \min_{\pi \in \Pi(\nu_1, \nu_2)} \left(\iint_{X \times X} \|x - y\|^p d\pi(x, y) \right)^{1/p}.$$

As a special case, 1-Wasserstein distance in one dimension (W_1 in 1-D) has a simple closed form such that

$$d_W^1(\nu_1, \nu_2) = \int_{\mathbb{R}} |\mathcal{C}_{\nu_1}(x) - \mathcal{C}_{\nu_2}(x)| \, dx, \tag{3}$$

where $\mathcal{C}_\alpha(x) := \int_{-\infty}^{x} d\alpha$ is the cumulative distribution function (CDF) that maps $\mathbb{R} \to [0,1]$ for a measure α. Throughout this paper, d_W is used to denote W_1 in 1-D and is illustrated in Figure 2.

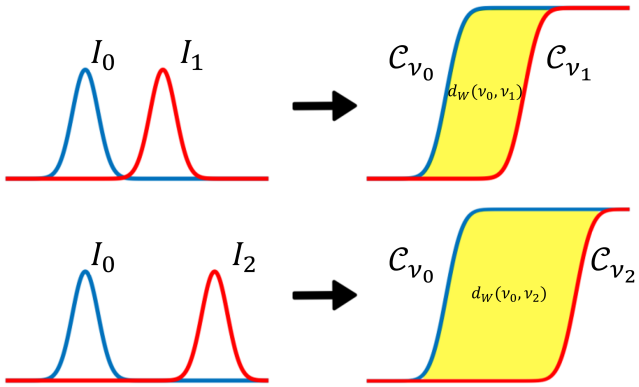

Figure 2: The computation of the optimal transport between two 1-D measures using CDFs. The Wasserstein distance $d_W(\nu_0, \nu_1)$ and $d_W(\nu_0, \nu_2)$ represent the areas between the CDFs $\mathcal{C}_{\nu_0}$ and $\mathcal{C}_{\nu_1}$, and $\mathcal{C}_{\nu_0}$ and $\mathcal{C}_{\nu_2}$, respectively. As the distance between the distributions increases, so does the Wasserstein distance d_W. This figure has been adapted from Rohde, Li, and Kolouri (2022).

In this paper, although unconventional, we utilize the difference in Kingman's formula values, henceforth referred to as the 'Kingman distance (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 input parametric models $\nu = (\rho, c_a^2, c_s^2, \mu)$:

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

where ν represents the input model characterizing the interarrival times A and servive times S, i.e., $\boldsymbol{\xi} = (A, S)$ that follow paramteric probability distributions parameterized by the utilization ρ, and the coefficients of variation for interarrival time and service time, c_a and c_s, respectively. Here, μ is the average service rate. The approximation is precise when the utilization is high, and exact when interarrival times follow an exponential distribution: 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:

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

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 Wasserstein distance.

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

As discussed in Section 2, Kingman's formula provides the exact waiting time for the $M/G/1$ queue, and an approximation for the $G/G/1$ queue. However, given our interest in the worst-case scenarios of the queueing system under high utilization such as $\rho = 0.9$, and considering (3), Figure 3 suggests that the largest Wasserstein distance does not necessarily correspond to the worst-case distribution that produces the most extreme expected waiting time. In other words, if the Kingman distance exhibits a monotonic relationship with the simulated waiting time, while the Wasserstein distance does not, then it can be concluded that it is easier to identify the worst-case scenario using the Kingman distance rather than the Wasserstein distance.

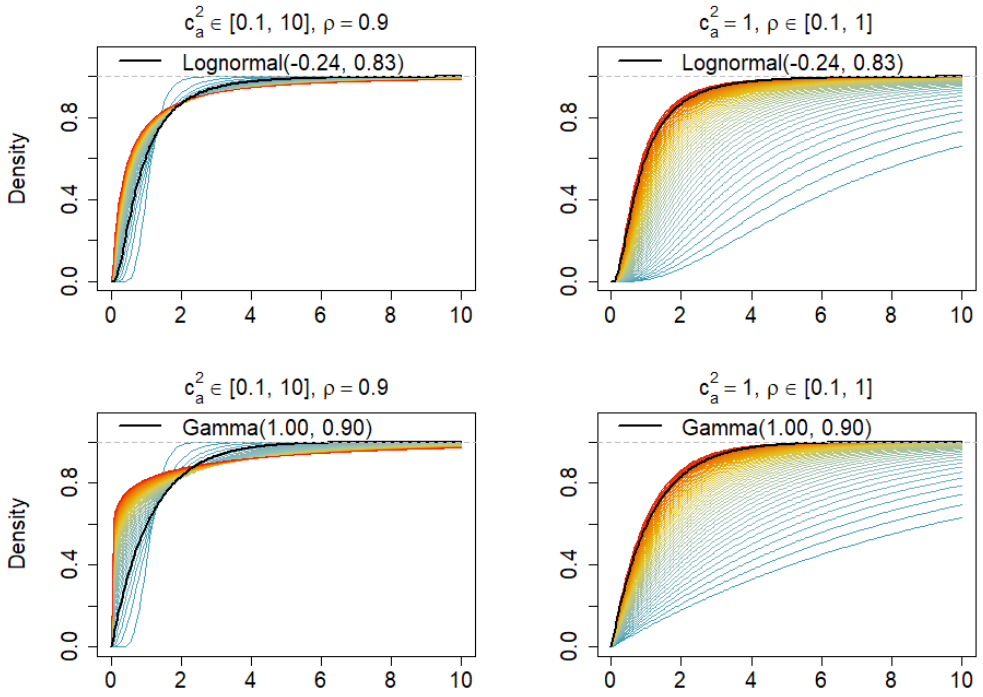

Figure 3: CDFs of Lognormal and Gamma distributions with various parameters. The plots in the left column demonstrate that increasing c_a^2 initially causes the area between the reference distributions and the nominal distribution to decrease quickly. However, this area begins to increase once the CDF crosses the nominal distribution. In contrast, the plots in the right column show that the gap decreases as ρ increases, with the most of the reference distributions converging towards the nominal distribution.

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 log-normal distribution. We assume that uncertainty in the system arises solely 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 (Abate et al. 1993). The shape parameter α is $1/c_a^2$ and the rate parameter λ is $\alpha\rho$; thus, we vary c_a^2 and ρ in our experiment so that $c_a^2 \in \{0.1, 0.5, 1, 2, 5, 10\}$ and $\rho \in \{0.7, 0.8, 0.9\}$, and take the full factorial of two parameters. Given the

sensitivity of the queueing system to utilization and arrival time variation, this approach, varying c_a^2 and ρ rather than shape and rate parameters, should offer more tractable insights.

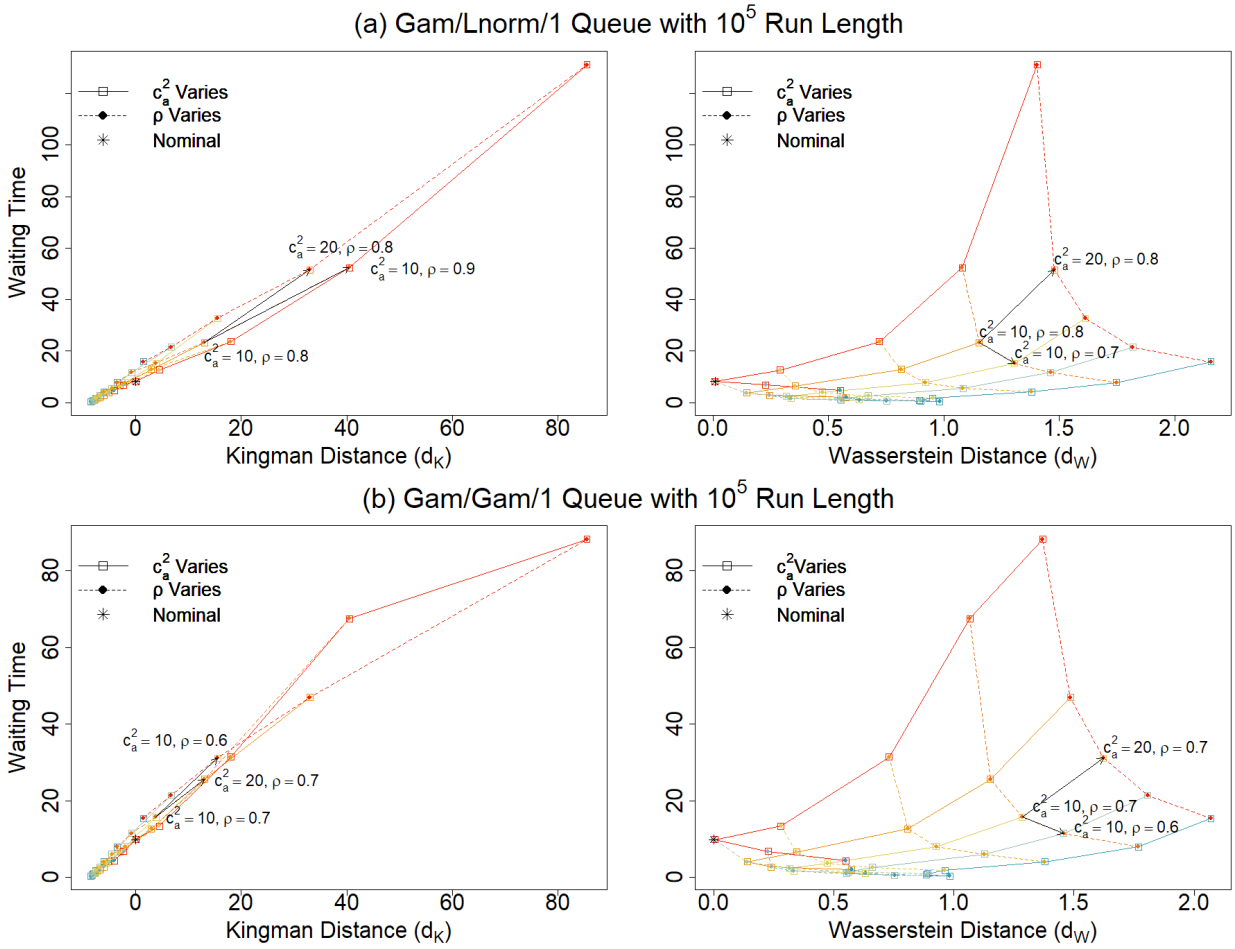

Figure 4: Simulated waiting times for a $G/G/1$ queue under two conditions: (a) arrival distribution is a Gamma distribution with service times following a lognormal distribution, and (b) arrival distribution is a Gamma distribution with service times following a Gamma distribution. The Kingman distance exhibits an almost monotonic relationship with the simulated waiting times, whereas the Wasserstein distance does not exhibit a consistent pattern.

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, 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 0.8 to 0.7, while maintaining c_a^2 at 10, also increases the Wasserstein distance but results in a decrease in waiting time. These outcomes illustrate that similar changes in the Wasserstein distance can have divergent effects on waiting times, depending on the specific adjustments made to the parameters. Such variability underscores the necessity of searching the entire ambiguity set based on the Wasserstein distance to identify the worst-case scenario. In contrast, with the Kingman distance, it is generally sufficient to pinpoint the distribution that yields the maximum (positive) Kingman distance from the nominal distribution. This distinction highlights the differing robustness and sensitivity of these metrics to parameter changes within queuing systems.

3.2 Simulation with Bootstrap Ambiguity Sets

One setting for robust optimization is when the distribution uncertainty comes from the finiteness of samples used to fit the distributions. In this case bootstrapping can be an important way to define ambiguity sets (Bertsimas and Van Parys 2022). Suppose we have a dataset for the interarrival time from an unknown distribution. By applying bootstrapping to this data, we can generate multiple sample sets considered to be drawn from distributions in the neighborhood of the nominal probability distribution, introducing some perturbations to reflect input uncertainty. For each bootstrapped sample set, the Kingman distance can be computed using its mean and variance, as described in (4). For the Wasserstein distance, we define the empirical cumulative distribution functions (CDFs) for both the original and a bootstrapped sample set as follows:

$$\mathcal{C}_{\nu_1}(x) = \sum_{i=1}^{m} \mathbf{1}\{a_i \leq x\}, \quad \mathcal{C}_{\nu_2}(x) = \sum_{i=1}^{m} \mathbf{1}\{b_i \leq x\}$$

where a_i and b_i represent the ith data point in their respective samples, and m is the sample size. The Wasserstein distance of order 1 in one dimension is computed as described in (3).

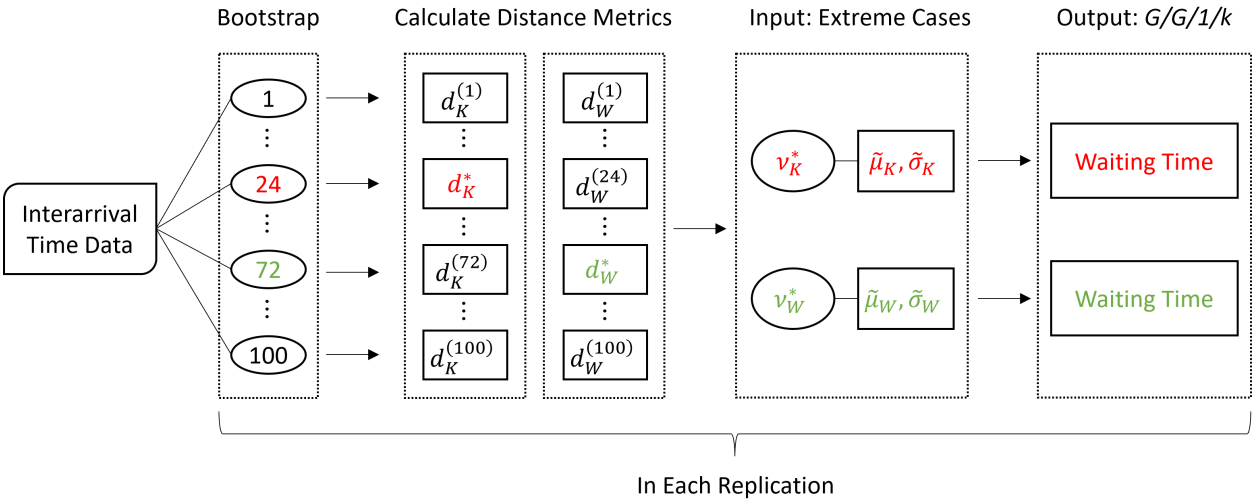

Figure 5: In each simulation replication, the experiment generates 100 bootstrap sample sets. For each sample set, the Wasserstein and Kingman distances are calculated. Sample sets producing the maximum distances—such as the 24th set for the Kingman and the 72nd for the Wasserstein—are selected. Parameters of the interarrival distribution are estimated from the mean and standard deviation of these sets, and waiting times are simulated accordingly. Note that the same set could be identified by both distance metrics.

In this section, we revisit the implications of a full-factorial experiment comparing the effectiveness of the Wasserstein and Kingman distances in identifying worst-case scenarios. Additionally, we assess the impact of using the Wasserstein distance over the Kingman distance for system performance evaluation.

Figure 5 illustrates the overall experimental design. Here, both the interarrival and service time distributions are assumed to follow gamma and lognormal distributions, respectively. Unlike in the initial experiment, the waiting time is simulated for a $G/G/1/k$ queue with varying capacities $k \in \{100, 500, 1000\}$. This variable capacity is introduced as a factor because the mean of the bootstrapped samples might be less than 1, indicating a potential for queue explosion where $\rho = \mathbb{E}[S]/\mathbb{E}[A] = 1/\mathbb{E}[A] > 1$.

The results of the experiment are depicted in Figures 6 and 7. We observe that in most cases, the quartiles derived from ν_W^* are lower than those from ν_K^*. This suggests that, given that the Kingman distance has almost monotonic and positive relationship with the waiting time, the Wasserstein distance fails often in identifying the worst-case scenario. In comparison, the Kingman distance more accurately

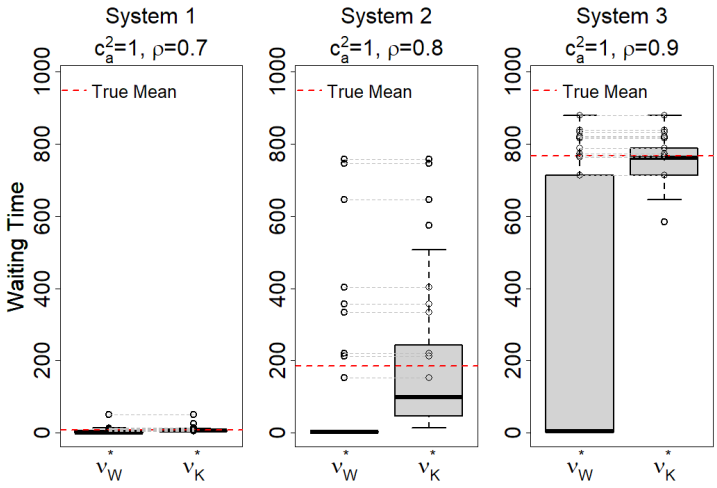

Figure 6: The graph shows simulated waiting times under worst-case distributions from 50 replications, each lasting 10,000 time units. The results are for a queue capacity of 1,000, with interarrival times exhibiting $c_a^2 = 1$ and $\rho \in \{0.7, 0.8, 0.9\}$. Dotted lines connect pairs of points where the same bootstrap sample set was chosen by both Wasserstein and Kingman metrics. The 'True Mean' indicates the average longest waiting time across 50 replications. The waiting times calculated from the distributions selected by the Kingman distance align more closely with the 'True Mean'.

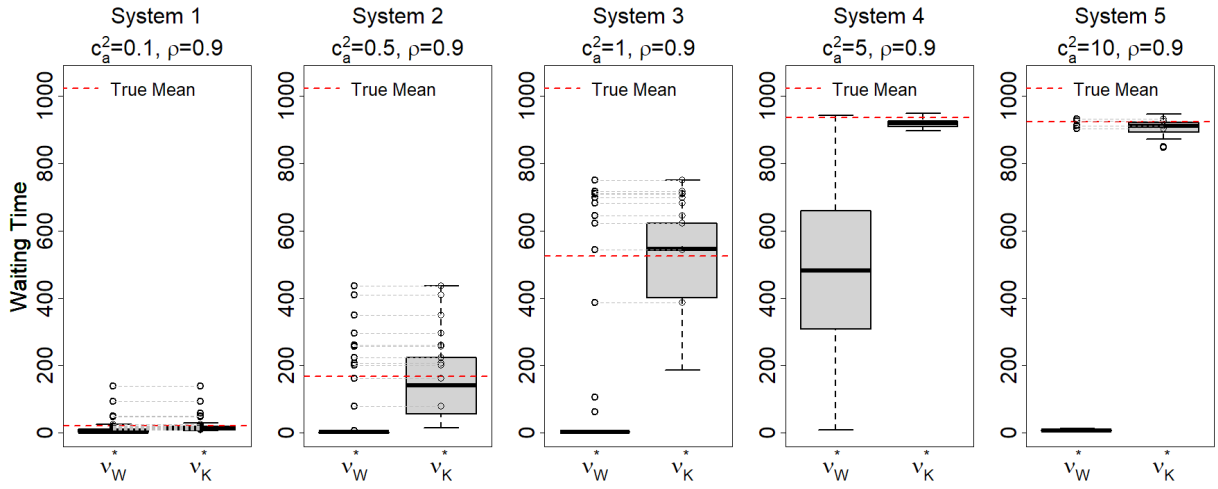

Figure 7: Simulated waiting times under the same experimental settings, where the original interarrival time distributions vary, characterized by $c_a^2 \in \{0.1, 0.5, 1, 5, 10\}$ and a constant utilization rate of $\rho = 0.9$. Again, in this figure, the waiting times calculated from the distributions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

identifies worst cases when compared to the mean of the true longest waiting times across all replications. The performance difference between the two metrics is further supported by the results in Table 1.

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

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

Capacity k	c_a^2 ($\rho = 0.9$)					ρ ($c_a^2 = 1$)		
	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

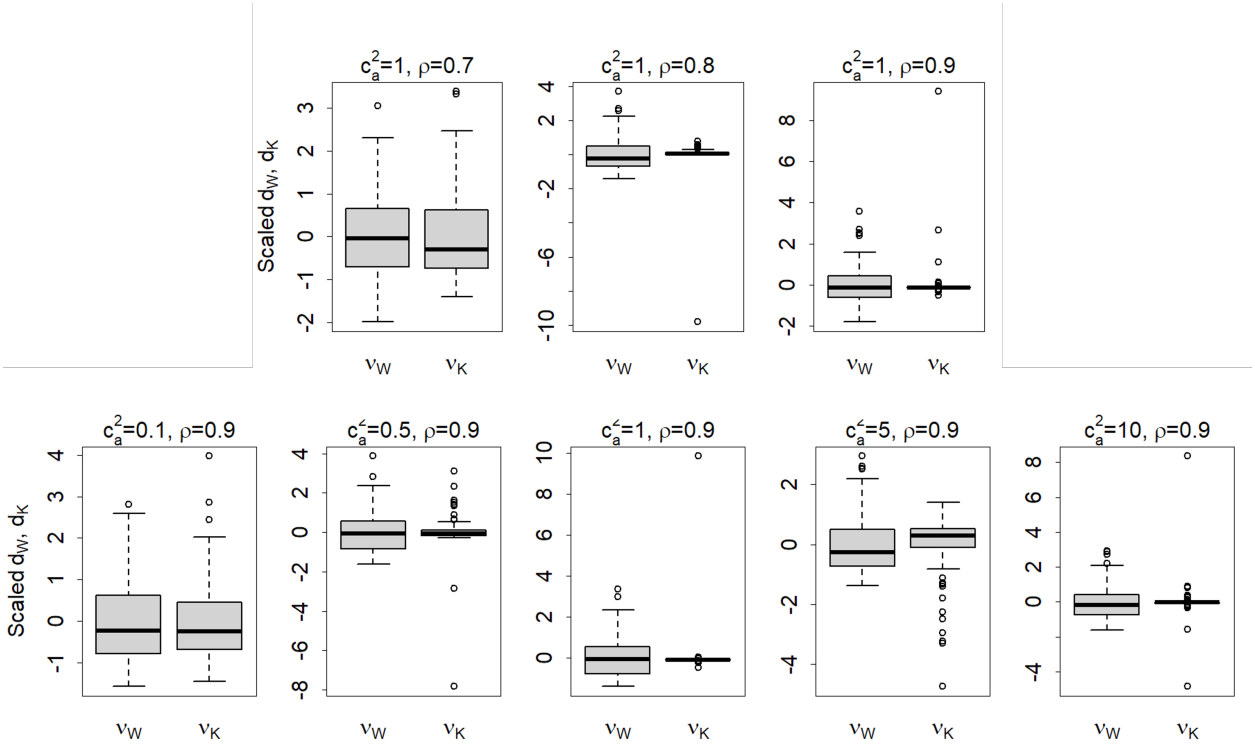

Figure 8: Box plots of the scaled distance obtained from 100 bootstrap sample sets. The queuing capacity is set to 1,000. The spreads in the Kingman distance are smaller than in the Wasserstein distance.

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. 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, 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 superiot 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.

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.

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. 2022. "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/.