Math. Program., Ser. B
DOI 10.1007/s10107-012-0567-2

FULL LENGTH PAPER

Tractable stochastic analysis in high dimensions via robust optimization

Chaithanya Bandi · Dimitris Bertsimas

Received: 16 November 2011 / Accepted: 1 June 2012
© Springer and Mathematical Optimization Society 2012

Abstract Modern probability theory, whose foundation is based on the axioms set forth by Kolmogorov, is currently the major tool for performance analysis in stochastic systems. While it offers insights in understanding such systems, probability theory, in contrast to optimization, has not been developed with computational tractability as an objective when the dimension increases. Correspondingly, some of its major areas of application remain unsolved when the underlying systems become multidimensional: Queueing networks, auction design in multi-item, multi-bidder auctions, network information theory, pricing multi-dimensional options, among others. We propose a new approach to analyze stochastic systems based on robust optimization. The key idea is to replace the Kolmogorov axioms and the concept of random variables as primitives of probability theory, with uncertainty sets that are derived from some of the asymptotic implications of probability theory like the central limit theorem. In addition, we observe that several desired system properties such as incentive compatibility and individual rationality in auction design are naturally expressed in the language of robust optimization. In this way, the performance analysis questions become highly structured optimization problems (linear, semidefinite, mixed integer) for which there exist efficient, practical algorithms that are capable of solving problems in high dimensions. We demonstrate that the proposed approach achieves computationally tractable methods for (**a**) analyzing queueing networks, (**b**) designing multi-item, multi-bidder auctions with budget constraints, and (**c**) pricing multi-dimensional options.

C. Bandi · D. Bertsimas (✉)
Operations Research Center, Massachusetts Institute of Technology,
E40-147, Cambridge, MA 02139, USA
e-mail: dbertsim@mit.edu

C. Bandi
e-mail: cbandi@mit.edu

Published online: 17 June 2012

Keywords Stochastic analysis · Robust optimization · Queueing · Mechanism design · Option pricing

Mathematics Subject Classification 90-02

1 Introduction

Probability theory has a long and distinguished history that dates back to the beginning of the seventeenth century. Games involving randomness led to an exchange of letters between Pascal and Fermat in which the fundamental principles of probability theory were formulated for the first time. The Dutch scientist Huygens, learned of this correspondence and in 1657 published the first book on probability entitled *De Ratiociniis in Ludo Aleae*. In 1812 Laplace introduced a host of new ideas and mathematical techniques in his book *Theorie Analytique des Probabilities*. Laplace applied probabilistic ideas to many scientific and practical problems. The theory of errors, actuarial mathematics, and statistical mechanics are examples of some of the important applications of probability theory developed in the nineteenth century. Many researchers have contributed to the theory since Laplace's time; among the most important are Chebyshev, Markov, von Mises, and Kolmogorov.

One of the difficulties in developing a mathematical theory of probability has been the need to arrive at a definition of probability that is precise enough for use in mathematics, yet comprehensive enough to be applicable to a wide range of phenomena. The search for a widely acceptable definition took nearly three centuries. The matter was finally resolved in the 1933 monograph of Kolmogorov who outlined an axiomatic approach that forms the basis for the modern theory. With the publication in 1933 of Kolmogorov's book *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Kolmogorov laid the foundations of an abstract theory, designed to be used as a mathematical model for certain classes of observable events. The fundamental concept of the theory is the concept of a probability space (Ω, A, P), where Ω is a space of points ω which are denoted as elementary events, while A is a σ-algebra of sets in Ω, and P is a probability measure defined for all A-measurable events, i.e., for all sets S belonging to A. Kolmogorov's three axioms form the basis of this theory

1. $\mathbb{P}(S) \geq 0, \ \forall S \in A$.
2. $\mathbb{P}(\Omega) = 1$.
3. If $S_i \in A, \ i \geq 1$, are pairwise disjoint, then $\mathbb{P}(\cup_{i=1}^{\infty} S_i) = \sum_{i=1}^{\infty} \mathbb{P}(S_i)$.

Another important primitive of probability theory is the notion of a random variable as a quantity that takes values with certain probabilities. A key objective of probability theory is to estimate the probability distributions of a random variable Y, which is a function of n primitive random variables $X_1, \ldots, X_n$, that is $Y = f(X_1, \ldots, X_n)$, given information on the joint probability distribution of the primitive random variables $X_1, \ldots, X_n$. For example, suppose that we are given n independent, random variables X_i uniformly distributed in [0, 1], and we are interested in evaluating the distribution the random variable $Y = \sum_{i=1}^{n} X_i$. Specifically, we are interested in the quantity $P(Y \leq t)$, $0 \leq t \leq n$. Even for a modest value of $n = 10$, this is a complex calculation involving convolutions. Perhaps the simplest way to calculate the Laplace transform of Y, which is the product (because of independence) of the Laplace transforms of the X_i's and then numerically invert the transform. Note that in order to estimate a probability of a relative simple event, we need to invoke rather heavy machinery from complex analysis and inverse of transforms.

The situation we described is not an isolated instance. Consider a single class queueing network (see Sect. 3), that has been used in the latter part of the twentieth century to model computer and communication networks. Suppose we are interested in the expected value of the number of jobs waiting in one of the queues in the network. If the distribution of interarrival and service times is not exponential, we do not know how to calculate this expectation exactly, and two avenues available to make progress are *simulation* and *approximation*. Simulation can take a considerable amount of time in order for the results to be statistically significant, and in addition, if the simulation model is complex as it is often the case, then it is difficult to isolate and understand the key insights in the model.

On the other hand, approximation methods can potentially lead to results that are not very close to the true answers. Given these considerations, it is fair to say that after more than 50 years of research we really do not have a satisfactory answer to the problem of performance analysis of queueing networks. J.F.C. Kingman, one of the pioneers of queueing theory in the twentieth century in his opening lecture at the conference entitled "100 Years of Queueing–The Erlang Centennial" [45], writes, "If a queue has an arrival process which cannot be well modeled by a Poisson process or one of its near relatives, it is likely to be difficult to fit any simple model, still less to analyze it effectively. So why do we insist on regarding the arrival times as random variables, quantities about which we can make sensible probabilistic statements? Would it not be better to accept that the arrivals form an irregular sequence, and carry out our calculations without positing a joint probability distribution over which that sequence can be averaged?".

The situation in queueing networks we discussed above is present in other examples. Shannon [64] characterized the capacity region and designed optimal coding and decoding methods in single-sender, single-receiver channels, but the extension to multi-sender, multi receiver channels with interference is by and large open. Myerson [57], in his Nobel Prize winning work, solved the problem of optimal market design in single item auctions, but the extension to multi-item case with bidders that have budget constraints has remained open. Black and Scholes [21], in their Nobel Prize winning work, solved the problem of pricing an option in an underlying security, but the extension to multiple securities with market frictions has not been resolved. In all of these and other problems, we see that we can solve the underlying problem in low dimensional problems, but we have been unable to solve the underlying problems when the dimension increases.

In our opinion, the reason for this is related to the history of probability theory as a scientific field. The multi-century effort that led to the development of modern probability theory aimed to lay the conceptual and foundational basis of the field. The primitives of probability theory, the Kolmogorov axioms and the notion of a random variable, while powerful for modeling purposes, have not been developed with computational tractability as an objective when the dimension increases. In contrast, consider the development of optimization as a scientific field in the second part of the twentieth century. From its early years (Dantzig [31]), modern optimization has had as an objective to solve multi-dimensional problems efficiently from a practical point of view. The notion of efficiency used, however, is not the same as theoretical efficiency (polynomial time solvability) developed in the 1970s ([27,43]). The Simplex method, for instance, has proven over many decades to be practically efficient, but not theoretically efficient. It is exactly this notion of practical efficiency we use in this work: it is the ability to solve problems of realistic size relative to the application we address. For example, queueing networks with hundreds of nodes, auctions with hundreds of items and bidders with budget constraints, network information theory with hundreds of thousands of codewords, and option pricing problems with hundreds of securities.

Given the success of optimization to solve multi-dimensional problems, it is natural, in our opinion, to formulate probability problems as optimization problems. For this purpose, we utilize robust optimization, arguably one of the fastest growing areas of optimization in the last decade, to accomplish this. In this effort, we are guided by the words of Dantzig [33], who in the opening sentence of his book Linear Programming and Extensions writes “The final test of any theory is its capacity to solve the problems which originated it.” In this respect, we report in this paper our progress to date in the three areas that originated the need to develop the proposed theory:

- **(a)** Analyzing queueing networks in Sect. 3.
- **(b)** Designing multi-item, multi-bidder auctions with budget constraints in Sect. 4.
- **(c)** Pricing multi-dimensional options in Sect. 5.

The research program surveyed here aims to develop a tractable theory for analyzing stochastic systems in high dimensions via robust optimization. In addition to the applications covered here, we have applied this approach to:

- **(a)** *Network Information theory* (see [2]), where we present a robust optimization based framework to formulate and solve the central problem of characterizing the capacity region and constructing matching optimal codes for multi-user channels with interference. In particular, we solve the open problems of characterizing the capacity regions of the multi-user Gaussian interference channel, the multicast and the multi-access Gaussian channels and construct matching optimal codes by solving semidefinite optimization problems with rank one constraints.
- **(b)** *Transient Analysis of queueing networks* (see [7]), where we concentrate on the transient analysis of single class queues, and feed-forward networks, and derive closed form expressions for the transient behavior of such systems.
- **(c)** *Analysis of Multi-class queueing networks* (see [5]), under various scheduling policies (FCFS and priority based).

Given the broad spectrum of applications we cover, our aim is to introduce the key concepts and algorithms, and provide the main theoretical and empirical evidence that illustrates the strength of the method. We do not provide all the mathematical proofs but give reference to our specific papers that give more details. In the next section, we outline the building blocks of the proposed approach.

2 The building blocks of our approach

One of the major successes of probability theory is the development of limit laws. As an illustration consider the central limit theorem that asserts that if X_i, $i = 1, \ldots, n$ are independent, identically distributed random variables with mean μ and standard deviation σ, then as $n \to \infty$, the random variable $S_n = \sum_{i=1}^n X_i$ is asymptotically distributed as a standard normal, that is

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \le t\right) = \mathbb{P}\left(Z \le t\right),$$

where $Z \sim N(0, 1)$. The importance of the limit laws in the theory of probability can be emphasized by quoting Kolmogorov [39]: "All epistemological value of the theory of probability is based on this: that large scale random phenomena in their collective action create strict, non random regularity."

The key building block in our approach is that rather than assuming as primitives the axioms of probability theory (Kolmogorov axioms and the notion of a random variable), we assume as primitives the conclusions of probability theory, namely its limit laws. Let us give a motivating example. From the central limit theorem $(S_n - n\mu)/\sigma\sqrt{n}$ is asymptotically standard normal. We know that a standard normal Z satisfies

$$\mathbb{P}(|Z| \le 2) \approx 0.95,\ \mathbb{P}(|Z| \le 3) \approx 0.99.$$

We therefore assume that the quantities X_i take values such that

$$\left|\sum_{i=1}^n X_i - n\mu\right| \le \Gamma\sigma\sqrt{n},$$

where Γ is a small numerical constant 2 or 3 that is selected adaptively to make a good fit empirically. In other words, we do not describe the uncertain quantities X_i as random variables, rather they take values in an uncertainty set

$$\mathcal{U} = \left\{(X_1, \ldots, X_n) \,\middle|\, \left|\sum_{i=1}^n X_i - n\mu\right| \le \Gamma\sigma\sqrt{n}\right\}. \tag{1}$$

In specific situations we can augment the uncertainty set $\mathcal{U}$ by using additional asymptotic laws as we illustrate in Sect. 2.2.

2.1 The connection with optimization

Suppose we are interested in estimating/analyzing $E[f(X_1, \ldots, X_n)]$, where $(X_1, \ldots, X_n)$ are random variables. Using asymptotic laws of probability, we construct an uncertainty set $\mathcal{U}$. We have already seen an example in Eq. (1). Then, we estimate $E[f(X_1, \ldots, X_n)]$ by solving the constrained optimization problems

$$\max f(x_1, x_2, \ldots, x_n) \quad \min f(x_1, x_2, \ldots, x_n)$$
$$\text{s.t. } (x_1, x_2, \ldots, x_n) \in \mathcal{U}, \quad \text{s.t. } (x_1, x_2, \ldots, x_n) \in \mathcal{U},$$

In other words, we transform the performance analysis question to a constrained optimization problem, arguably a problem we can solve efficiently in high dimensions, and we use the asymptotic laws of probability theory, arguably the most insightful aspect of probability theory, to construct the constrained set in the optimization problem.

Suppose that we are interested in a design problem involving design parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$ and uncertain parameters $\mathbf{x} = (x_1, \ldots, x_n) \in \mathcal{U}$ and we are interested in solving the problem

$$\max_{\boldsymbol{\theta}} E[f(\mathbf{x}, \boldsymbol{\theta})]. \tag{2}$$

Problems like (2) have been studied in the literature of Stochastic Programming that started with the work of Dantzig [32] and continues to be strong to this date. For a more comprehensive review of Stochastic Programming, we refer to [20].

In the Robust Optimization (RO) paradigm, we model the uncertainty of the parameters $\mathbf{x}$ by the uncertainty set $\mathcal{U}$ where the parameters $\mathbf{x}$ take values and solve the optimization problem

$$\max_{\boldsymbol{\theta}} \min_{\mathbf{x} \in \mathcal{U}} f(\mathbf{x}, \boldsymbol{\theta}). \tag{3}$$

RO is one of the fastest growing areas of optimization in the last decade. It addresses the problem of optimization under uncertainty, in which the uncertainty model is not stochastic, but rather deterministic and set-based. RO models are typically tractable computationally. For example, [9,11,12,36], and [37], proposed linear optimization models with ellipsoidal uncertainty sets, whose robust counterparts correspond to conic quadratic optimization problems. [10,18,19] proposed RO models with polyhedral uncertainty sets that can model linear/integer variables, and whose robust counterparts correspond to linear/integer optimization models. For a more thorough review we refer the reader to [8], and [15].

Within the interface of Robust Optimization and Stochastic Programming, [8,58] (pp 27–60), [13], and [73] propose alternative tractable approaches to model chance constrained problems.

2.2 Constructing uncertainty sets

In this section, we outline the principles for constructing uncertainty sets we use in this paper.

Using historical data and the central limit theorem Suppose that we have estimated the mean μ and the standard deviation σ of i.i.d. random variables $(X_1, \ldots, X_n)$. We expect that the central limit theorem holds, and we model uncertainty by the uncertainty set given in Eq. (1).

Modeling correlation information Suppose that the random variables $\mathbf{X} = (X_1, \ldots, X_n)$ are correlated. Specifically, there are $m < n$ i.i.d. random variables $\mathbf{Y} = (Y_1, \ldots, Y_m)$ with mean μ_Y and standard deviation σ_Y such that $\mathbf{X} = \mathbb{A}\mathbf{Y} + \boldsymbol{\epsilon}$, where $\mathbb{A}$ is an $n \times m$ matrix and $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)$ of i.i.d. random variables that have mean zero and standard deviation σ_ϵ. Then the uncertainty set is given by

$$\mathcal{U}^{Corr} = \left\{ \mathbf{X} \,\middle|\, \mathbf{X} = \mathbb{A}\mathbf{Y} + \boldsymbol{\epsilon},\ \left| \sum_{i=1}^{m} Y_i - m\mu_Y \right| \le \Gamma \sigma_Y \sqrt{m},\ \left| \sum_{i=1}^{n} \epsilon_i \right| \le \Gamma \sigma_\epsilon \sqrt{n} \right\}. \tag{4}$$

Stable laws The central limit theorem belongs to a broad class of weak convergence theorems. These theorems express the fact that a sum of many independent random variables tend to be distributed according to one of a small set of stable distributions. When the variance of the variables is finite, the stable distribution is the normal distribution. In particular, these stable laws allow us to construct uncertainty sets for heavy-tailed distributions.

Theorem 1 (Nolan [60]) *Let $Y_1, Y_2, \ldots$ be a sequence of i.i.d. random variables, with mean μ and undefined variance. If $Y_i \sim Y$, where Y is a stable distribution with parameter $\alpha \in (1, 2]$ then*

$$\frac{\sum_{i=1}^{n} Y_i - n\mu}{n^{1/\alpha}} \sim Y.$$

Motivated by this result, we construct an uncertainty set $\mathcal{U}^{\mathrm{HT}}$ representing the random variables $\{Y_i\}$ as follows

$$\mathcal{U}^{\mathrm{HT}} = \left\{ (Y_1, Y_2, \ldots, Y_n) \,\middle|\, -\Gamma^{\mathrm{HT}} \le \frac{\sum_{i=1}^{n} Y_i - n\mu}{n^{1/\alpha}} \le \Gamma^{\mathrm{HT}} \right\}, \tag{5}$$

where Γ^{HT} can be chosen based on the distributional properties of the random variable Y. Note that $\mathcal{U}^{\mathrm{HT}}$ is again a polyhedron.

Incorporating distributional information using typical sets In this section, we illustrate how to construct uncertainty sets that utilize knowledge of the specific probability distribution. We use the idea of a typical set $\mathcal{U}^{\mathrm{Typical}}$, introduced by Shannon [64] in the context of his seminal work in information theory.

(a) $\mathbb{P}\left[\tilde{\mathbf{Z}} \in \mathcal{U}^{\mathrm{Typical}}\right] \to 1$, as $n \to \infty$.

(b) The conditional pdf $h(\tilde{\mathbf{Z}}) = f(\tilde{\mathbf{Z}} | \tilde{\mathbf{Z}} \in \mathcal{U}^{\mathrm{Typical}})$ satisfies:

$$\left| \frac{1}{n} \log h(\tilde{\mathbf{Z}}) + H_f \right| \le \epsilon_n,$$

for some H_f (the entropy of the distribution) and $\epsilon_n \to 0$, as $n \to \infty$.

Property (a) means that the typical set has probability nearly 1, while Property (b) means that all elements of the typical set are nearly equiprobable, see [28]. We next show that (Proposition 1), for a probability density $f(\cdot)$, the typical set is given by

$$\mathcal{U}_\epsilon^f = \left\{ \mathbf{Z} \middle| -\Gamma_\epsilon^f \leq \frac{\sum_{i=1}^n \log f(Z_i) + nH_f}{\sigma_f \sqrt{n}} \leq \Gamma_\epsilon^f \right\}, \tag{6}$$

where

$$H_f = -\int_{-\infty}^{\infty} f(x) \log f(x)\, \mathrm{d}x, \; \sigma_f^2 = \int_{-\infty}^{\infty} f(x) \left(\log f(x) + H_f\right)^2 \mathrm{d}x,$$

and Γ_ϵ^f is chosen such that

$$\mathbb{P}\left[\left| \sum_{i=1}^n \log f(Z_i) + nH_f \right| \leq \Gamma_\epsilon^f \cdot \sigma_f \sqrt{n} \right] = 1 - \epsilon. \tag{7}$$

Proposition 1 *For a distribution $f(\cdot)$, $\mathcal{U}_\epsilon^f$ defined in Eq. (6) satisfies*

(a) $\mathbb{P}\left[\tilde{\mathbf{Z}} \notin \mathcal{U}_\epsilon^f\right] \leq \epsilon$.

(b) *The conditional pdf $h(\tilde{\mathbf{Z}}) = f\left(\tilde{\mathbf{Z}} \middle| \tilde{\mathbf{Z}} \in \mathcal{U}_\epsilon^f\right)$ satisfies:*

$$\left| \frac{1}{n} \log h(\tilde{\mathbf{Z}}) + H_f \right| \leq \phi(\epsilon),$$

with $\phi(\epsilon) \to 0$, as $n \to \infty$.

Proof **(a)** We have $\mathbb{P}\left[\tilde{\mathbf{Z}} \notin \mathcal{U}_\epsilon^f\right] = 1 - \mathbb{P}\left[\tilde{\mathbf{Z}} \in \mathcal{U}_\epsilon^f\right]$ and by (7), we have

$$\mathbb{P}\left[\tilde{\mathbf{Z}} \in \mathcal{U}_\epsilon^f\right] = \mathbb{P}\left[\left| \sum_{i=1}^n \log f(Z_i) + nH_f \right| \leq \Gamma_\epsilon^f \cdot \sigma_f \sqrt{n} \right] = 1 - \epsilon.$$

Therefore, $\mathbb{P}\left[\tilde{\mathbf{Z}} \notin \mathcal{U}_\epsilon^f\right] \leq \epsilon$.

(b) Let $\tilde{\mathbf{Z}} \in \mathcal{U}_\epsilon^f$. Then,

$$h(\tilde{\mathbf{Z}}) = f(Z_1)\, f(Z_2) \ldots f(Z_n).$$

Therefore, since $\tilde{\mathbf{Z}} \in \mathcal{U}_\epsilon^f$, we have

$$\left|\frac{1}{n}\log h(\tilde{\mathbf{Z}}) + H_f\right| = \left|\frac{1}{n}\sum_{j=1}^{n}\log f(Z_j) + H_f\right| \leq \frac{\Gamma_\epsilon^f \cdot \sigma_f}{\sqrt{n}} \to 0, \quad \text{as } n \to \infty.$$

□

In order to obtain stronger intuition on the nature of the uncertainty sets, we specialize Proposition 1 for the cases of normal, exponential, uniform and binary distributions.

Corollary 1 [Typical Sets for Normal, Exponential, Uniform and Binary Distributions]

(a) *The typical set for normally distributed i.i.d. random variables* $\tilde{Z}_i \sim N(0, \sigma)$ *is given by*

$$\mathcal{U}_\epsilon^G = \left\{\mathbf{Z}\;\middle|\; -\Gamma_\epsilon^G \leq \|\mathbf{Z}\|^2 - n\sigma^2 \leq \Gamma_\epsilon^G\right\}, \tag{8}$$

(b) *The typical set for correlated normally distributed random variables* $\tilde{\mathbf{Z}} \sim N(\mathbf{0}, \Sigma)$ *is given by*

$$\mathcal{U}_\epsilon^{CG} = \left\{\mathbf{Z}\;\middle|\; -\Gamma_\epsilon^{CG} \leq \|\Sigma^{-\mathbf{1}}\mathbf{Z}\|^2 - n \leq \Gamma_\epsilon^{CG}\right\}, \tag{9}$$

(c) *The typical set for exponentially distributed random variables* $\tilde{Z}_i \sim Exp(\lambda)$ *is given by*

$$\mathcal{U}_\epsilon^E = \left\{\mathbf{Z}\;\middle|\; \frac{n}{\lambda} - \frac{\sqrt{n}}{\lambda}\cdot\Gamma_\epsilon^E \leq \sum_{j=1}^{n} Z_j \leq \frac{n}{\lambda} + \frac{\sqrt{n}}{\lambda}\cdot\Gamma_\epsilon^E,\ \mathbf{Z} \geq \mathbf{0}\right\}, \tag{10}$$

(d) *The typical set for uniformly distributed random variables* $\tilde{Z}_i \sim U[a, b]$ *is given by*

$$\mathcal{U}_\epsilon^U = \left\{\mathbf{Z}\;\middle|\; \begin{array}{c} n\dfrac{a+b}{2} - \Gamma_\epsilon^U\sqrt{n} \leq \displaystyle\sum_{j=1}^{n} Z_j \leq n\dfrac{a+b}{2} + \Gamma_\epsilon^U\sqrt{n}, \\ a \leq Z_j \leq b,\ j = 1, \ldots, n, \end{array}\right\}, \tag{11}$$

(e) *The typical set for binary random variables* $\tilde{Z}_i \sim Bin(p)$ *is given by*

$$\mathcal{U}_\epsilon^B = \left\{\mathbf{Z}\;\middle|\; \begin{array}{c} np - \Gamma_\epsilon^B\sqrt{n} \leq \displaystyle\sum_{j=1}^{n} Z_j \leq np + \Gamma_\epsilon^B\sqrt{n}, \\ Z_j \in \{0, 1\},\ j = 1, \ldots, n, \end{array}\right\}, \tag{12}$$

where $\varGamma_\epsilon^G,\ \varGamma_\epsilon^{CG},\ \varGamma_\epsilon^E,\ \varGamma_\epsilon^U,\ \varGamma_\epsilon^B$ *are chosen such that*

$$\mathbb{P}\left[\mathcal{U}_\epsilon^G\right] = \mathbb{P}\left[\mathcal{U}_\epsilon^{CG}\right] = \mathbb{P}\left[\mathcal{U}_\epsilon^E\right] = \mathbb{P}\left[\mathcal{U}_\epsilon^U\right] = \mathbb{P}\left[\mathcal{U}_\epsilon^B\right] = 1 - \epsilon. \tag{13}$$

3 Performance analysis of queueing networks

The origin of queueing theory dates back to the beginning of the twentieth century, when Erlang [38] published his fundamental paper on congestion in telephone traffic. In addition to formulating and solving several practical problems arising in telephony, Erlang laid the foundations for queueing theory in terms of the nature of assumptions and techniques of analysis that are being used to this day. In the second part of the twentieth century, a very substantial literature of queueing theory was developed modeling queueing primitives as renewal processes.

From the time of Erlang, the Poisson process has played a very significant role in modeling the arrival process of a queue. When combined with exponentially distributed service times, the resulting $M/M/m$ queue with m servers is tractable to analyze in steady-sate. While exponentiality leads to a tractable theory, assuming general distributions, on the other hand, yields considerable difficulty with respect to performing a near-exact analysis of the system. The $GI/GI/m$ queue with independent and generally distributed arrivals and services is by and large intractable. Currently, there does not exist a method that is capable of producing accurate numerical answers, let alone closed form expressions, for arbitrary distributions.

The situation becomes even more challenging if one considers analyzing the performance of queueing networks. A key result that allows generalizations to networks of queues is Burke's theorem (Burke [24]) which states that the departure process from an $M/M/m$ queue is Poisson. This property allows one to analyze queueing networks and leads to product form solutions as in Jackson [41]. However, when the queueing system is not $M/M/m$, the departure process is no longer a renewal process, i.e., the interdeparture times are dependent. With the departure process lacking the renewal property, the state-of-the-art theory provides no means to determine performance measures exactly, even for a simple network with queues in tandem. The two avenues in such cases are *simulation* and *approximation*. Simulation can take a considerable amount of time in order for the results to be statistically significant. In addition, simulation models are often complex, which makes it difficult to isolate and understand key qualitative insights. On the other hand, approximation methods can potentially lead to results that are not very close to the true answers.

Given these challenges, it is fair to say that the key problem of performance analysis of queueing networks has remained open under the probabilistic framework. Our objective in this section, is to analyze queueing networks by replacing the primitives of stochastic processes with uncertainty sets. In this section, we summarize the results of Bandi et al. [6] for single class queueing networks. Extensions to multiclass queueing networks are contained in [5].

3.1 An alternate model of a queue

We introduce the notion of a *robust queue* where we model the arrival and service processes by uncertainty sets instead of assigning probability distributions. We denote the interarrival time between the $(i-1)$st and ith customers by T_i and the service time of customer i by X_i. We propose the following uncertainty sets on the interarrival and service processes.

Assumption 1 We make the following assumptions

(a) The interarrival times belong to the uncertainty set

$$\mathcal{U}^a = \left\{ (T_1, T_2, \ldots, T_n) \left| \frac{\left| \sum_{i=k+1}^{n} T_i - \frac{(n-k)}{\lambda} \right|}{(n-k)^{1/\alpha_a}} \leq \Gamma_a, \ \forall k \leq k_0 \right. \right\},$$

where $1/\lambda$ is the expected interarrival time, Γ_a is a parameter that captures variability information and $1 < \alpha_a \leq 2$ models possibly heavy-tailed probability distributions.

(b) The service times for an m-server belong to the uncertainty set

$$\mathcal{U}_m^s = \left\{ (X_{i \cdot m+r})_{i=1}^{v} \left| \frac{\left| \sum_{i=k+1}^{v} X_{im+r} - \frac{(v-k)}{\mu} \right|}{(v-k)^{1/\alpha_s}} \leq \Gamma_s, \ \forall k \leq k_0 \right. \right\},$$

where $0 \leq r < m$, $1/\mu$ is the expected service time, Γ_s is a parameter that captures variability information and $1 < \alpha_s \leq 2$ models possibly heavy-tailed probability distributions.

For the case of a single server queue, that is, when $m = 1$, the uncertainty set is given by

$$\mathcal{U}^s = \left\{ (X_1, X_2, \ldots, X_n) \left| \frac{\left| \sum_{i=k+1}^{n} X_i - \frac{(n-k)}{\mu} \right|}{(n-k)^{1/\alpha_s}} \leq \Gamma_s, \ \forall k \leq k_0 \right. \right\}.$$

The value of k_0 is chosen so that the central limit theorem is valid for the variables $X_1, X_2, \ldots, X_{k_0}$. A typical value would be $k_0 = n - 30$. We note that the case of independent and identically distributed interarrival and service times corresponds to $\alpha = 2$.

3.2 Waiting time in a robust queue

Consider a single server queue in which the interarrival and service times belong to the sets $\mathcal{U}^a$ and $\mathcal{U}^s$, respectively. In this section, we assume that $\alpha_a = \alpha_s = \alpha$. Let

W_i, $i \geq 1$ be the waiting time of the ith customer in such a queue. The waiting times are linked by the recursion (Lindley [52])

$$W_i = \max\left(W_{i-1} + X_{i-1} - T_i, 0\right) = \max_{1\leq k\leq i}\left(\sum_{j=k}^{i-1} X_j - \sum_{j=k+1}^{i} T_j, 0\right). \qquad (14)$$

In our framework, the worst case waiting time W_n of the nth customer can be obtained by solving the optimization problem

$$\max_{\mathbf{T}\in\mathcal{U}^a, \mathbf{X}\in\mathcal{U}^s} \max_{1\leq k\leq n}\left(\sum_{i=k}^{n-1} X_i - \sum_{j=k+1}^{n} T_j, 0\right).$$

This problem allows a closed form solution.

Theorem 2 *Under Assumption 1, the worst case waiting time* W_n

(a) *in a single server queue with traffic intensity* $\rho = \lambda/\mu < 1$, *is given by:*

$$W_n \leq \frac{\alpha - 1}{\alpha^{\alpha/(\alpha-1)}} \cdot \frac{\lambda^{1/(\alpha-1)}\left(\Gamma_T + \Gamma_X\right)^{\alpha/(\alpha-1)}}{(1-\rho)^{1/(\alpha-1)}}. \qquad (15)$$

(b) *in a multi-server queue with* m *servers and traffic intensity* $\rho = \lambda/m\mu < 1$, *is given by:*

$$W_n \leq \frac{\alpha - 1}{\alpha^{\alpha/(\alpha-1)}} \cdot \frac{\lambda^{1/(\alpha-1)}\left(\Gamma_T + \Gamma_X/m^{1/\alpha}\right)^{\alpha/(\alpha-1)}}{(1-\rho)^{1/(\alpha-1)}}. \qquad (16)$$

Proof **(a)** The waiting time of the nth customer can be expressed recursively in terms of the interarrival and service times and using Eq. (14), can be written as

$$\begin{aligned}\overline{W}_n &= \max_{\mathbf{X}\in\mathcal{U}^s, \mathbf{T}\in\mathcal{U}^a} \max_{1\leq j\leq n}\left(\sum_{\ell=j}^{n-1} X_\ell - \sum_{\ell=j+1}^{n} T_\ell, 0\right)\\ &= \max_{1\leq j\leq n} \max_{\mathbf{X}\in\mathcal{U}^s, \mathbf{T}\in\mathcal{U}^a}\left(\sum_{\ell=j}^{n-1} X_\ell - \sum_{\ell=j+1}^{n} T_\ell, 0\right).\end{aligned} \qquad (17)$$

From Assumption 1, for any $j \leq k_0$, we know that the sums of the service times and interarrival times are bounded by

$$\sum_{\ell=j}^{n-1} X_\ell \leq \frac{n-j}{\mu} + \Gamma_s(n-j)^{1/\alpha}, \quad \sum_{\ell=j+1}^{n} T_\ell \geq \frac{n-j}{\lambda} - \Gamma_a(n-j)^{1/\alpha}. \qquad (18)$$

Combining Eqs. (17) and (18), we obtain an one-dimensional concave maximization problem (since $1 < \alpha \leq 2$)

$$\max_{1 \leq j \leq n} \left\{ (\Gamma_a + \Gamma_s)(n-j)^{1/\alpha} - \frac{1-\rho}{\lambda}(n-j) \right\}. \tag{19}$$

Making the transformation $x = n - j$, Eq. (19) becomes

$$\max_{1 \leq x \leq n} \beta \cdot x^{1/\alpha} - \gamma \cdot x = \frac{\alpha - 1}{\alpha^{\alpha/(\alpha-1)}} \frac{\beta^{\alpha/(\alpha-1)}}{\gamma^{1/(\alpha-1)}}, \tag{20}$$

with $\beta = \Gamma_a + \Gamma_s$ and $\gamma = (1-\rho)/\lambda > 0$, given $\rho < 1$. Note that Eq. (20) is maximized at

$$n - j^* = x^* = \left(\frac{\beta}{\alpha\gamma}\right)^{\alpha/(\alpha-1)} = \left(\frac{\lambda(\Gamma_a + \Gamma_s)}{\alpha(1-\rho)}\right)^{\alpha/(\alpha-1)}. \tag{21}$$

As $\rho \to 1$, we have

$$j^* = n - x^* = n - \left(\frac{\lambda(\Gamma_a + \Gamma_s)}{\alpha(1-\rho)}\right)^{\alpha/(\alpha-1)} \leq n - 30 = k_0,$$

which implies that Eq. (18) is valid for j^*. Substituting β and γ by their respective expressions in Eq. (20) yields Eq. (15) after some straightforward algebraic manipulations.

(b) The proof is very similar, and we omit it.

□

Implications and insights

(a) **Qualitative insights:** The robust queue behaves qualitatively the same as the traditional queue. For instance, the classical i.i.d. arrival and service processes with finite variance can be modeled by setting $\alpha = 2$. For the single server queue, Eq. (16) becomes

$$\overline{W}_n \leq \frac{\lambda}{4} \cdot \frac{(\Gamma_a + \Gamma_s)^2}{(1-\rho)}, \tag{22}$$

and for the multi-server queue

$$\overline{W}_n \leq \frac{\lambda}{4} \cdot \frac{\left(\Gamma_a + \Gamma_s/m^{1/2}\right)^2}{1-\rho}. \tag{23}$$

In traditional queueing theory, Kingman [44] provides insightful bounds on the expected waiting time in steady state for the $GI/GI/1$ queue

$$\mathbb{E}[W_n] \leq \frac{\lambda}{2} \cdot \frac{\sigma_a^2 + \sigma_s^2}{1-\rho}, \tag{24}$$

and for the $GI/GI/m$ queue

$$\mathbb{E}[W_n] \leq \frac{\lambda}{2} \cdot \frac{\sigma_a^2 + \sigma_s^2/m + (1/m - 1/m^2)/\mu^2}{1-\rho}. \tag{25}$$

Contrasting Kingman's bounds (24) and (25) with the bounds (22) and (23), we observe that they have the same functional dependence on $\lambda/(1-\rho)$ and on the variability parameters Γ_a^2, Γ_s^2/m, (correspondingly σ_a^2, σ_s^2/m). In this sense, both approaches lead to the same qualitative insights.

(b) Heavy-tailed behavior: During the past decade, studies have shown the heavy-tailed behavior of internet traffic ([30,42,49,51,71]). The absence of closed form expressions for queueing systems with heavy tail behavior has made it difficult to make progress in the area of communication networks scheduling. Using our approach, we are able to provide closed form expressions for the waiting times [Eqs. (15) and (16)], which to the best of our knowledge, are not available under stochastic heavy-tailed assumptions on arrivals and services.

3.3 Analysis of single class queueing networks

In this section, we analyze single class queueing networks in our framework. Consider a network of J queues serving a single class of customers. Each customer enters the network through some queue j, and either leaves the network or departs towards another queue right after completion of his service. In order the analyze the waiting time in a particular queue j in the network, we need to characterize the overall arrival process to queue j and then apply Theorem 2. The arrival process in queue j is the superposition of different processes, each of which is either a process from the outside world, or a departure process from another queue, or a thinning of a departure process from another queue, or a thinning of an external arrival process. Correspondingly, in order to analyze the network, we need to characterize the effect that the following operations have on the arrival process:

(a) Passing through a queue: Under this operation, we characterize the departure process $\{D_i\}_{i\geq 1}$ when an arrival process $\{T_i\}_{i\geq 1} \in \mathcal{U}^a$ passes through a queue. We show that the interdeparture times belong to an uncertainty set that has the same form as the uncertainty set for the interarrival times. This is the generalization of Burke's theorem (Burke [24]) for an M/M/m queue. This is a significant advantage of our approach compared to modeling queues with arrival processes from a renewal process. While, the departure process satisfies the same properties as the arrival process under our framework, in traditional queueing networks the departure process fails to be a renewal process.

(b) **Superposition of arrival processes:** Under this operation, m arrival processes $\{T_i^j\}_{i\geq 1} \in \mathcal{U}_j^a$, $j = 1, \ldots, m$ combine to form a single arrival process. Proposition 2 characterizes the uncertainty set of the combined arrival process.

(c) **Thinning of a process with probability p:** Under this operation, an arrival from a given arrival process is classified as type I with probability p and type II with probability $1 - p$. In Proposition 3, we characterize the uncertainty set of the resulting thinned type I process.

Passing through a queue The next theorem is the analog of Burke's theorem in our framework.

Theorem 3 *If $\{T_i\}_{i\geq 1} \in \mathcal{U}^a$, $\{X_i\}_{i\geq 1} \in \mathcal{U}^s$, $\alpha_a = \alpha_s = \alpha$ and $\rho < 1$, then the interdeparture times $\{D_i\}_{i\geq 1}$ belong to the uncertainty set*

$$\mathcal{U}^d = \left\{ (D_1, \ldots, D_n) \left| \frac{\left|\sum_{i=k+1}^n D_i - \frac{n-k}{\lambda}\right|}{(n-k)^{1/\alpha}} \leq \Gamma_a + c_{n-k},\ \forall k \leq k_0 \right.\right\}, \tag{26}$$

where

$$c_k = \frac{1}{k^{1/\alpha}} \left(\frac{\alpha - 1}{\alpha^{\alpha/(\alpha-1)}} \cdot \frac{\lambda^{1/(\alpha-1)} \cdot (\Gamma_a + \Gamma_s)^{\alpha/(\alpha-1)}}{(1-\rho)^{1/(\alpha-1)}} + \frac{1}{\lambda} \right) = \mathcal{O}\left(\frac{1}{k^{1/\alpha}}\right).$$

Proof The nth interdeparture time is expressed as $D_n = T_n + W_n - W_{n-1} + X_n - X_{n-1}$, thus

$$\sum_{i=k+1}^n D_i = \sum_{i=k+1}^n T_i + W_n - W_k + X_n - X_k \tag{27}$$

$$\leq \sum_{i=k+1}^n T_i + W_n + X_n. \tag{28}$$

Combining Eqs. (14) and (28), we obtain

$$\sum_{i=k+1}^n D_i \leq \sum_{i=k+1}^n T_i + \max_{1\leq j\leq n} \left(\sum_{\ell=j}^n X_\ell - \sum_{\ell=j+1}^n T_\ell \right).$$

We seek to maximize the right-hand side over sets $\mathcal{U}^s$ and $\mathcal{U}^a$

$$\sum_{i=k+1}^n D_i \leq \max_{\mathbf{X}\in\mathcal{U}^s, \mathbf{T}\in\mathcal{U}^a} \left\{ \sum_{i=k+1}^n T_i + \max_{1\leq j\leq n} \left(\sum_{\ell=j}^n X_\ell - \sum_{\ell=j+1}^n T_\ell \right) \right\},$$

$$= \frac{n-k}{\lambda} + \Gamma_a (n-k)^{1/\alpha} + \max_{\mathbf{X}\in\mathcal{U}^s, \mathbf{T}\in\mathcal{U}^a} \left\{ \max_{1\leq j\leq n} \left(\sum_{\ell=j}^n X_l - \sum_{\ell=j+1}^n T_l \right) \right\}.$$

A similar procedure is done as in Eq. (17) by switching the maximization operators and bounding the sums of service times and interarrival times by Assumption 1, the maximum term simplifies to the one dimensional concave maximization problem

$$\max_{1\leq j\leq n}\left\{\Gamma_a\,(n-j)^{1/\alpha}+\Gamma_s(n-j+1)^{1/\alpha}-(n-j)\,\frac{1-\rho}{\lambda}+\frac{1}{\mu}\right\},\ \text{for } j\leq k_0 \tag{29}$$

which is of the form

$$\max_{1\leq x\leq n}\beta\cdot x^{1/\alpha}+\delta\,(x+1)^{1/\alpha}-\gamma\cdot x\leq\max_{1\leq x\leq n}(\beta+\delta)(x+1)^{1/\alpha}-\gamma(x+1)+\gamma$$
$$=\frac{\alpha-1}{\alpha^{\alpha/(\alpha-1)}}\frac{(\beta+\delta)^{\alpha/(\alpha-1)}}{\gamma^{1/(\alpha-1)}}+\frac{1-\rho}{\lambda},\tag{30}$$

where $\beta=\Gamma_a$, $\delta=\Gamma_s$, $\gamma=(1-\rho)/\lambda>0$, given $\rho<1$. Note that Eq. (30) is maximized at

$$x^*+1=\left(\frac{\beta+\delta}{\alpha\gamma}\right)^{\alpha/(\alpha-1)}=\left(\frac{\lambda(\Gamma_a+\Gamma_s)}{\alpha(1-\rho)}\right)^{\alpha/(\alpha-1)}.\tag{31}$$

Therefore, we obtain the upper bound

$$\frac{\sum_{i=k+1}^{n}D_i-\frac{n-k}{\lambda}}{(n-k)^{1/\alpha}}\leq\Gamma_a+c_{n-k}.$$

The lower bound is obtained similarly. □

Superposition of multiple processes The next proposition (for a proof see [6]) characterizes the resulting uncertainty set when two processes combine.

Proposition 2 *The superposition of arrival processes characterized by the uncertainty sets*

$$\mathcal{U}_j^a=\left\{(T_1,\ldots,T_n)\left|\frac{\left|\sum_{i=k+1}^{n}T_i-\frac{n-k}{\lambda_j}\right|}{(n-k)^{1/\alpha}}\leq\Gamma_{a,j},\ \forall k\leq k_0\right.\right\},\ 1\leq j\leq J,\tag{32}$$

results in a merged arrival process characterized by the uncertainty set

$$\mathcal{U}_{sup}^a=\left\{(T_1,\ldots,T_n)\left|\frac{\left|\sum_{i=k+1}^{n}T_i-\frac{n-k}{\lambda_{sup}}\right|}{(n-k)^{1/\alpha}}\leq\Gamma_{a,sup},\ \forall k\leq k_0\right.\right\},$$

where

$$\lambda_{sup} = \sum_{j=1}^{J} \lambda_j, \quad \Gamma_{a,sup} = \frac{\left(\sum_{j=1}^{J} \left(\lambda_j \Gamma_{a,j}\right)^{\alpha/(\alpha-1)}\right)^{(\alpha-1)/\alpha}}{\sum_{j=1}^{J} \lambda_j}. \tag{33}$$

Thinning of a process We consider the case in which an arrival process is thinned, that is a fraction $(1-p)$ of arrivals of the original process are discarded. For a proof see [6].

Proposition 3 *When an arrival process characterized by the uncertainty set*

$$\mathcal{U}^a = \left\{ (T_1, \ldots, T_n) \;\middle|\; \frac{\left| \sum_{i=k+1}^{n} T_i - \dfrac{n-k}{\lambda} \right|}{(n-k)^{1/\alpha}} \le \Gamma_a, \;\; \forall k \le k_0 \right\},$$

is thinned with probability p, the resulting arrival process is described by the uncertainty set

$$\mathcal{U}^a_{split} = \left\{ (T_1^{split}, \ldots, T_n^{split}) \;\middle|\; \frac{\left| \sum_{i=k+1}^{n} T_i^{split} - \dfrac{n-k}{\lambda_{split}} \right|}{(n-k)^{1/\alpha}} \le \Gamma_{a,split}, \;\; \forall k \le k_0 \right\},$$

where $\lambda_{split} = \lambda \cdot p$ *and* $\Gamma_{a,split} = \Gamma_a \cdot \left(\dfrac{1}{p}\right)^{1/\alpha}$.

The analysis of single class queueing networks We consider a single class queueing network of single servers with the following data:

- **(a)** External arrival processes with parameters $\left(\lambda_j, \Gamma_{a,j}, \alpha_{a,j}\right)$ that arrive to each node $j = 1, \ldots, J$.
- **(b)** Service processes with parameters $\left(\mu_j, \Gamma_{s,j}, \alpha_{s,j}\right)$, and the number of servers $m_j, \; j = 1, \ldots, J$.
- **(c)** Routing matrix $\mathbf{P} = [P_{ij}], i, j = 1, \ldots, J$ with the interpretation that after completing service in queue i, a customer is routed to queue j with probability P_{ij} and leaves the network with probability $1 - \sum_j P_{ij}$.

The following theorem combines Theorem 3, and Propositions 2 and 3.

Theorem 4 *The behavior of a single class queueing network is equivalent to that of a collection of independent queues, with the arrival process to node j characterized by the uncertainty set*

$$\mathcal{U}^a_j = \left\{ (T_1^j, \ldots, T_n^j) \;\middle|\; \frac{\left| \sum_{i=k+1}^{n} T_i^j - \dfrac{n-k}{\overline{\lambda}_j} \right|}{(n-k)^{1/\alpha}} \le \overline{\Gamma}_{a,j}, \;\; \forall k \le k_0 \right\}, \quad j = 1 \ldots, J,$$

where $\{\overline{\lambda}_1, \overline{\lambda}_2, \ldots, \overline{\lambda}_J\}$ and $\{\overline{\Gamma}_{a,1}, \overline{\Gamma}_{a,2}, \ldots, \overline{\Gamma}_{a,j}\}$ *satisfy the set of equations for all* $j = 1, \ldots, J$

$$\overline{\lambda}_j = \lambda_j + \sum_{i=1}^{J} \left(\overline{\lambda}_i P_{ij}\right), \tag{34}$$

$$\overline{\Gamma}_{a,j} = \frac{\left[\left(\lambda_j \cdot \Gamma_{a,j}\right)^{\alpha/(\alpha-1)} + \sum_{i=1}^{J} \left(\overline{\lambda}_i \cdot \overline{\Gamma}_{a,i}\right)^{\alpha/(\alpha-1)} \cdot P_{ij}\right]^{(\alpha-1)/\alpha}}{\overline{\lambda}_j}. \tag{35}$$

By applying Theorem 2 using the parameters we compute in Theorem 4, we can now compute performance measures in a single class queueing network.

Queues with asymmetric heavy-tailed arrivals and services All the results presented in this section assumed that the arrival and the service process have the same tail coefficients. In [6], we present results for the case of asymmetric heavy-tailed arrival and service processes, that is, when $\alpha_a \neq \alpha_s$. We present analogs of Theorems 3, 2, and Propositions 2 and 3 that allow us to analyze queueing networks composed of queues with arbitrary values for α_a's and α_s's.

3.4 Computational results

In this section, we present computational results to demonstrate the effectiveness of our approach in analyzing queueing networks. We shall refer to our approach as the Robust Queueing Network Analyzer (RQNA). We compare the results obtained by RQNA with the results obtained from simulation and the Queueing Network Analyzer (QNA) proposed by Whitt [70], and investigate the relative performance of RQNA with respect to system's network size, degree of feedback, maximum traffic intensity, and diversity of external arrival distributions.

In view of comparing our approach to simulation and QNA, we consider instances of stochastic queueing networks with the following primitive data:

(a) The distributions of the external arrival processes with parameters $(\lambda_j, \sigma_{a,j}, \alpha_{a,j})$ with coefficients of variation $c_{a,j}^2 = \lambda_j^2 \sigma_{a,j}^2$, $j = 1, \ldots, J$.
(b) The distributions of the service processes with parameters $\left(\mu_j, \sigma_{s,j}, \alpha_{s,j}\right)$ with coefficients of variation $c_{s,j}^2 = \mu_j^2 \sigma_{s,j}^2$ and the number of servers m_j, $j = 1, \ldots, J$.
(c) The routing matrix $\mathbf{P} = [P_{ij}]$, $i, j = 1, \ldots, J$ with the interpretation that after completing service in queue i, a customer is routed to queue j with probability P_{ij} and leaves the network with probability $1 - \sum_j P_{ij}$.

To apply RQNA on stochastic queueing networks, we first need to translate the stochastic primitive data given above into robust primitive data, namely uncertainty sets with appropriate variability parameters $(\Gamma_{a,j}, \Gamma_{s,j})$ for each $j = 1, \ldots, J$. To achieve this goal, we next describe how we use simulation on a single isolated queue to construct parameters (Γ_a, Γ_s) given arrival and service distributions. This enables us to transform the stochastic data into uncertainty sets over external arrival and service processes.

Derived variability parameters Along the lines of QNA, we use simulation to construct appropriate functions for the variability parameters. To do so, we consider a single queue with m servers characterized by $(\rho, \sigma_a, \sigma_s, \alpha_a, \alpha_s)$ and model its variability parameters $(\varGamma_a, \varGamma_s)$ as follows

$$\varGamma_a = \sigma_a \quad \text{and} \quad \varGamma_s = f(\rho, \sigma_a, \sigma_s, \alpha_a, \alpha_s). \tag{36}$$

Motivated by Kingman's bound [see Eq. (25)], we consider the following functional form for $f(\cdot)$

$$f(\rho, \sigma_s, \sigma_a, \alpha_a, \alpha_s) = \left(\theta_0 + \theta_1 \cdot \sigma_s^2/m + \theta_2 \cdot \sigma_a^2 \rho^2 m\right)^{(\alpha-1)/\alpha} - \sigma_a m^{(\alpha-1)/\alpha},$$

where $\alpha = \min\{\alpha_a, \alpha_s\}$.

We run simulation over multiple instances of a single queue while varying parameters $(\rho, \sigma_a, \sigma_s, \alpha_a, \alpha_s)$ for different arrival and service distributions. We employ linear regression to generate appropriate values for θ_0, θ_1 and θ_2 such that the values obtained for $\overline{W}_n$ by Theorem 2 are adapted according to the expected values of the waiting time obtained from simulation.

The RQNA Algorithm Having derived the required primitive data for our robust approach, we next describe the RQNA algorithm we employ to compute performance measures of a given network of queues.

Algorithm 1 ***Robust Queueing Network Analyzer (RQNA)***
Input: *Parameters* $\left\{\left(\lambda_j, \sigma_{a,j}, \alpha_{a,j}\right), \left(\mu_j, \sigma_{s,j}, \alpha_{s,j}\right), \mathbf{P} = [P_{ij}]\right\}, \ 1 \leq i, j \leq J.$
Output: *Waiting times* $\overline{W}_n$ *at each node* $j, \ 1 \leq j \leq J.$
Algorithm:

1. *For each external arrival process* i *in the network, set* $\varGamma_{a,i} = \sigma_{a,i}$.
2. *For each queue* j *in the network with parameters* $(\mu_j, \sigma_{s,j}, \alpha_{s,j})$, *compute*
   - **(a)** *the effective parameters* $\left(\overline{\lambda}_j, \overline{\varGamma}_{a,j}, \overline{\alpha}_{a,j}\right)$ *according to Theorem 4 and set* $\rho_j = \overline{\lambda}_j / \mu_j$,
   - **(b)** *the variability parameter* $\varGamma_{s,j} = f\left(\rho_j, \overline{\varGamma}_{a,j}, \sigma_{s,j}, \overline{\alpha}_{a,j}, \alpha_{s,j}\right)$, *and*
   - **(c)** *the waiting time* $\overline{W}_n$ *at node* j *using Theorem 2.*

Note that, in Step 2(b), we treat each queue j in the network separately as a single isolated queue with an effective arrival process described by the variability parameter $\overline{\varGamma}_{a,j}$. Note that we use $\overline{\varGamma}_{a,j}$ as an input to $f(\cdot)$ in place of the standard deviation. This is motivated from our use of $\varGamma_a = \sigma_a$ for the single queue case [see Eq. (36)].

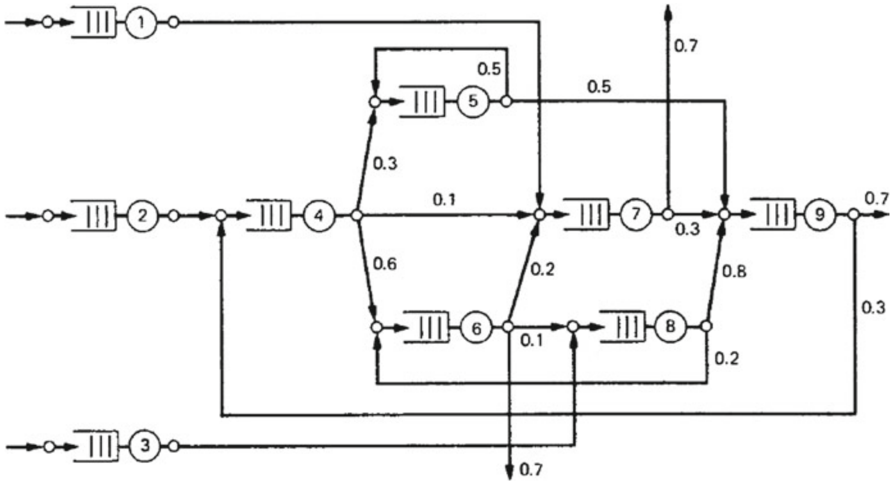

Fig. 1 The Kuehn's network (see Kuehn [48])

Table 1 Single-server network: sojourn time percent errors relative to simulation

Case	Pareto distribution		Normal distribution	
(c_a^2, c_s^2)	QNA	RQNA	QNA	RQNA
(0.25, 0)	22.78	3.291	15.28	1.389
(0.25, 1)	18.48	−3.478	12.08	3.869
(0.25, 4)	20.13	−3.052	11.57	−3.882
(1,0)	19.01	1.056	12.68	−3.797
(1,1)	14.06	1.799	5.84	−2.555
(1,4)	10.15	2.893	−10.45	−0.681
(4,0)	21.82	−1.934	10.95	1.290
(4,1)	23.71	−2.139	14.18	−3.508
(4,4)	17.51	−2.974	11.55	1.671

Performance of RQNA in comparison to QNA and simulation We consider the network shown in Fig. 1 and perform computations assuming queues have either single or multiple servers, with normal or Pareto distributed service times.

Table 1 reports the percentage errors between the expected sojourn times calculated by simulation and those obtained by each of QNA and RQNA, assuming all nine queues in the network have a single server. Note that the sojourn time is defined as the time elapsed between the arrival of a customer to the network until his departure from the network. Table 2 summarizes the percentage errors for RQNA relative to simulation for queues with 3, 6, and 10 servers. We observe that RQNA produces results that are often significantly closer to simulated values compared to QNA. RQNA is fairly insensitive to the heavy-tailed nature of the service distributions. In fact, the sojourn time percentage errors for both the Pareto and normally distributed services

Table 2 Multi-server network: sojourn time percent errors

Case	3 Servers		6 Servers		10 Servers	
$(c^2_{a,j}, c^2_{s,j})$	Normal	Pareto	Normal	Pareto	Normal	Pareto
(0.25, 0)	2.095	−2.732	2.628	−3.475	2.844	−3.655
(0.25, 1)	3.255	−0.803	4.034	−1.065	4.419	−0.992
(0.25, 4)	−2.067	1.416	−2.557	1.793	−2.760	1.911
(1, 0)	−2.254	−3.663	−2.886	−4.609	−2.877	−4.731
(1, 1)	−3.183	1.725	−4.133	2.232	−3.975	2.230
(1, 4)	3.859	1.529	4.978	1.936	5.118	1.998
(4, 0)	−3.852	4.628	−5.823	5.358	−5.429	5.309
(4, 1)	−3.272	−4.283	−4.372	−4.83	−4.228	−5.667
(4, 4)	−3.282	−4.123	−5.823	−5.823	−5.834	−6.129

Table 3 Single-server networks: RQNA percent error as a function of network size and degree of feedback

% Feedback loops/no. of nodes	N = 10	N = 15	N = 20	N = 25	N = 30
Feed-forward networks 0 %	2.86	2.94	3.03	2.92	3.21
20 %	3.12	3.25	3.29	3.71	3.64
35 %	3.74	3.81	4.02	4.07	4.14
50 %	4.42	4.63	4.84	5.23	5.65
70 %	4.85	5.16	5.34	5.68	5.86

Table 4 Multi-server networks: RQNA percent error as a function of network size and degree of feedback

% Feedback loops/no. of nodes	N = 10	N = 15	N = 20	N = 25	N = 30
Feed-forward networks 0 %	3.594	3.546	3.756	3.432	3.846
20 %	3.696	4.014	4.02	4.392	4.452
35 %	4.32	4.776	4.956	5.034	4.878
50 %	4.95	4.806	5.358	5.67	6.192
70 %	5.016	5.556	5.934	5.958	6.03

are within the same order. Furthermore, RQNA's performance is generally stable with respect to the number of servers at each queue, yielding errors within the same range for instances with 3 to 10 servers per queue.

Performance of RQNA as a function of network parameters We investigate the performance of RQNA (for the service dependent adaptation regime) as a function of the system's parameters (network size, degree of feedback, maximum traffic intensity among all queues and number of distinct distributions for the external arrival processes) in families of randomly generated queueing networks. Tables 3 and 4 report the sojourn time percentage errors of RQNA relative to simulation as a function of the

Table 5 Single-server networks: RQNA percent error as a function of traffic intensity and variety of external arrival distributions

No. of different distributions	$\rho = 0.95$	$\rho = 0.9$	$\rho = 0.8$	$\rho = 0.65$	$\rho = 0.5$
1	3.34	3.26	3.17	3.02	2.72
2	6.38	5.85	5.47	4.87	3.24
3	7.43	7.09	6.04	5.88	4.53
4	7.56	6.98	6.81	6.29	5.18

Table 6 Multi-server networks: RQNA percent error as a function of traffic intensity and variety of external arrival distributions

No. of different distributions	$\rho = 0.95$	$\rho = 0.9$	$\rho = 0.8$	$\rho = 0.65$	$\rho = 0.5$
1	4.05	4.092	3.618	3.678	3.228
2	5.082	7.104	6.42	6.108	3.714
3	5.916	6.318	6.9	7.344	5.676
4	7.672	8.644	7.284	6.852	5.37

size of the network and the degree of feedback for queues with single and multiple servers, respectively. In the case of multi-server queueing networks, we randomly assign 3, 6 or 10 servers to each of the queues in the network independently of each other.

Tables 5 and 6 present the sojourn time percentage errors for RQNA relative to simulation as a function of the maximum traffic intensity among all queues in the network and the number of distinct distributions for the external arrival processes. Table 5 presents the results for networks with only single server queues, while Table 6 presents the results for networks in which each queue was randomly assigned 3, 6 or 10 servers. Specifically, we design four sets of experiments in which we use one type (normal), two types (Pareto and normal), three types (Pareto, normal and Erlang) and four types (Pareto, normal, Erlang and exponential) of arrival distributions. We observe that

- **(a)** The percentage errors are slightly higher for multi-server networks compared to single-server networks.
- **(b)** The performance of RQNA is generally stable with an increased degree of feedback with errors below 6.2 %.
- **(c)** RQNA is fairly insensitive to network size with a very slight increase in percent errors between 10-node and 30-node networks.
- **(d)** RQNA presents slightly improved results for lower traffic intensity levels. It is nevertheless fairly stable with respect to higher traffic intensity levels.
- **(e)** The percentage errors generally increase with diversity of external arrival distributions, but still are below 8.5 % relative to simulation.

3.5 Extensions

In this section, we have introduced our approach in the context of queueing networks with a single class of customers. For more details, please refer to [6]. We have also extended our approach to analyze more involved queueing systems, mainly in two directions: performance analysis of multi-class queueing networks in [5] and performance analysis of queueing systems in the transient domain in [7]. In [5], we consider networks under FCFS and priority policies. Our conclusions for the multi-class setting parallel that of single class queueing networks. In [7], we concentrate on the transient analysis of single class queues, and feed-forward networks, and derive closed form expressions for the transient behavior of such systems.

4 Optimal mechanism design for multi-item auctions

The optimal design of auctions is a central problem in Economics which arises when an auctioneer is interested in selling multiple items to multiple buyers with private valuations for the items. The auctioneer is faced with the task of designing the rules of the auction so as to maximize revenue, while also incentivizing the buyers to reveal their true valuations when they participate in the auction. Building on the work of Vickrey [68], Myerson [57] considers the optimal auction design problem for the sale of a single item to buyers with unlimited budgets. He considers this problem in a probabilistic setting, that is, he assumes that the buyers' valuations are drawn from independent, but not necessarily identical, probability distributions. These distributions are assumed to be common knowledge, so that all buyers and the auctioneer know the distribution from which each buyer's valuation is drawn. He obtains a characterization of the optimal solution as a second price auction with buyer dependent reservation prices, which for the special case of identical buyers, reduces to that of a second price auction with a single reservation price.

In the past decade, auction theory has also attracted the attention of researchers in Theoretical Computer Science. In what follows, we present a brief review of the relevant literature around the predominant modeling paradigms mentioned earlier. For a more comprehensive review, we refer the readers to [46,47,69] for the Economics and [59] for the Computer Science perspective, respectively.

The probabilistic approach

This approach has been widely studied (see [26,50,54,61,67,72]). The key primitive assumptions are:

- **(a)** Buyers' valuations are sampled from a joint probability distribution;
- **(b)** The auctioneer has exact knowledge of this joint distribution;
- **(c)** The auctioneer is risk neutral and seeks to obtain a mechanism in order to maximize the expected revenue.

In this setting, we divide the literature based on the problem that was solved.

Public budget constraints (Problem P1): The analysis of budget constrained auctions was first done by [50], where they assume that all buyers have the same common knowledge budget constraint and derive the subsidy-free (i.e., payments are non-negative) optimal auction. Under the same assumption of equal budgets, Maskin [55] obtained the optimal auction that maximizes social surplus. Malakhov and Vohra [53] relaxed the assumption of symmetrical budgets and obtained the revenue maximizing auction for the case of two buyers, only one of whom is budget constrained. Chawla et al. [25] obtained the first approximation algorithm for the general problem where they show that a sequential all-pay mechanism is a 4-approximation to the revenue of the optimal truthful mechanism with a discrete valuation space for each bidder. They also show that a sequential posted price mechanism is an $O(1)$-approximation to the revenue of the optimal truthful mechanism, when the valuation space of each bidder is a product distribution that satisfies the standard hazard rate condition. Dobzinski et al. [35] shows that an adaptive version of the "clinching auction" ([1]) is Pareto-optimal and incentive compatible. Moreover, they show that it is the unique auction with these properties, when there are exactly two bidders. The more general problem, however, remains open in the setting of public budget constraints under probabilistic assumptions.

Private budget constraints (Problem P2): Dobzinski et al. [35] show that there is no incentive compatible, individual rational and Pareto-optimal deterministic auction, for any finite number $m > 1$ of units of a single indivisible good and any $n > 2$ players, when the budgets are private. In the same setting, [22] showed that it is impossible to design a non-trivial truthful auction which allocates all units. Instead they provide the design of an asymptotically revenue-maximizing truthful mechanism which may allocate only some of the units. Furthermore, Pai and Vohra [61] shows several interesting qualitative properties of such auctions by discretizing the valuation space and formulating a linear optimization problem, whose dimension is exponential in the number of buyers. Based on these results, there is a need to consider other notions of optimality in order to obtain computationally tractable auction mechanisms.

Correlated valuations (Problem P3): For the case of correlated buyers which was left open by Myerson [57], some of the early work was done by Cremer [29] who solved it in a weak sense, that is, using auctions that are individually rational only in expectation. However, the computational complexity of designing the optimal ex-post individually rational auction for correlated valuations has been open until recently, when [62] obtained a polynomial time algorithm for the two buyer case and established an inapproximability result for three or more buyers.

The adversarial approach

The objective in the adversarial approach is to identify a single mechanism that always has good performance, e.g., under any distributional assumption. There have been broadly three approaches that have been used so far:

(a) The *resource augmentation* approach, also known as, the *bicriteria* approach which was introduced in [23], is based on the observation that in some cases increasing competition, e.g., by recruiting more agents, and running the second price auction mechanism increases revenue when compared to running the (optimal) Myerson mechanism in the original setting.

(b) The main idea in the *average-case* approach is to show that, for a large class of distributions and settings, there is a single mechanism that approximates the revenue of the Bayesian optimal mechanism. For example, when the probability distribution is known, the second price auction mechanism with a particular way (the so-called monopoly reservation price) of calculating the reservation prices is approximately optimal by a constant factor (of 2). Dhangwatnotai et al. [34] relaxes the need to know the probability distribution of the valuation and uses a sampling-based approach to calculate the reservation price. For the case of correlated buyers, [63] proposed a mechanism for the correlated case that achieves half of the optimum revenue.

(c) The *worst-case* approach, where the idea is to define an appropriate performance benchmark and attempt to obtain mechanisms that approximate this benchmark on any worst-case valuation vector. Goldberg [40], in a negative result, showed that when the adversary knows all of the buyers' valuations exactly, then no incentive compatible auction can obtain more than a vanishingly small fraction of its revenue in the worst case. Under this approach, it is desirable to identify the right kind of performance benchmarks, but this problem is still open.

All the aforementioned results have been for the cases of buyers without budget constraints and, except for the result in [63], they all assume independent valuations. Thus, Problems (P1)–(P3) are open under the adversarial approach.

In what follows, we revisit the auction design problem for multi-item auctions with budget constrained buyers by using a robust optimization approach to model **(a)** concepts such as incentive compatibility and individual rationality that are naturally expressed in the language of robust optimization and **(b)** the auctioneer's beliefs on the buyers' valuations of the items. In this setting, we provide a characterization of the optimal solution as an auction with reservation prices, thus extending the work of Myerson [57] from single item without budget constraints, to multiple items with budgets, potentially correlated valuations and uncertain budgets. We report computational evidence that suggests the proposed approach **(a)** is numerically tractable for large scale auction design problems, **(b)** leads to improved revenue compared to the classical probabilistic approach when the true distributions are different from the assumed ones, and **(c)** leads to higher revenue when correlations in the buyers' valuations are explicitly modeled.

4.1 The robust optimization approach

In this section, we summarize the major results of Bandi and Bertsimas [3].

Models of valuations

We consider a setting where n buyers, indexed by $i \in \mathcal{N}$, are interested in a set of m items, indexed by $j \in \mathcal{M}$, made available by an auctioneer. Each buyer $i \in \mathcal{N}$ has a valuation v_{ij} associated with each of the items $j \in \mathcal{M}$, which is not known to the auctioneer. Additionally the buyers are also budget constrained with budgets $\{B_1, B_2, \ldots, B_n\}$. Before proceeding further, we introduce the following notation.

For each item $j \in \mathcal{M}$, let $\mathbf{v}_j = \left(v_{1j}, v_{2j}, \ldots, v_{nj}\right) \in \mathbb{R}^n$ be the vector of valuations for the *j*th item by the *n* bidders. We let $\mathbf{v} = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m)$ denote the concatenation of the vectors $\mathbf{v}_j$, $j \in \mathcal{M}$. With a slight abuse of notation, let $\mathbf{v}_i = (v_{i1}, v_{i2}, \ldots, v_{im})$ be the vector of valuations for the *i*th bidder for all items. In the same vein, we let $\mathbf{v}_{-i,j} = \left(v_{1j}, \ldots, v_{i-1,j}, v_{i+1,j}, \ldots, v_{nj}\right) \in \mathbb{R}^{n-1}$ be the vector of valuations of all bidders except i, for item j, $\forall j \in \mathcal{M}$. And let $\mathbf{v}_{-i} = \left(\mathbf{v}_{-i,1}, \ldots, \mathbf{v}_{-i,m}\right) \in \mathbb{R}^{(n-1)\times m}$ be the concatenation of the vectors $\left(\mathbf{v}_{-i,j}\right)_{j\in\mathcal{M}}$. Finally, we write $\mathbf{v} \in \mathcal{U}$ to denote $\mathbf{v}_j \in \mathcal{U}_j$, $j \in \mathcal{M}$.

For each item $j \in \mathcal{M}$, we model the auctioneer's beliefs on valuations for this item using an uncertainty set $\mathcal{U}_j \in \mathbb{R}^n$, from which the n dimensional valuation vector $\mathbf{v}_j$ is derived. We can construct such an uncertainty set in multiple ways, depending on the type of information that we have access to. Specifically, we use the following uncertainty sets, if valuations are (a) i.i.d Eq. (1), (b) correlated Eq. (4), and (c) normally distributed Eq. (8). See Sect. 2 for more discussion.

Optimization formulation

We next introduce the concept of worst case optimality and show how the resulting auction design problem can be formulated as a robust linear optimization problem. In this case, the objective is to maximize the worst case revenue over all valuation vectors $\mathbf{v}$ lying in an uncertainty set $\mathcal{U}$. We introduce the decision variables $\mathbf{x}^{\mathbf{v}}$ and $\mathbf{p}^{\mathbf{v}}$ that represent the allocation and the payment rules, respectively, for all valuation vectors $\mathbf{v} \in \mathcal{U}$. That is, if the realized valuation vector is $\mathbf{v}$, then we allocate a fraction $x_{ij}^{\mathbf{v}}$ of item j to buyer i, and charge a total of $p_i^{\mathbf{v}}$ to the *i*th buyer. Note that we do not account for payments of buyer i relative to item j, but only account for the total payment of buyer i.

The allocation and payment rules should be chosen to satisfy the following properties:

- **(a)** *Individual Rationality (IR)* : This property ensures that the buyers do not derive negative utility by participating in the auction when they bid truthfully.
- **(b)** *Budget Feasibility (BF)* : This property ensures that each buyer is charged within his budget constraints.
- **(c)** *Incentive Compatibility (IC)* : This property ensures that the total utility of the *i*th buyer under truthful bidding, which is given by

$$U\left(\mathbf{v}_i, \mathbf{v}_{-i}\right) = \sum_{j\in\mathcal{M}} v_{ij} x_{ij}^{(\mathbf{v}_i,\mathbf{v}_{-i})} - p_i^{(\mathbf{v}_i,\mathbf{v}_{-i})} \geq \sum_{j\in\mathcal{M}} v_{ij} x_{ij}^{(\mathbf{u}_i,\mathbf{v}_{-i})} - p_i^{(\mathbf{u}_i,\mathbf{v}_{-i})},$$

  is greater than the total utility that Buyer i derives by bidding any other other bid vector $\mathbf{u}_i$.

The optimal auction design problem with these properties, leads to the following linear optimization model:

$$Z^* = \max \quad W \tag{37}$$

$$\text{s.t.} \quad W - \sum_{i \in \mathcal{N}} p_i^{\mathbf{v}} \le 0, \quad \forall \mathbf{v} \in \mathcal{U}, \tag{38}$$

$$\sum_{i \in \mathcal{N}} x_{ij}^{\mathbf{v}} \le 1, \quad \forall j \in \mathcal{M}, \forall \mathbf{v} \in \mathcal{U}, \tag{39}$$

$$\sum_{j \in \mathcal{M}} v_{ij} \cdot x_{ij}^{(\mathbf{u}_i, \mathbf{v}_{-i})} - p_i^{(\mathbf{u}_i, \mathbf{v}_{-i})} - \sum_{j \in \mathcal{M}} v_{ij} \cdot x_{ij}^{(\mathbf{v}_i, \mathbf{v}_{-i})} \tag{40}$$

$$+ p_i^{(\mathbf{v}_i, \mathbf{v}_{-i})} \le 0, \quad \forall (\mathbf{v}_i, \mathbf{v}_{-i}) \in \mathcal{U}, \forall (\mathbf{u}_i, \mathbf{v}_{-i}) \in \mathcal{U}, \forall i \in \mathcal{N},$$

$$p_i^{\mathbf{v}} \le B_i, \quad \forall i \in \mathcal{N}, \forall \mathbf{v} \in \mathcal{U}, \tag{41}$$

$$p_i^{\mathbf{v}} \le \sum_{j \in \mathcal{M}} v_{ij} \cdot x_{ij}^{\mathbf{v}}, \quad \forall i \in \mathcal{N}, \forall \mathbf{v} \in \mathcal{U}, \tag{42}$$

$$\mathbf{x}^{\mathbf{v}} \ge \mathbf{0}.$$

The objective value (37) and the constraint (38) represent the fact that we are interested in maximizing the worst case revenue. Constraint (39) expresses the fact that at most one unit of item j can be assigned to all bidders. Constraints (40), (41), (42) implement the IC, BF and IR properties, respectively. We next present the dual problem of (37)–(42), by using the dual variables $\omega_{\mathbf{v}}$, $\xi_{j,\mathbf{v}}$, $\beta_{i,\mathbf{v}_{-i},\mathbf{v}_i,\mathbf{u}_i}$, $\eta_{i,\mathbf{v}}$, $\theta_{i,\mathbf{v}}$ that correspond to the constraints (38)–(42), respectively.

$$\min \sum_{\mathbf{v} \in \mathcal{U}} \left(\sum_{j=1}^{m} \xi_{j,\mathbf{v}} + \sum_{i=1}^{n} \eta_{i,\mathbf{v}} B_i \right) \tag{43}$$

$$\text{s.t.} \quad \xi_{j,(\mathbf{v}_i,\mathbf{v}_{-i})} + \sum_{\mathbf{u}_i} u_{ij} \cdot \beta_{i,\mathbf{v}_{-i},\mathbf{u}_i,\mathbf{v}_i} - v_{ij} \cdot \sum_{\mathbf{u}_i} \beta_{i,\mathbf{v}_{-i},\mathbf{v}_i,\mathbf{u}_i}$$

$$- v_{ij} \cdot \theta_{i,(\mathbf{v}_i,\mathbf{v}_{-i})} \ge 0, \quad \forall (\mathbf{v}_i, \mathbf{v}_{-i}) \in \mathcal{U},$$

$$\sum_{\mathbf{u}_i} \beta_{i,\mathbf{v}_{-i},\mathbf{v}_i,\mathbf{u}_i} - \sum_{\mathbf{u}_i} \beta_{i,\mathbf{v}_{-i},\mathbf{u}_i,\mathbf{v}_i} - \omega_{(\mathbf{v}_i,\mathbf{v}_{-i})} + \eta_{i,(\mathbf{v}_i,\mathbf{v}_{-i})}$$

$$+ \theta_{i,(\mathbf{v}_i,\mathbf{v}_{-i})} = 0, \quad \forall (\mathbf{v}_i, \mathbf{v}_{-i}) \in \mathcal{U},$$

$$\sum_{\mathbf{v} \in \mathcal{U}} \omega_{\mathbf{v}} = 1,$$

$$\omega_{\mathbf{v}} \ge 0, \boldsymbol{\xi}_{\mathbf{v}} \ge \mathbf{0}, \beta_{i,\mathbf{v}_{-i},\mathbf{v}_i,\mathbf{u}_i} \ge 0, \boldsymbol{\eta}_{\mathbf{v}} \ge \mathbf{0}, \boldsymbol{\theta}_{\mathbf{v}} \ge \mathbf{0}.$$

4.2 A robust optimal mechanism

In this section, we present a mechanism, that we call *ROM* (Robust Optimal Mechanism), that constitutes an optimal solution to the optimization problem (37). *ROM* consists of Algorithms 2 and 3, respectively. In Algorithm 2, which occurs prior to the realization of a specific bid vector $\mathbf{v}$, we compute the quantity R^*, which stands for the worst case revenue obtained when one uses *ROM*. In Algorithm 3, when the bid vector $\mathbf{v}$ is realized, we calculate the allocation vector $\left\{a_{ij}^{\mathbf{v}}\right\}_{i\in\mathcal{N},j\in\mathcal{M}}$ and the payments $\left\{p_i^{\mathbf{v}}\right\}_{i\in\mathcal{N}}$.

Algorithm 2 Calculation of the *worst case revenue.*
Input: *Uncertainty set $\mathcal{U}$, and budgets $B_1, \ldots, B_n$.*
Output: *Worst case revenue R^*.*
Algorithm:

1. *Compute the* worst case valuation vector $\mathbf{z} = \left\{z_{ij}\right\}_{i\in\mathcal{N},j\in\mathcal{M}}$ *given by*

$$\mathbf{z} = \arg\min_{\mathbf{v}\in\mathcal{U}} \left\{ \begin{array}{ll} \max\limits_{x_{ij},r_i} & \sum\limits_{i\in\mathcal{N}} r_i \\ s.t. & \sum\limits_{j\in\mathcal{M}} x_{ij}\cdot v_{ij} \le B_i, \ \forall i \in \mathcal{N}, \\ & r_i \le \sum\limits_{j\in\mathcal{M}} x_{ij}\cdot v_{ij}, \ \forall i\in\mathcal{N}, \\ & \sum\limits_{i\in\mathcal{N}} x_{ij} \le 1, \ \forall j\in\mathcal{M}, \\ & \mathbf{x} \ge \mathbf{0}. \end{array} \right\}. \tag{44}$$

2. *Compute* $\left(\left\{\xi_j^*\right\}_{j\in\mathcal{M}}, \left\{\eta_i^*\right\}_{i\in\mathcal{N}}, \left\{\theta_i^*\right\}_{i\in\mathcal{N}}\right)$ *given by*

$$(\boldsymbol{\xi}, \boldsymbol{\eta}, \boldsymbol{\theta}) = \arg \left\{ \begin{array}{ll} \min\limits_{\{\xi_j,\eta_i,\theta_i\}} & \sum\limits_{j\in\mathcal{M}} \xi_j + \sum\limits_{i\in\mathcal{N}} \eta_i B_i \\ s.t. & \xi_j + z_{ij}\cdot\eta_i \ge z_{ij}\cdot\theta_i, \ \forall i\in\mathcal{N}, \ \forall j\in\mathcal{M}, \\ & \theta_i = 1, \ \forall i\in\mathcal{N}, \\ & \boldsymbol{\xi}, \boldsymbol{\eta}, \boldsymbol{\theta} \ge \mathbf{0}. \end{array} \right\}. \tag{45}$$

3. *Compute the worst case revenue given by*

$$R^* = \sum_{j\in\mathcal{M}} \xi_j^* + \sum_{i\in\mathcal{N}} \eta_i^* B_i. \tag{46}$$

Algorithm 3 Calculation of *allocations* and *payments*.
Input: *Bid vector* $\mathbf{v} = \{v_{ij}\}_{i\in\mathcal{N}, j\in\mathcal{M}}$, *worst case revenue* R^*.
Output *Allocation vector* $\left\{a_{ij}^{\mathbf{v}}\right\}_{i\in\mathcal{N}, j\in\mathcal{M}}$ *and the payments* $\{p_i^{\mathbf{v}}\}_{i\in\mathcal{N}}$.
Algorithm:

1. *If* $\mathbf{v} \notin \mathcal{U}$, *then do not allocate any item and charge zero, otherwise proceed to Step 2.*
2. *Calculate the quantities* $\left(\left\{y_{ij}^{\mathbf{v}}\right\}_{i\in\mathcal{N}, j\in\mathcal{M}}, \{r_i^{\mathbf{v}}\}_{i\in\mathcal{N}}\right)$ *and* $\left(\left\{y_{ij,k}^{\mathbf{v}_{-k}}\right\}_{i\in\mathcal{N}, j\in\mathcal{M}}, \left\{r_{i,k}^{\mathbf{v}_{-\mathbf{k}}}\right\}_{\mathcal{N}}\right)_{k\in\mathcal{N}}$ *given by*

$$\left(\left\{y_{ij}^{\mathbf{v}}\right\}_{i\in\mathcal{N}, j\in\mathcal{M}}, \{r_i^{\mathbf{v}}\}_{i\in\mathcal{N}}\right) = \arg\max_{(\mathbf{y},\mathbf{r})\in\mathcal{P}^{\mathbf{v}}} \sum_{i\in\mathcal{N}} \left(\sum_{j\in\mathcal{M}} y_{ij}\cdot v_{ij} - r_i\right),$$

$$\left(\left\{y_{ij,k}^{\mathbf{v}_{-k}}\right\}_{i\in\mathcal{N}, j\in\mathcal{M}}, \left\{r_{i,k}^{\mathbf{v}_{-k}}\right\}_{i\in\mathcal{N}}\right) = \arg\max_{(\mathbf{y},\mathbf{r})\in\mathcal{P}^{\mathbf{v}}} \sum_{i\in\mathcal{N}\setminus\{k\}} \left(\sum_{j\in\mathcal{M}} y_{ij}\cdot v_{ij} - r_i\right),$$

where

$$\mathcal{P}^{\mathbf{v}} = \left\{ \left(\{x_{ij}\}_{i\in\mathcal{N}, j\in\mathcal{M}}, \{r_i\}_{i\in\mathcal{N}}\right) \;\middle|\; \begin{array}{ll} \sum_{j\in\mathcal{M}} x_{ij}v_{ij} \le B_i, & \forall i\in\mathcal{N}, \\ r_i \le \sum_{j\in\mathcal{M}} x_{ij}v_{ij}, & \forall i\in\mathcal{N}, \\ \sum_{i\in\mathcal{N}} x_{ij} \le 1, & \forall j\in\mathcal{M}, \\ \sum_{i\in\mathcal{N}} r_i \ge R^*, & \\ \mathbf{x} \ge \mathbf{0}. & \end{array} \right\}.$$

3. *Compute the allocation vector* $\{a_k^{\mathbf{v}}\}_{k\in\mathcal{N}}$ *and the payments* $\{p_k^{\mathbf{v}}\}_{k\in\mathcal{N}}$ *as follows*

$$a_k^{\mathbf{v}} = y_k^{\mathbf{v}},$$
$$p_k^{\mathbf{v}} = r_k^{\mathbf{v}} + \sum_{i\in\mathcal{N}\setminus\{k\}} \left(\sum_{j\in\mathcal{M}} y_{ij}^{\mathbf{v}_{-k}}\cdot v_{ij} - r_{i,k}^{\mathbf{v}_{-k}}\right) - \sum_{i\in\mathcal{N}\setminus\{k\}} \left(\sum_{j\in\mathcal{M}} y_{ij}^{\mathbf{v}}\cdot v_{ij} - r_i^{\mathbf{v}}\right).$$

We next present the main theorem that *ROM* gives worst case revenue of at least Z^*, the worst case optimal revenue computed by the optimization problem (37).

Theorem 5 ROM ***(a)*** *is budget feasible,* **(b)** *is individually rational and* **(c)** *achieves a worst case revenue of at least* Z^*.

4.3 Solving *ROM*

The computationally intensive step in *ROM* involves solving the bilinear optimization problems (44). Bilinear problems are NP-Hard ([65]) for general uncertainty sets $\mathcal{U}$. However, if the uncertainty set $\mathcal{U}$ has a polynomial number of extreme points, then we can obtain a polynomial time algorithm that solves (44). This follows from Proposition 4, which states that there exists an optimal solution to these problems. Thus, we can solve the problems (44) in polynomial time, by simply enumerating all the extreme points.

Proposition 4 *There exists an optimal solution to Problems (44) that is an extreme point of* $\mathcal{U}$.

We next describe an algorithm to solve the bilinear optimization problem (44). This algorithm, motivated from the Generalized Benders Decomposition algorithm presented in [14], is presented in Algorithm 4.

Algorithm 4 ***Generalized Benders decomposition algorithm for Problem (44).***
Input: *Problem (44), accuracy parameter* ϵ.
Output: *Approximate optimal solution* $\mathbf{z}$.
Algorithm:

1. *Set parameters* $UB = \infty,\ LB = 0,\ k = 0$.
2. *Compute* $v_1^0 = \min\limits_{\mathbf{v}\in\mathcal{U}} v_1$ *and for each* $i = 2, \ldots, |\mathcal{N}|$

$$v_i^0 = \min_{(v_1^0,\ldots,v_{i-1}^0,v_i,\ldots,v_n)\in\mathcal{U}} v_i.$$

3. *While* $UB - LB \geq \epsilon$,
   - **(a)** *Solve the inner linear maximization problem in (44) using* $\mathbf{v} = \mathbf{v}^k$. *Set* $\mathbf{x}^k$ *to be an optimal solution to this problem and update the value of* UB *to the value of of* $\mathbf{x}^k$.
   - **(b)** *Solve the outer linear minimization problem in (44) using* $\mathbf{x} = \mathbf{x}^k$. *Set* $\mathbf{v}^{k+1}$ *to be an optimal solution to this problem, and update the value of* LB *to the value of* $\mathbf{v}^{k+1}$.
   - **(c)** *Increment* k.
   - **(d)** *Add the constraint*

$$\sum_{i\in\mathcal{N}} p_i \leq UB$$

     to the inner maximization problem.
   - **(e)** *Add the constraint*

$$\sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{M}} x_{ij}v_{ij} \geq LB$$

     to the outer minimization problem.
4. *Output* $\mathbf{v}^k$.

4.4 Auctions without budget constraints

In this section, we consider a special case of the auction design problem in which the buyers do not have any budget constraints. In the absence of budget constraints, the auction design problem for multiple items reduces to the auction design problem for a single item. Consequently we consider the auction design problem for a single item without budget constraints. Myerson [57] solved this problem in a probabilistic setting for buyers with uncorrelated valuations and showed the optimal mechanism takes the form of a second price auction with a reservation price. We recover Myerson's result in a more general setting that allows correlated buyers and obtain an optimal mechanism that also takes the form of a second price auction with a reservation price.

The robust optimal mechanism for single item auctions without budget constraints

By specializing *ROM* to the case $B_i = \infty$, $\forall i \in \mathcal{N}$ and $|\mathcal{M}| = 1$, we derive the optimal mechanism for single item auctions without budget constraints, that we will refer to as *ROM-Si*. *ROM-Si* consists of Algorithms 5 and 6, respectively.

Algorithm 5 Calculation of the *reservation price.*

Input: *Uncertainty set* $\mathcal{U}$.

Output: *Reservation price* r^*.
Algorithm:

1. *Compute the* worst case valuation vector $\mathbf{z} = \{z_i\}_{i\in\mathcal{N}}$ *given by*

$$\mathbf{z} = \arg\min_{\mathbf{v}\in\mathcal{U}} \left\{ \begin{array}{lll} \max\limits_{\{x_i\}_{i\in\mathcal{N}}} & \sum\limits_{i\in\mathcal{N}} x_i \cdot v_i & \\ s.t. & \sum\limits_{i\in\mathcal{N}} x_i \leq 1, & \\ & x_i \geq 0, & \forall i \in \mathcal{N}. \end{array} \right\}$$

2. *Compute* reservation price r^* *given by*

$$r^* = \arg \left\{ \begin{array}{ll} \min\limits_{r} & r \\ s.t. & r \geq z_i, \quad \forall i \in \mathcal{N}. \end{array} \right\}$$

Algorithm 6 ***Calculation of*** **allocations** ***and*** **payments.**
Input: *Bid vector* $\mathbf{v} = \{v_i\}_{i\in\mathcal{N}}, r^*$.
Output: *Allocation vector* $\left\{a_k^{\mathbf{v}}\right\}_{k\in\mathcal{N}}$ *and the payments* $\left\{p_k^{\mathbf{v}}\right\}_{k\in\mathcal{N}}$.
Algorithm:

1. *Calculate the quantities* $\left(\left\{y_i^{\mathbf{v}}, r_i^{\mathbf{v}}\right\}_{i\in\mathcal{N}}, \left\{y_{i,k}^{\mathbf{v}_{-k}}, r_{i,k}^{\mathbf{v}_{-k}}\right\}_{i\in\mathcal{N}\setminus\{k\}}\right)$ *given by*

$$\begin{aligned}\left\{y_i^{\mathbf{v}}, r_i^{\mathbf{v}}\right\}_{i\in\mathcal{N}} &= \arg\max_{(\mathbf{y},\mathbf{r})\in\mathcal{P}} \sum_{i\in\mathcal{N}} y_i \cdot v_i - r_i,\\ \left\{y_{i,k}^{\mathbf{v}_{-k}}, r_{i,k}^{\mathbf{v}_{-k}}\right\}_{i\in\mathcal{N}\setminus\{k\}} &= \arg\max_{(\mathbf{y},\mathbf{r})\in\mathcal{P}} \sum_{i\in\mathcal{N}\setminus\{k\}} y_i \cdot v_i - r_i,\end{aligned}$$

where

$$\mathcal{P} = \left\{ \{x_i, r_i\}_{i\in\mathcal{N}} \;\middle|\; \begin{array}{l} \sum_{i\in\mathcal{N}} x_i \le 1, \\ \sum_{i\in\mathcal{N}} r_i \ge R^*, \\ x_i \ge 0, \quad \forall i \in \mathcal{N}, \end{array} \right\}.$$

2. *Compute the allocation vector* $\left\{a_k^{\mathbf{v}}\right\}_{k\in\mathcal{N}}$ *and the payments* $\left\{p_k^{\mathbf{v}}\right\}_{k\in\mathcal{N}}$ *as follows*

$$\begin{aligned} a_k^{\mathbf{v}} &= y_k^{\mathbf{v}},\\ p_k^{\mathbf{v}} &= r_k^{\mathbf{v}} + \sum_{i\in\mathcal{N}\setminus\{k\}} \left(y_{i,k}^{\mathbf{v}_{-k}} v_i - r_{i,k}^{\mathbf{v}_{-k}}\right) - \sum_{i\in\mathcal{N}\setminus\{k\}} \left(y_i^{\mathbf{v}} v_i - r_i^{\mathbf{v}}\right). \end{aligned}$$

Comparison with the Myerson auction

ROM-Si and the Myerson auction have the same structure, that of a second price auction with a reservation price. However, the mechanisms differ in the way they calculate the reservation prices. In the case of the Myerson auction the reservation price is calculated by solving a non-linear equation

$$\frac{1 - F(r)}{f(r)} = r, \tag{47}$$

where $F(\cdot)$ is the cdf and $f(\cdot)$ is the pdf of the probability distribution that the auctioneer assumes the valuations are sampled from. On the other hand, in *ROM-Si*, the reservation price is calculated using the linear optimization problem

$$\begin{aligned} \min_{r,\mathbf{v}} \quad & r \\ \text{s.t.} \quad & r \geq v_i, \ \forall i \in \mathcal{N}, \\ & (v_1, v_2, \ldots, v_n) \in \mathcal{U}. \end{aligned} \tag{48}$$

In this section, we compare *ROM-Si* and the Myerson auction with respect to the following aspects:

(a) Computational complexity The computationally intensive step in both *ROM-Si* and the Myerson auction is the calculation of the reservation price. Once the reservation price is calculated, both these mechanisms solve linear optimization problems to carry out the auction. While the Myerson auction solves the non-linear equation (47), *ROM-Si* solves the optimization problem (48) to calculate the reservation price. As long as the uncertainty set $\mathcal{U}$ is polyhedral (for example given as in Eqs. (4),(10)), this optimization problem is efficiently solvable. In particular, when $\mathcal{U}$ is a polyhedron, the optimization problem reduces to a linear optimization problem.

(b) Robustness to mis-specification The values of the reservation prices obtained by *ROM-Si* and by the Myerson auction differ in general. For example, when we use the uncertainty set $\mathcal{U}^{\text{CLT}}$ with parameters μ and σ, ROM-Si gives us a single value for the reserve price of

$$\mu - \frac{\Gamma \sigma}{\sqrt{n}} \approx \mu \text{ (for large n).}$$

On the other hand, the Myerson auction gives different values of reservation prices for different distributions. For uniform and exponential distributions, the reservation prices obtained by the auctions match, while for other distributions, the reservation prices are different. This dependence of reservation prices on the distribution may lead to lack of robustness on the part of the Myerson auction, when the assumed distribution differs from the realized distribution.

In order to study this, we design the following experiments. We first assume that valuations are normally distributed with parameters $\mu = 1$, and $\sigma = 0.5, 1, 2$ and carry out *ROM-Si* and the Myerson auctions with these parameters. Then, we investigate how these auctions compare with each other, when the realized distributions are different from the assumed distributions. This is done by computing the quantity *Relative Revenue* defined as

$$\text{Relative Revenue } = \frac{\text{ROM-Single Revenue} - \text{Myerson Revenue}}{\text{Myerson Revenue}}, \tag{49}$$

which when positive, indicates that the proposed auction results in a greater revenue than the Myerson auction.

In Table 7, we compare the expected revenues (obtained by simulation) of the *ROM-Si* and of the Myerson auction, when the realized distribution is Gamma, Beta,

Table 7 Myerson versus *ROM-Si*: the *relative revenue*, defined in (49) for different distributions with the same mean and standard deviation

Distribution	(μ, σ)		
	(1, 0.5)	(1, 1)	(1, 2)
Gamma	0.529	0.696	1.038
Beta	0.387	0.507	0.799
Triangle	0.271	0.376	0.526
Uniform	0.498	0.697	0.959

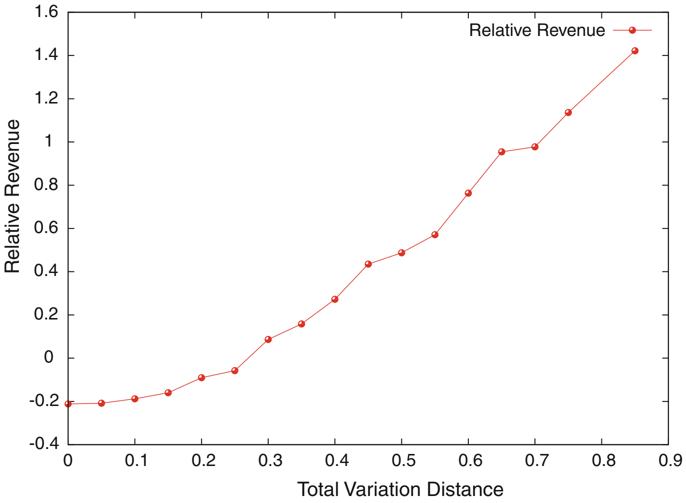

Fig. 2 Robustness of *ROM-Si*

Uniform and Triangle with the same mean and standard deviation. We find that the proposed approach has very significant benefits with revenue improvements in the range of [27 %, 103 %]. To amplify this further, we perform another experiment where we vary the distributions at a slower pace. In particular, we consider a series of distributions $\mathcal{F}$ that are increasingly different from $N(1, 0.5)$ with respect to the value of *total variation distance*. The *total variation distance* between probability measures $\mathcal{F}_1$ and $\mathcal{F}_2$ is defined as the largest possible difference between the probabilities that $\mathcal{F}_1$ and $\mathcal{F}_2$ can assign to the same event, that is,

$$\|\mathcal{F}_1 - \mathcal{F}_2\|_{TV} = \sup_{A \in \Omega} |\mathcal{F}_1(A) - \mathcal{F}_2(A)| .$$

We plot the Relative Revenue against the values of *total variation distances* in Fig. 2. We observe that *ROM-Si* performs better than the Myerson auction when the total variation distance becomes larger than 0.22.

In Table 8, we compare the expected revenues of the *ROM-Si* and of the Myerson auction when the distribution is still normal with the same mean but with different standard deviations. We find that the proposed approach still has potentially significant benefits in the range of [2.8 %, 54 %]. In Table 9, we investigate the situation when

Table 8 Myerson versus *ROM-Si*: the *relative revenue* under the same mean but different standard deviations

Standard deviation	(μ, σ)		
	(1, 0.5)	(1, 1)	(1, 2)
$\sigma/4$	0.108	0.134	0.196
$\sigma/2$	0.0357	0.042	0.068
$3\sigma/2$	0.0282	0.039	0.062
2σ	0.141	0.187	0.261
5σ	0.247	0.334	0.542

Table 9 Myerson versus *ROM-Si*: the *relative revenue* under the same standard deviation but different means

Mean	(μ, σ)		
	(1, 0.5)	(1, 1)	(1, 2)
$\mu/4$	0.178	0.22	0.335
$\mu/4$	0.053	0.064	0.09
2μ	−0.176	−0.242	−0.392
3μ	−0.312	−0.416	−0.58

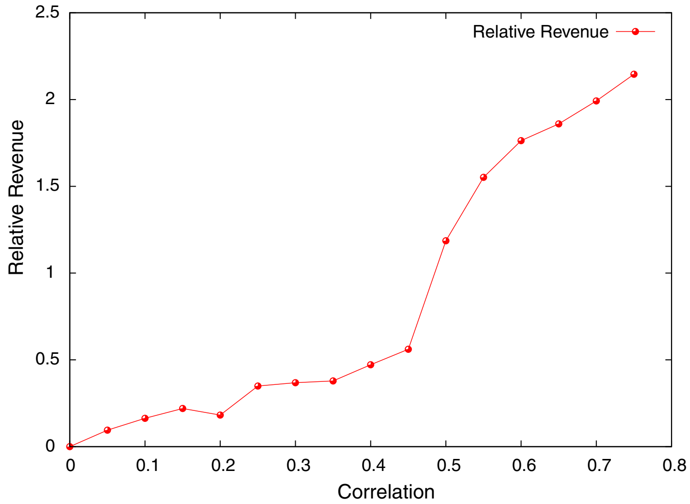

Fig. 3 Effect of correlations in valuations on the auction revenue

the realized distribution is again normal with the same standard deviation but with different means. We find that the results in this case are mixed. When the realized mean is smaller (larger) than the assumed mean, the proposed approach outperforms (underperforms) when compared with the Myerson auction. In summary, we feel that *ROM-Si* has stronger robustness properties when the distribution or the standard deviation is misspecified.

(c) Effect of correlations on the revenue In this experiment, we demonstrate how incorporating correlation information can lead to revenue gains. We compare the revenue of the Myerson auction, which ignores any correlation information, with the revenue obtained by using *ROM* with the uncertainty set $\mathcal{U}^{Corr}$ [see Eq. (4)]. We compute the *Relative Revenue* as defined in (49) and plot it against the value of correlation in Fig. 3. We observe that there are significant benefits in incorporating correlation information, which *ROM* allows us to do.

5 Pricing multi-dimensional options

The problem of pricing and hedging derivative securities has been one of the most well studied problems in Financial Economics. The most important breakthrough, to this date, has been the celebrated Black and Scholes [21], and Merton [56] option-pricing formula, which is based on the principle of dynamic replication that allows one to look for a portfolio of simpler securities which is self-financing and whose value at the end of the time horizon matches the payoff of the option. Such a portfolio of simpler securities is known as a replicating portfolio, and the value of this portfolio at the beginning is the no-arbitrage price of the option. For example, under the assumption of geometric Brownian motion for price dynamics, Black and Scholes [21] shows that a European call-option on a stock can be replicated exactly by a dynamic-hedging strategy involving only stocks and risk-free borrowing and lending. This replication allowed them to give the first closed-form expression for the price of a European Option.

The Black–Scholes formula, in spite of its popularity, has some well-known deficiencies. Empirically, the model prices appear to differ from market prices in certain systematic ways. These discrepancies are usually ascribed to the strong assumption that the underlying security follows a stationary geometric Brownian motion. Apart from the problems with the price-dynamics, there are other factors that arise mostly due to transaction costs and liquidity issues, that make it intractable for one to look for an exact replicating portfolio. This suggests that we consider the natural trade-off between exactness and tractability and to look for nearly exact replications, which can be found in a tractable manner. Motivated by the inability to exactly replicate different types of options (American style options, among others), [17] have proposed the idea of ϵ-arbitrage pricing using dynamic programming under which we seek a self-financing dynamic portfolio strategy that most closely approximates the payoff of an option.

In this section, we summarize the major results of Bandi et al. [4] and propose to model the underlying price dynamics with uncertainty sets, and then apply robust optimization as opposed to dynamic programming to solve the ϵ-arbitrage problem. We use the ℓ_1-norm to measure the error in replication which when combined with polyhedral uncertainty sets results in linear optimization problems. This approach also allows us to easily model transaction costs, very general pricing dynamics, accommodate high dimensional problems that today can only be handled by simulation methods.

5.1 The option pricing problem

An option is a contract defined on a set of predetermined underlying securities, and is associated with a payoff function. The payoff function determines the value of the option after the realization of random returns of the underlying securities. The option pricing problem refers to the problem of calculating the "value" of an option before the realization of the random returns. The payoff function, denoted here by

$$P\left(\left\{\widetilde{S}_\tau^1, \widetilde{S}_\tau^2, \ldots, \widetilde{S}_\tau^M\right\}, \{K_1, K_2, \ldots, K_r\}\right),$$

depends on

(a) $\left\{\widetilde{S}_\tau^1, \widetilde{S}_\tau^2, \ldots, \widetilde{S}_\tau^M\right\}$: vector of prices of the set of M underlying securities at time τ.
(b) τ : time at which the option is exercised.
(c) $\{K_1, K_2, \ldots, K_r\}$: a set of other parameters (e.g. strike price, dividend etc.).

For example, a European Call option's payoff is defined to be $(\widetilde{S}_T - K)^+ = \max\{\widetilde{S}_T - K, 0\}$, where $\widetilde{S}_T$ denotes the price of the underlying security at the time of expiry T, and K denotes the strike price.

To determine the value of the option before the realization of the random returns, we seek to obtain a "replicating portfolio" to capture the payoff dynamics of the option. We construct this portfolio out of the set of stocks and a risk free asset.

5.2 The underlying primitive for price dynamics

We consider a discrete model of price movements where the price of the stock changes at discrete points of time. Let $\widetilde{r}_t^S$ be the return at t; i.e., the return from period $[t, t+1)$. Applying the central limit theorem to the random variables $\left\{\widetilde{r}_1^S, \widetilde{r}_2^S, \ldots, \widetilde{r}_\tau^S\right\}$, we conclude that

$$\frac{\sum_{i=1}^{\tau} \log\left(1+\widetilde{r}_i^S\right) - \tau \cdot \mu_{\log}}{\sigma_{\log} \cdot \sqrt{\tau}}$$

obeys a standard normal distribution as $\tau \to \infty$, where $\mu_{\log}, \ \sigma_{\log}$ are mean and standard deviation of $\log\left(1+\widetilde{r}_i^S\right)$, respectively. We assume as a primitive that the stock returns obey

$$\left|\frac{\log \widetilde{R}_\tau^S - \tau \cdot \mu_{\log}}{\sigma_{\log} \cdot \sqrt{\tau}}\right| \leq \Gamma_\tau \qquad \forall \tau \geq \tau_0, \tag{50}$$

where $\widetilde{R}_\tau^S = \prod_{i=1}^{\tau}(1+\widetilde{r}_i^S)$, is the cumulative return up to time τ, and τ_0 is chosen such that Eq. (50) holds with high probability. A typical value of $\tau_0 = 30$ is chosen. The parameter Γ_τ can be seen to represent the risk averseness in this context. Note that Eq. (50) is equivalent to

$$e^{\tau\mu_{\log} - \Gamma_\tau\sqrt{\tau}\sigma_{\log}} \leq \widetilde{R}_\tau^S \leq e^{\tau\mu_{\log} + \Gamma_\tau\sqrt{\tau}\sigma_{\log}}, \qquad \forall \tau, \tag{51}$$

which defines a box uncertainty set for the cumulative returns. In addition, we assume some bounds on the single period return $\widetilde{r}_\tau^S$ and since $\left(1+\widetilde{r}_\tau^S\right) = \widetilde{R}_\tau^S / \widetilde{R}_{\tau-1}^S$, we have:

$$\mu_r - \Gamma_\tau \sigma_r \leq \frac{\widetilde{R}_\tau^S}{\widetilde{R}_{\tau-1}^S} \leq \mu_r + \Gamma_\tau \sigma_r, \qquad \forall \tau. \tag{52}$$

In summary, we assume that the cumulative stock returns belong to the following uncertainty set (a polytope) defined by Eqs. (51)–(52)

$$\mathcal{U}^S = \left\{ \widetilde{R}_t^S \;\middle|\; \begin{array}{ll} \underline{R_t^S} \leq \widetilde{R}_t^S \leq \overline{R_t^S}, & \forall t = 1, \ldots, T \\ \underline{r_t^S} \cdot \widetilde{R}_{t-1}^S \leq \widetilde{R}_t^S \leq \overline{r_t^S} \cdot \widetilde{R}_{t-1}^S, & \forall t = 1, \ldots, T \end{array} \right\}, \tag{53}$$

where

$$\underline{R_t^S} = e^{t \cdot \mu_{\log} - \Gamma \cdot \sqrt{t} \cdot \sigma_{\log}}, \; \overline{R_t^S} = e^{t \cdot \mu_{\log} + \Gamma \cdot \sqrt{t} \cdot \sigma_{\log}}, \; \underline{r_t^S} = \mu_r - \Gamma_t \cdot \sigma_r, \; \overline{r_t^S} = \mu_r + \Gamma_t \cdot \sigma_r,$$
$$\underline{R_{t,\tau}^S} = (t-\tau) \cdot \mu_r - \Gamma_t \cdot \sigma_r \cdot \sqrt{t-\tau}, \quad \text{and} \quad \overline{R_{t,\tau}^S} = (t-\tau) \cdot \mu_r + \Gamma_t \cdot \sigma_r \cdot \sqrt{t-\tau}.$$

The values of μ, σ, $\mu_{\log}$, $\sigma_{\log}$ can be obtained from empirical data on single period returns.

Note that the uncertainty sets are defined using the cumulative returns $\widetilde{R}_t^S$ instead of single period return $\widetilde{r}_t^S$. This is mainly to avoid being overly conservative in each period, instead the uncertainty set $\mathcal{U}^S$ allows the price to change by a potentially large amount in a single period but forces the change to average out in a longer period of time. As observed in [4], this choice of modeling the cumulative returns leads to linear optimization formulations for the option pricing problem for many kinds of options.

5.3 Option pricing problem as an optimization problem

The idea of our approach is to find a replicating portfolio that consists of the underlying stock $\mathbb{S}$ and a risk-free asset $\mathbb{B}$ so that the value of this portfolio at the time of exercise matches the payoff of the option as closely as possible. We refer to the difference as the replication error and can be seen as a form of arbitrage. It is given by:

$$\left| P\left(S_T, K\right) - W_T \right|,$$

where W_T is the value of the portfolio at the time of exercise T. In a robust optimization setting, our goal is to find a portfolio that minimizes the worst case replication error (denoted by ϵ), between the portfolio wealth and the option payoff, over all possible returns that lie in a predetermined uncertainty set $\mathcal{U}^S$ defined in Section 6.2. The optimal portfolio thus obtained, will have payoff that is within $\pm\epsilon$ from the actual option payoff for all possible realizations of the returns lying in $\mathcal{U}^S$. The price of the option would thus be the initial value of this replicating portfolio. Note that, as $T \to \infty$, assuming complete markets, we are guaranteed to have a replication error of 0. For finite T, we obtain non-zero replication errors but are often close to zero, see [16] for a thorough discussion.

The associated optimization problem can be represented as follows

$$\begin{aligned} \min_{\{x_t^S, x_t^B, y_t\}} \max_{\{\widetilde{R}_t^S \in \mathcal{U}^S\}} \quad & \left| P\left(S_T, K\right) - W_T \right| \\ \text{s.t.} \quad & W_T = x_T^S + x_T^B, \\ & x_t^S = \left(1 + \widetilde{r}_{t-1}^S\right)\left(x_{t-1}^S + y_{t-1}\right), \quad \forall t = 1, \ldots, T, \\ & x_t^B = \left(1 + r_{t-1}^B\right)\left(x_{t-1}^B - y_{t-1}\right), \quad \forall t = 1, \ldots, T, \end{aligned} \tag{54}$$

where x_t^S is the amount invested in the underlying security, x_t^B is the amount invested in the risk-less asset, and y_t is the amount traded from the underlying security to the risk-less asset during the period $[t, t+1)$. In the optimization problem (54), we seek to minimize the worst case replication error. After finding the portfolio, the price of the option would then be given by $x_0^S + x_0^B$, which is the value of the portfolio at time $t = 0$.

5.4 Pricing a European option: an illustration

In this section, we present our approach in the context of a European call option. A European call option gives the option holder the right to buy the stock at a predetermined price K, at T. Hence, the payoff function for a European Call is $P\left(\widetilde{S}_T, K\right) = \left(\widetilde{S}_T - K\right)^+$.

Using the same set of decision variables and data as in (54) and with payoff function $P\left(\widetilde{S}_T, K\right) = \left(\widetilde{S}_T - K\right)^+$, the resulting optimization problem becomes

$$\begin{aligned} \min_{\{x_t^S, x_t^B, y_t\}} \max_{\{\widetilde{R}_t^S \in \mathcal{U}^S\}} \quad & \left| \left(\widetilde{S}_T - K\right)^+ - W_T \right| \\ \text{s.t.} \quad & W_T = x_T^S + x_T^B, \\ & x_t^S = \left(1 + \widetilde{r}_{t-1}^S\right)\left(x_{t-1}^S + y_{t-1}\right), \quad \forall t = 1, \ldots, T, \\ & x_t^B = \left(1 + r_{t-1}^B\right)\left(x_{t-1}^B - y_{t-1}\right), \quad \forall t = 1, \ldots, T. \end{aligned}$$

Let α_t^S, α_t^B, β_t be given by

$$\alpha_t^S = \frac{x_t^S}{\widetilde{R}_t^S}, \; \alpha_t^B = \frac{x_t^B}{R_t^B}, \; \beta_t = \frac{y_t}{\widetilde{R}_t^S}, \; \text{where } \widetilde{R}_t^S = \prod_{i=0}^{t-1}\left(1 + \widetilde{r}_i^S\right), \; R_t^B = \prod_{i=0}^{t-1}\left(1 + r_i^B\right).$$

Using these variables, we obtain the following equivalent formulation:

$$\begin{aligned} \min_{\{\alpha_t^S,\alpha_t^B,\beta_t\}} \max_{\{\widetilde{R}_t^S\in\mathcal{U}^S\}} & \left|\left(S_0\widetilde{R}_T^S-K\right)^+ - \left(\widetilde{R}_T^S\alpha_T^S + R_T^B\alpha_T^B\right)\right| \\ \text{s.t.} \quad & \alpha_t^S = \alpha_{t-1}^S + \beta_{t-1}, \quad \forall t=1,\ldots,T, \\ & \alpha_t^B = \alpha_{t-1}^B - \beta_{t-1}\frac{\widetilde{R}_{t-1}^S}{R_{t-1}^B}, \ \forall t=1,\ldots,T. \end{aligned} \tag{55}$$

Note that the variables $\alpha_t^S,\ \beta_t$ depend on the uncertain parameters $\widetilde{R}_t^S$, and α_t^B depends on R_t^B. However, we are only interested in $\alpha_0^S,\ \beta_0,\ \alpha_0^B$ as far as the strategy we will implement. Since $\alpha_0^S = x_0^S$, $\beta_0 = y_0$, $\alpha_0^B = x_0^B$, after solving problem (55), we only implement $\alpha_0^S,\ \beta_0,\ \alpha_0^B$ which are well defined. Moreover, the price of the option given by $x_0^S + x_0^B = \alpha_0^S + \alpha_0^B$ is also well defined.

Substituting all intermediate $\alpha_t^B,\ \alpha_t^S$, we obtain

$$\begin{aligned} \min_{\{\alpha_t^S,\alpha_t^B,\beta_t\}} \max_{\{\widetilde{R}_t^S\in\mathcal{U}^S\}} & \left|\left(S_0\widetilde{R}_T^S-K\right)^+ - \left(\alpha_0^S+\sum_{t=1}^T\beta_{t-1}\right)\widetilde{R}_T^S\right. \\ & \left. - \alpha_0^B R_T^B + \sum_{t=0}^{T-1}\beta_t\frac{R_T^B}{R_t^B}\widetilde{R}_t^S\right|. \end{aligned} \tag{56}$$

We next describe the steps involved in obtaining a linear formulation that is equivalent to (56). Let

$$\delta = \max_{\{\widetilde{R}_t^S\in\mathcal{U}^S\}} \left|\left(S_0\widetilde{R}_T^S-K\right)^+ - \left(\alpha_0^S+\sum_{t=1}^T\beta_{t-1}\right)\widetilde{R}_T^S - \alpha_0^B R_T^B + \sum_{t=0}^{T-1}\beta_t\frac{R_T^B}{R_t^B}\widetilde{R}_t^S\right|,$$

and note that δ is the optimal solution of the problem

$$\begin{aligned} \min \quad & \kappa \\ \text{s.t.} \quad & \kappa \ge \left(S_0\widetilde{R}_T^S-K\right)^+ - \left(\alpha_0^S+\sum_{t=1}^T\beta_{t-1}\right)\widetilde{R}_T^S - \alpha_0^B R_T^B \\ & \qquad + \sum_{t=1}^T\beta_{t-1}\frac{R_T^{\mathrm{B}}}{R_t^B}\widetilde{R}_t^S, \quad \forall \tilde{R}_t^S\in\mathbb{U}^1, \\ & \kappa \ge -\left(S_0\widetilde{R}_T^S-K\right)^+ + \left(\alpha_0^S+\sum_{t=1}^T\beta_{t-1}\right)\widetilde{R}_T^S + \alpha_0^B R_T^B \\ & \qquad - \sum_{t=1}^T\beta_{t-1}\frac{R_T^{\mathrm{B}}}{R_t^B}\widetilde{R}_t^S, \quad \forall \tilde{R}_t^S\in\mathbb{U}^1, \end{aligned}$$

Moreover, observing that

$$\left(S_0\widetilde{R}_T^S - K\right)^+ = \begin{cases} 0, & \forall \widetilde{R}_T^S \le \frac{K}{S_0}, \\ S_0\widetilde{R}_T^S - K, & \forall \widetilde{R}_T^S \ge \frac{K}{S_0}, \end{cases}$$

we partition the uncertainty set $\mathcal{U}$ according to whether $\widetilde{R}_T^S \ge \dfrac{K}{S_0}$. Let

$$\begin{aligned} \mathcal{U}_a^S &= \mathcal{U}^S \cap \left\{ \widetilde{R}_T^S \ge \frac{K}{S_0} \right\}, \\ \mathcal{U}_b^S &= \mathcal{U}^S \cap \left\{ \widetilde{R}_T^S \le \frac{K}{S_0} \right\}. \end{aligned}$$

Using this partition, we next obtain another equivalent formulation of (56):

$$\begin{aligned} \min_{\{\alpha_0^S, \alpha_0^B, \beta_t\}} \quad & \epsilon \\ \text{s.t.} \quad & \\ & \epsilon \ge \left(S_0\widetilde{R}_T^S - K\right) - \left(\alpha_0^S + \sum_{t=1}^{T} \beta_{t-1}\right)\widetilde{R}_T^S \\ & \quad -\alpha_0^B R_T^B + \sum_{t=1}^{T} \beta_{t-1} \frac{R_T^{\mathrm{B}}}{R_t^B} \widetilde{R}_t^S, \quad \forall \tilde{R}_t^S \in \mathbb{U}_a^1, \\ & \epsilon \ge -\left(S_0\widetilde{R}_T^S - K\right) + \left(\alpha_0^S + \sum_{t=1}^{T} \beta_{t-1}\right)\widetilde{R}_T^S \\ & \quad +\alpha_0^B R_T^B - \sum_{t=1}^{T} \beta_{t-1} \frac{R_T^{\mathrm{B}}}{R_t^B} \widetilde{R}_t^S, \quad \forall \tilde{R}_t^S \in \mathbb{U}_a^1, \\ & \epsilon \ge -\left(\alpha_0^S + \sum_{t=1}^{T} \beta_{t-1}\right)\widetilde{R}_T^S - \alpha_0^B R_T^B + \sum_{t=1}^{T} \beta_{t-1} \frac{R_T^{\mathrm{B}}}{R_t^B} \widetilde{R}_t^S, \quad \forall \tilde{R}_t^S \in \mathbb{U}_b^1, \\ & \epsilon \ge \left(\alpha_0^S + \sum_{t=1}^{T} \beta_{t-1}\right)\widetilde{R}_T^S + \alpha_0^B R_T^B - \sum_{t=1}^{T} \beta_{t-1} \frac{R_T^{\mathrm{B}}}{R_t^B} \widetilde{R}_t^S, \quad \forall \tilde{R}_t^S \in \mathbb{U}_b^1. \end{aligned}$$

Since the uncertainty set $\mathcal{U}^S$ is a polyhedron, the above formulation leads to an equivalent linear optimization problem. In particular, Bandi et al. [4] shows that the size of the linear optimization problem scales linearly with T.

5.5 Modeling flexibility

In this section, we discuss the other main advantage of this framework (other than computational tractability)—modeling flexibility. Apart from the ability to model transaction costs and liquidity effects, which comes from our use of linear optimization as the main tool, we also show how our framework allows us to capture the implied volatility smile, defined below, in an intuitive way. This is achieved by adjusting the parameter Γ in order to account for the risk attitude of the investor.

Risk aversion as an explanation for the implied volatility smile As the Black–Scholes model became popular, many people started using the model to calculate the volatilities in the market from the market prices observed. This quantity known as the *implied volatility* of an option is simply that volatility that makes the model price exactly equal to the observed market price. Each option has a unique implied volatility, and traders started to quote options in terms of implied volatilities. The main reason is that as the underlying asset price changes through the day, the implied volatility does not have to be adjusted as much as the option prices, which change all the time. The implied volatilities started to appear as fundamental quantities associated with an option.

When the implied volatilities are plotted across strike prices for options with the same time to expiration and on the same underlying stock, it was observed that these plots exhibit *smiles* or *smirks*. According to the Black–Scholes formula, the plot should be a flat line because only one volatility parameter governs the underlying stochastic process on which all options are priced. The same holds for European-style options on the U.S. S&P 500 Index, which were flat from the start of their exchange-based trading in April 1986 until the U.S. stock market crash of October 1987. After the crash, however, volatility smiles became skewed; that is, volatility smiles became downward sloping as the strike price increased. Other markets also often exhibit volatility smiles.

Many explanations have been offered to explain the downward sloping and the U-shaped volatility smiles. As noted by Taylor [66], there is no economic intuition behind many of these explanations. In particular, Taylor critiques the stochastic volatility models and suggests that it does not capture the true dynamics of the smile. We believe that risk attitude of an investor may be a possible explanation for the existence of the smiles. This is supported from the widely noted empirical observation that the phenomenon of volatility smile started appearing after the crash of 1987.

One of the key features of our pricing methodology is the ability to capture the risk attitude of an option writer, when pricing an option. This is achieved with the help of the parameter Γ which characterizes the uncertainty set that is used for pricing. A lower value for Γ implies that the user is willing to take higher risk by ignoring the variability of stock prices. On the other hand, a higher value of Γ indicates that the user seeks a price that will allow him to replicate the payoff of the option for a larger range of stock prices. Therefore, Γ becomes a natural way to express one's risk aversion.

In tune with the concept of implied volatility, we define the quantity *Implied Coefficient of Risk Aversion* $\left(\Gamma_{\text{implied}}\right)$, as the value of the parameter Γ, which leads to a model price that matches the market price. We observe, from our experiments,

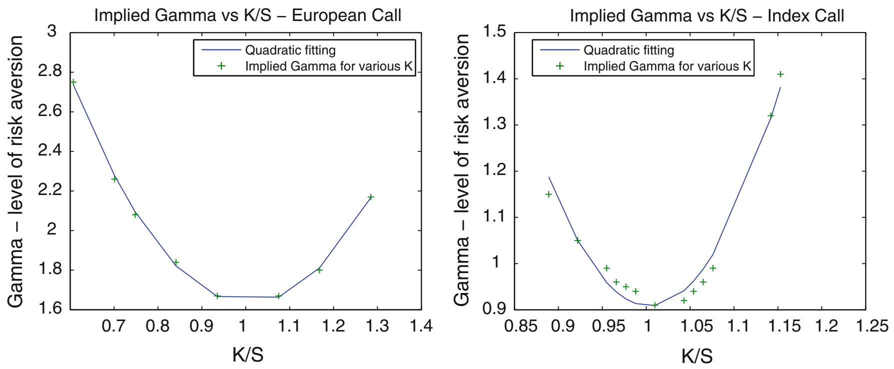

Fig. 4 Quadratic nature of coefficients of risk aversion—European simple call and European Index call

that Γ_{implied} behaves a lot like the implied volatility. When plotted against the strike prices, it displays a U-shaped behavior and downward sloping. This observation can be explained by simply recalling the meaning of the parameter Γ which stands for the risk aversion of the option writer.

In particular, we observe, from our experiments, that Γ_{implied} varies in a near-quadratic manner with K. When we model this quadratic dependence, we observed that the vertex of the parabola lies very close to the spot price S_0. This suggests that as K moves away from S_0, Γ_{implied} increases indicating the increase of risk aversion of the option writer as K moves away from S_0.

In the experiments, we show empirically that a quadratic variation of Γ_{implied} with K/S_0 would be adequate to characterize the risk aversion of an investor towards different strike prices. We use the following function to describe the relationship:

$$\Gamma(K) = \theta_0 + \theta_1 \frac{K - S_0}{S_0} + \theta_2 \left(\frac{K - S_0}{S_0} \right)^2, \ \theta_2 \geq 0. \tag{57}$$

The quantity $(K - S_0)/S_0$ captures the distance between the strike and the spot price and is also called as *moneyness* in the literature. We use a quadratic regression model to compute the coefficients $\{\theta_0, \theta_1, \theta_2\}$ so that the prices obtained using these Γ's match the market prices of our training set of strikes. We then use the resulting quadratic model $\Gamma(K)$ to calculate Γ and input it to yield the price for options with other strike prices (Fig. 4).

Modeling transaction costs and other market constraints Our framework also has the ability to model transaction costs and other market constraints. When we account for the transaction costs, the optimal replication problem (54) changes to

$$\begin{aligned} \min_{\{x_t^S, x_t^B, y_t\}} \max_{\{\widetilde{R}_t^S \in \mathcal{U}^S\}} \quad & \left| P(S_T, K) - \left(x_T^S + x_T^B\right) \right| \\ \text{s.t.} \quad & x_t^S = \left(1 + \widetilde{r}_{t-1}^S\right)\left(x_{t-1}^S - u_{t-1} + v_{t-1}\right), \quad \forall t = 1, \ldots, T, \\ & x_t^B = \left(1 + r_{t-1}^B\right)\left(x_{t-1}^B + (1 - c_{sell})\, u_{t-1} - \left(1 + c_{buy}\right) v_{t-1}\right), \forall t, \end{aligned}$$

where u_t is the amount removed from the stock and v_t is the amount added to the stock during the time period $[t, t+1)$. The parameters c_{sell}, c_{buy} represent the transaction costs. Introduce variables defined by

$$\alpha_t^S = \frac{x_t^S}{R_t^S}, \quad \alpha_t^B = \frac{x_t^B}{R_t^B}, \quad \beta_t^1 = \frac{u_t}{R_t^S}, \text{ and } \beta_t^2 = \frac{v_t}{R_t^S},$$

to obtain the following equivalent formulation

$$\begin{aligned}
\min_{\{\alpha_t^S, \alpha_t^B, \beta_t\}} \max_{\{\widetilde{R}_t^S \in \mathcal{U}^S\}} \quad & \left| P\left(S_T, K\right) - \left(\widetilde{R}_T^S \alpha_T^S + R_T^B \alpha_T^B\right) \right| \\
\text{s.t.} \quad & \alpha_t^S = \alpha_{t-1}^S + \beta_{t-1}^1 - \beta_{t-1}^2, \quad \forall t = 1, \ldots, T \\
& \alpha_t^B = \alpha_{t-1}^B + \left((1 - c_{\text{sell}})\, \beta_{t-1}^1 - \left(1 + c_{\text{buy}}\right) \beta_{t-1}^2\right) \frac{\widetilde{R}_{t-1}^S}{R_{t-1}^B}, \ \forall t.
\end{aligned}$$

Observing that

$$\begin{aligned}
\alpha_T^S &= \alpha_0^S + \sum_{t=1}^T \left(\beta_{t-1}^1 - \beta_{t-1}^2\right), \\
\alpha_T^B &= \alpha_0^B + \sum_{t=1}^T \left\{ \left((1 - c_{\text{sell}})\, \beta_{t-1}^1 - \left(1 + c_{\text{buy}}\right) \beta_{t-1}^2\right) \frac{\widetilde{R}_{t-1}^S}{R_{t-1}^B} \right\},
\end{aligned}$$

we obtain equivalent formulations which can then be formulated as linear optimization problems. Other market constraints such as limits on shorting and limits on the leverage ratio can also be modeled by linear constraints.

5.6 Computational results

In this section, we give examples on the accuracy of the method relative to prices that are observed in the market.

Comparison with market prices for American put options In the first experiment, we aim to price Microsoft (MSFT) 25 week American put options, with spot price $S_0 = \$24.8$ and strike prices in the range \$7.5–\$50. In the training stage, we compute $\Gamma_{\text{implied}}(K)$ for each of the options with strikes in the training set. Then we fit a quadratic function to these values, thus obtaining the coefficients of the quadratic function in (57). We report out of sample results in Table 10.

Comparison with market prices for European Index options In this experiment, we consider the 1/100 Dow Jones Industrial Average Index Options (DJIA). The underlying value of these options is based on the level of the Dow Jones Industrial Average, a price-weight stock market index calculated from the stock prices of thirty of the largest public companies in the US. We consider 8 week options with spot $S_0 = \$90.8$, for strike prices in the range \$74–\$105. We report out of sample results in Table 11.

Table 10 Price and replication error for American put options

No.	T	K/S	Γ_{implied}	Mkt price	Model price	Error	ϵ
Out of sample							
1	25	0.605	2.62	0.17	0.201	0.031	0.3982
2	25	0.806	1.83	0.695	0.589	−0.106	0.9996
3	25	0.968	1.6	1.895	1.764	−0.132	1.9487
4	25	1.008	1.59	2.365	2.266	−0.099	2.2109
5	25	1.21	1.9	5.85	5.939	0.089	3.5781
6	25	1.411	2.87	10.5	10.703	0.203	5.9713
7	25	1.512	3.63	12.975	13.2	0.225	6.832
8	25	1.815	7.7	20.45	20.303	−0.147	12.5842

Table 11 Price and replication error for European Index call options

No.	T	K/S	Γ_{implied}	Mkt price	Model price	Error	ϵ
Out of sample							
1	8	0.944	1.01	6	5.918	−0.082	2.002
2	8	0.999	0.92	2.69	2.647	−0.043	2.5717
3	8	1.021	0.91	1.75	1.762	0.012	2.6012
4	8	1.032	0.92	1.38	1.402	0.022	2.5615
5	8	1.098	1.07	0.225	0.254	0.029	1.9665
6	8	1.109	1.13	0.16	0.17	0.01	1.881

In these and other experiments, we have found that the errors between the model and the market prices for the single asset European options, and European style Index options are smaller than the errors obtained in other types of options. In general, when the option type is simpler (e.g. European or Asian) the replication error is small. The largest replication error occurs in the American put option where the price we produce tries to protect the holder against the worst choice for exercising the option, thus adding an additional layer of conservativeness.

6 Concluding remarks

We revisited some of the major successes of stochastic analysis in the twentieth century. In all these examples, together with considerable success, came challenges when researchers and practitioners desired to generalize the problems studied to multiple dimensions. It is our contention that stochastic analysis, based on the primitives of the Kolmogorov axioms and the concept of random variables was not intended to provide a tool for efficient computation. Rather it was intended to put the theory on a firm ground and give insights on the modeling of stochastic phenomena. In retrospect, given the historical developments and intentions of the originators, the computational challenges that stochastic analysis has faced, when attempting to solve problems in multiple dimensions, should have been anticipated.

Possibly because the development of modern optimization happened at about the same time as the development of the digital computer, optimization had from its very beginning efficient computation as its intention. Correspondingly, optimization has succeeded remarkably to solve problems in high dimensions. Given its success, it seems natural to us to apply optimization to solve problems of stochastic analysis in multiple dimensions. We also remark that in real world problems, what is available is data. Modeling stochastic phenomena with probability theory is a choice, that is probability distributions are not inherent to the problem. Given the computational difficulties in high dimensions, we feel we should consider alternative, computationally tractable approaches in high dimensions. We have proposed such an approach in this paper.

In all three examples we addressed in this paper, as well as others reported elsewhere ([2,5,7]), we have implemented the proposed approach and have included in the paper tables and figures with computational evidence in concrete examples in order to show that the approach of stochastic analysis based on optimization is capable of solving problems numerically in ways that, in our opinion, go beyond the current state of the art of stochastic analysis. The types of optimization problems that were required to be solved ranged from linear, discrete, and bilinear optimization problems. We anticipate that this research program, in addition to advancing stochastic analysis, will also advance optimization as it will reveal new optimization problems that need to be addressed.

Acknowledgments We thank the three reviewers of the paper for several insightful comments.

References

1. Ausubel, L.M.: An efficient ascending-bid auction for multiple objects. Am. Econ. Rev. **94**(5), 1452–1475 (2004)
2. Bandi, C., Bertsimas, D.: Network information theory via robust optimization. Working Paper (2011)
3. Bandi, C., Bertsimas, D.: Optimal design for multi-item auctions: a robust optimization approach. Working Paper (2012)
4. Bandi, C., Bertsimas, D., Chen, S.: Robust option pricing. Working Paper (2011)
5. Bandi, C., Bertsimas, D., Youssef, N.: Robust multi-class queueing theory. Working Paper (2012)
6. Bandi, C., Bertsimas, D., Youssef, N.: Robust queueing theory. Working Paper (2012)
7. Bandi, C., Bertsimas, D., Youssef, N.: Robust transient queueing theory. Working Paper (2012)
8. Ben-Tal, A., El-Ghaoui, L., Nemirovski, A.: Robust Optimization. Princeton University Press, Princeton (2009)
9. Ben-Tal, A., Nemirovski, A.: Robust convex optimization. Math. Oper. Res. **23**(4), 769–805 (1998)
10. Ben-Tal, A., Nemirovski, A.: Robust solutions of uncertain linear programs. Oper. Res. Lett. **25**, 1–13 (1999)
11. Ben-Tal, A., Nemirovski, A.: Robust solutions to uncertain programs. Oper. Res. Lett. **25**, 1–13 (1999)
12. Ben-Tal, A., Nemirovski, A.: Robust solutions of linear programming problems contaminated with uncertain data. Math. Program. **88**, 411–424 (2000)
13. Ben-Tal, A., Nemirovski, A.: On safe tractable approximations of chance-constrained linear matrix inequalities. Math. Oper. Res. **34**, 1–25 (2009)
14. Beran, E., Vandenberghe, L., Boyd, S.P.: A global bmi algorithm based on the generalized benders decomposition. In: Proceedings of the European Control Conference, pp. 1074–1082 (1997)
15. Bertsimas, D., Brown, D., Caramanis, C.: Theory and applications of robust optimization. SIAM Rev. **53**, 464–501 (2011)
16. Bertsimas, D., Kogan, L., Lo, A.W.: When is time continuous. J. Financ. Econ. **55**, 173–204 (2000)

17. Bertsimas, D., Kogan, L., Lo, A.W.: Hedging derivative securities and incomplete markets: an $\epsilon-$arbitrage approach. Oper. Res. **49**(7), 372–397 (2001)
18. Bertsimas, D., Sim, M.: Robust discrete optimization and network flows. Math. Program. **98**, 49–71 (2003)
19. Bertsimas, D., Sim, M.: The price of robustness. Oper. Res. **52**(1), 35–53 (2004)
20. Birge, J.R., Louveaux, F.: Introduction to Stochastic Programming. Springer Series in Operations Research, Springer, New York, NY (1997)
21. Black, F., Scholes, M.: Pricing of options and corporate liabilities. J. Polit. Econ. **81**, 637–654 (1973)
22. Borgs, C., Chayes, J.T., Immorlica, N., Mahdian, M., Saberi, A.: Multi-unit auctions with budget constrained bidders. In: ACM Conference on Electronic Commerce, pp. 44–51 (2005)
23. Bulow, J., Klemperer, P.: Auctions versus negotiations. Am. Econ. Rev. **86**, 180–194 (1996)
24. Burke, P.J.: The output of a queueing system. Oper. Res. **4**(6), 699–704 (1956)
25. Chawla, S., Hartline, J.D., Malec, D.L., Sivan, B.: Multi-parameter mechanism design and sequential posted pricing. In: STOC, pp. 311–320 (2010)
26. Che, Y., Gale, J.: The optimal mechanism for selling to a budget constrained buyer. J. Econ. Theory **92**(2), 198–233 (2000)
27. Cook, S.: The complexity of theorem-proving procedures. In: Conference Record of Third Annual ACM Symposium on Theory of Computing, vol. 1, pp. 151–158 (1971)
28. Cover, T.M., Thomas, J.A.: Elements of Information Theory. Wiley, New York (2006)
29. Cremer, J., McLean, R.P.: Full extraction of the surplus in bayesian and dominant strategy auctions. Econometrica **56**(6), 1247–1257 (1988)
30. Crovella, M.: The relationship between heavy-tailed file sizes and self-similar network traffic. In: INFORMS Applied Probability Conference (1997)
31. Dantzig, G.B.: Programming of interdependent activities: II mathematical model. Econometrica **17**, 200–211 (1949)
32. Dantzig, G.B.: Linear programming under uncertainty. Manag. Sci. **1**, 197–206 (1955)
33. Dantzig, G.B.: Linear Programming and Extensions. Princeton University Press and the RAND Corporation, Princeton (1963)
34. Dhangwatnotai, P., Roughgarden, T., Yan, Q.: Revenue maximization with a single sample. In: Proceedings of 12th ACM Conference on Electronic Commerce (2010)
35. Dobzinski, S., Lavi, R., Nisan, N.: Multi-unit auctions with budget limits. In: FOCS, pp. 260–269 (2008)
36. El-Ghaoui, L., Lebret, H.: Robust solutions to least-square problems to uncertain data matrices. SIAM J. Matrix Anal. Appl. **18**, 1035–1064 (1997)
37. El-Ghaoui, L., Oustry, F., Lebret, H.: Robust solutions to uncertain semidefinite programs. SIAM J. Optim. **9**, 33–52 (1998)
38. Erlang, A.K.: The theory of probabilities and telephone conversations. Nyt. Tidsskr. Mat. Ser. B. **20**, 33–39 (1909)
39. Gnedenko, B.V., Kolmogorov, A.N.: Limit Distributions for Sums of Independent Random Variables. Addison Wesley, Reading, MA (1968)
40. Goldberg, A., Hartline, J., Karlin, A., Saks, M., Wright, A.: Competitive auctions. Games Econ. Behav. **55**(2), 242–269 (2006)
41. Jackson, J.R.: Networks of waiting lines. Oper. Res. **5**, 518–521 (1957)
42. Jelenkovic, P.R., Lazar, A.A., Semret, N.: The effect of multiple time scales and subexponentiality in mpeg video streams on queueing behavior. IEEE J. Sel. Areas Commun. **15**(6), 1052–1071 (1997)
43. Karp, R.M.: Complexity of Computer Computations, Chapter Reducibility Among Combinatorial Problems. pp. 85–103. Plenum Press, New York, NY (1972)
44. Kingman, J.F.C.: Inequalities in the theory of queues. J. R. Stat. Soc. **32**, 102–110 (1970)
45. Kingman, J.F.C.: 100 years of queueing. In: Proceedings of Conference on The Erlang Centennial, pp. 3–13 (2009)
46. Klemperer, P.: Auction theory: a guide to the literature. J. Econ. Surv. **13**(3), 227–286 (1999)
47. Krishna, V.: Auction Theory. Academic Press, San Diego (2002)
48. Kuehn, P.J.: Approximate analysis of general queueing networks by decomposition. IEEE Trans. Commun. **27**, 113–126 (1978)
49. Kumar, R., Raghavan, P., Rajagopalan, S., Sivakumar, D., Tomkins, A., Upfal, E.: Stochastic models for the web graph. In: Proceedings of the 41st Annual Symposium on Foundations of Computer Science, pp. 57–65 (2000)

50. Laffont, J.J., Robert, J.: Optimal auction with financially constrained buyers. Econ. Lett. **52**(2), 181–186 (1996)
51. Leland, W.E., Taqqu, M.S., Wilson, D.V.: On the self-similar nature of Ethernet traffic. ACM SIGCOMM Comput. Commun. Rev. **25**(1), 202–213 (1995)
52. Lindley, D.V.: The theory of queues with a single server. In: Mathematical Proceedings of the Cambridge Philosophical Society (1952)
53. Malakhov, A., Vohra, R.: Single and multi-dimensional optimal auctions—a network approach. CMS-EMS DP No. 1397, Northwestern University (2004)
54. Manellia, A.M., Vincent, D.R.: Multidimensional mechanism design: revenue maximization and the multiple-good monopoly. J. Econ. Theory **137**(1), 153–185 (2007)
55. Maskin, E.S.: Auctions, development, and privatization: efficient auctions with liquidity constrained buyers. Eur. Econ. Rev. **44**, 667–681 (2000)
56. Merton, R.C.: Option pricing when underlying stock returns are discontinuous. J. Financ. Econ. **3**, 125–144 (1976)
57. Myerson, R.B.: Optimal auction design. Math. Oper. Res. **6**(1), 58–73 (1981)
58. Nemirovski, A., Shapiro, A.: Convex approximations of chance constrained programs. SIAM J. Optim. **17**(4), 969–996 (2006)
59. Nisan, N., Roughgarden, T., Tardos, E., Vazirani, V.V.: Algorithmic Game Theory. Cambridge University Press, New York, NY, USA (2007)
60. Nolan, J.P.: Numerical calculation of stable densities and distribution functions: heavy tails and highly volatile phenomena, communications in statistics. Stoch. Models. **13**, 759–774 (1997)
61. Pai, M., Vohra, R.: Optimal auctions with financially constrained bidders. Working paper (2008)
62. Papadimitriou, C.H., Pierrakos, G.: On optimal single-item auctions. In: STOC (2011)
63. Ronen, A.: On approximating optimal auctions. In: ACM Conference on Electronic Commerce, pp. 11–17 (2001)
64. Shannon, C.E.: A mathematical theory of communication. Bell Syst. Tech. J. **27**, 379–423 (1948)
65. Sherali, H.D., Alameddine, A.: A new reformulation linearization technique for bilinear programming problems. J. Global Optim. **2**, 379–410 (1992)
66. Taylor, S.J.: Modelling stochastic volatility: a review and comparative study. Math. Finance **4**(2), 183–204 (1994)
67. Thanassoulis, J.: Haggling over substitutes. J. Econ. Theory **117**(2), 217–245 (2004)
68. Vickrey, W.: Counterspeculation, auctions, and competitive sealed tenders. J. Finance **16**(1), 8–37 (1961)
69. Vohra, R.: Mechanism Design: A Linear Programming Approach. Cambridge University Press, New York, NY, USA (2011)
70. Whitt, W.: The queueing network analyzer. Bell Syst. Tech. J. **62**, 2779–2813 (1983)
71. Willinger, W., Paxson, V., Taqqu, M.S.: Self-similarity and heavy tails: structural modeling of network traffic. A Practical Guide to Heavy Tails: Statistical Techniques and Applications (1998)
72. Wilson, R.B.: Nonlinear Pricing. Oxford University Press, Oxford (1997)
73. Yanikoglu, I., den Hertog, D.: Safe approximations of chance constraints using historical data. Technical Report, Tilburg University, Center for Economic Research (2011)