Proceedings of the 2013 Winter Simulation Conference
R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl, eds.

TWO-STAGE LIKELIHOOD ROBUST LINEAR PROGRAM WITH APPLICATION TO WATER ALLOCATION UNDER UNCERTAINTY

David Love

Program in Applied Mathematics
University of Arizona
Tucson, AZ 85721, USA

Güzin Bayraksan

Integrated Systems Engineering
The Ohio State University
Columbus, OH 43210, USA

ABSTRACT

We adapt and extend the likelihood robust optimization method recently proposed by Wang, Glynn, and Ye for the newsvendor problem to a more general two-stage setting. We examine the value of collecting additional data and the cost of finding a solution robust to an ambiguous probability distribution. A decomposition-based solution algorithm to solve the resulting model is given. We apply the model to examine a long-term water allocation problem in the southeast area of Tucson, AZ under ambiguous distribution of future available supply and demand and present computational results.

1 INTRODUCTION AND MOTIVATION

In practice, many optimization problems can be modeled by stochastic programs minimizing an expected value of an uncertain objective function. However, if the distribution of the uncertain parameters used in the model is incorrect, the stochastic program can give highly suboptimal results. Such problems have led to the development of distributionally robust optimization, a modeling technique that replaces the probability distribution by a set of distributions, and optimizes the expected cost relative to the worst distribution in the uncertainty set. One approach to this has been recently proposed by Wang, Glynn, and Ye (2010) is to use a set of distributions that are within a sufficiently high empirical likelihood for a given set of observations. This is called the Likelihood Robust Optimization (LRO) method. In this paper, we adapt and extend LRO to a more general setting of two-stage stochastic linear program with recourse and call this the two-stage likelihood robust linear program (LRLP-2). We examine the properties of the resulting model and develop a simple condition for assessing the value of collecting extra data. Finally, we present a modified Bender's Decomposition to solve the LRO and apply the above results to a water distribution planning problem.

LRLP-2 is an ambiguous stochastic program that is modeled on a two-stage minimax problem. Stochastic programs with uncertain objective functions have long been studied by applying the minimax approach to an expected cost; see, e.g., (Dupačová 1987). Shapiro and Kleywegt (2002) and Shapiro and Ahmed (2004) developed methods for converting stochastic minimax problems into equivalent stochastic programs with a certain distribution.

In recent years, there has been a growing interest in distributionally robust methods. Erdoğan and Iyengar (2006) study chance-constrained stochastic programs where the set of distributions considered is determined by the Prohorov metric. Calafiore and Campi (2005) develop a data-driven method for generating feasible solutions to chance constrained problems, and later Calafiore and El Ghaoui (2006) develop a method for converting distributionally robust chance constraints into second-order cone constraints. Jiang and Guan (2013) develop an exact approach to solving data-driven chance constrained programs. Delage and Ye (2010) provide methods for modeling uncertain distributions of a specific form (e.g., Gaussian, exponential, etc.) or using moment-based constraints. Ben-Tal et al. (2013) studies distributionally robust stochastic programs when the uncertainty region is defined by selecting distributions using a ϕ-divergence.

The method of Likelihood Robust Optimization proposed in Wang, Glynn, and Ye (2010) that we adapt in this paper is a special case of the ϕ-divergence measures of Ben-Tal et al. (2013). In particular, it defines the uncertainty set of distributions by the Kullback-Leibler (KL) divergence (a special case of ϕ-divergence) from a "nominal" distribution. The contributions of our work are that it provides (i) a simple condition to determine if an additional observation will change the worst-case distribution used in the optimal solution, (ii) asymptotic analysis to discuss conditions under which the optimal value and solution set LRLP-2 will converge to the two-stage stochastic program with recourse under the true distribution, (iii) a specialized decomposition-based algorithm to solve the resulting model, and (iv) application to water allocation problem under ambiguous uncertainty.

The recent work of Hu and Hong (2013) studies similar problems with ambiguous uncertainty either in the objective or the constraints where the uncertainty sets are defined by the KL-divergence. Hu and Hong (2013) differs from this work by considering a continuous distributions, and doesn't relate the nominal distribution to observational data. They produce a similar dual problem using the nominal distribution, which differs from this work by making use of the moment generating function. Our results also provide one quantification of the value of additional data, and apply the LRO method specifically to a two stage problem, and to water allocation.

The LRO is an attractive data-driven approach because it uses the data directly; and only those data points or scenarios of interest are used in the calculations. These scenarios can come from direct observation, results of simulation, from expert opinion regarding scenarios that the decision maker would especially like to be robust against. Because the LRO depends only on these scenarios, the size of the problem is polynomial in the sample size, making it computationally tractable.

We apply LRO to a generalized network model of Colorado River water allocation in Tucson, Arizona motivated by the CALVIN (CALifornia Value Integrated Network) optimal water allocation model of California created by Draper et al. (2003). Other models of Colorado River water distribution have also been studied, such as the Colorado River Reservoir Model (Christensen et al. 2004) and the Colorado River Budget Model (Barnett and Pierce 2009). Our model is modified to incorporate ambiguous future uncertainty by using the LRO approach.

This paper is organized as follows. Section 2 presents the extension of LRO model to a two-stage stochastic program with recourse. Section 3 describes basic properties of LRO and discusses how to select the level of robustness; Section 4 discusses the value of collecting additional data; and Section 5 discusses the asymptotic properties of LRO. In Section 6 we present a decomposition method for solving the LRO model; and in Section 7 we present a generalized network water model and computational results for its LRO model. Finally, we end in Section 8 with conclusions and future work.

2 LRLP-2 FORMULATION

We begin with a two-stage stochastic linear program with recourse (SLP-2). Let $\mathbf{x}$ be a vector of first stage decision variables with cost vector $\mathbf{c}$, constraint matrix A and right hand side $\mathbf{b}$. We assume a finite distribution given by probabilities p_ω of scenarios indexed by $\omega = 1,\ldots,n$. The SLP-2 is

$$\min_{\mathbf{x}} \left\{ \mathbf{c}\mathbf{x} + \sum_{\omega=1}^{n} p_\omega h^{\dagger}_{\omega}(\mathbf{x}) : \quad A\mathbf{x} = b, \mathbf{x} \geq 0 \right\}. \tag{1}$$

where

$$h^{\dagger}_{\omega}(\mathbf{x}) = \min_{y^\omega} \ \{\mathbf{q}^\omega y^\omega : \quad D^\omega y^\omega = B^\omega \mathbf{x} + \mathbf{d}^\omega, y^\omega \geq 0\} \tag{2}$$

We assume relatively complete recourse; i.e., the second-stage problems $h^{\dagger}_{\omega}(\mathbf{x})$ are feasible for every feasible solution $\mathbf{x}$ of the first-stage problem.

The SLP-2 formulation assumes that the distribution $\{p_\omega\}_{\omega=1}^{n}$ is known. However, in many applications, including our water planning, the distribution is itself unknown. One technique to deal with this is to replace the known distribution with an *ambiguity set* of distributions; i.e., a set of distributions which is believed to contain the true distribution. In the likelihood robust formulation, we assume scenario ω has been observed N_ω times, with $N=\sum_{\omega=1}^n N_\omega$ total observations. In SLP-2, this would correspond to probability of scenario ω to be $\hat{p}_\omega^N = N_\omega/N$, which is the empirical distribution and also the maximum likelihood distribution. The ambiguity set, however, is defined by the set of distributions with sufficiently high empirical likelihood $\prod_{\omega=1}^n p_\omega^{N_\omega}$. By replacing the specific distribution in SLP-2 with a set of distributions with high empirical likelihood, we create a model that we refer to as two-stage likelihood robust linear program with recourse (LRLP-2).

To derive the LRLP-2, we begin by writing SLP-2 given in (1)–(2) in extensive form

$$\begin{array}{llcll}
\min_{\mathbf{x},y^\omega} & \mathbf{cx} & +\sum_\omega p_\omega \mathbf{q}^\omega y^\omega & & \\
\text{s.t.} & A\mathbf{x} & & =\mathbf{b} & \\
 & -B^\omega \mathbf{x} & +D^\omega y^\omega & =\mathbf{d}^\omega, & \forall\omega \\
 & \mathbf{x} & & \geq 0 & \\
 & & y^\omega & \geq 0, & \forall\omega.
\end{array}$$

The SLP-2 formulation is then augmented by the set of distributions with sufficiently high likelihood. To be robust against all these possible distributions, the distribution that results in the maximum expected cost is considered. Then, the objective function is minimized with respect to this worst-case distribution selected from the ambiguity set of distributions. The resulting minimax formulation of LRLP-2 is

$$\min_{\mathbf{x},y^\omega} \max_{p}\ \mathbf{cx}+\sum_\omega p_\omega \mathbf{q}^\omega y^\omega \tag{3}$$

$$\text{s.t. } A\mathbf{x}=\mathbf{b},\quad \mathbf{x}\geq 0$$

$$-B^\omega \mathbf{x}+D^\omega y^\omega = \mathbf{d}^\omega,\ \forall\omega$$

$$\sum_\omega N_\omega \log p_\omega \geq \gamma \tag{4}$$

$$\sum_\omega p_\omega = 1 \tag{5}$$

$$y^\omega, p_\omega \geq 0,\ \forall\omega. \tag{6}$$

Following Wang, Glynn, and Ye (2010), we have introduced the *likelihood parameter* γ, and used it to construct the ambiguity set of distributions $\{p_\omega\}_{\omega=1}^n$ satisfying constraints (4), (5), and (6). Note that constraint (4) is equivalent to $\prod_{\omega=1}^n p_\omega^{N_\omega} \geq e^\gamma$, which explicitly states that the empirical likelihood should be above a certain desired level dictated by e^γ. Constraint (5), along with nonnegativity constraints on p_ω given in (6), simply ensure that $\{p_\omega\}_{\omega=1}^n$ constitutes a probability distribution. Let $0\leq\gamma'\leq 1$ be the *relative likelihood parameter* that expresses γ as a *proportion* of the maximum likelihood; i.e., $\gamma=\log(\gamma'\prod_\omega(\frac{N_\omega}{N})^{N_\omega})$. Using γ', constraint (4) can be rewritten as

$$\sum_{\omega=1}^n N_\omega \log\left(\frac{p_\omega}{N_\omega/N}\right)\geq \log\gamma'. \tag{7}$$

Taking the dual of the inner maximization problem, with dual variables λ and μ, of constraints (4) and (5), respectively, yields

$$\min_{\lambda,\mu}\ \mu+\bar{N}\lambda+\sum_\omega N_\omega\lambda(\log\lambda-\log(\mu-\mathbf{q}^\omega y^\omega))$$

$$\text{s.t. } \lambda\geq 0$$

$$\mu\geq \mathbf{q}^\omega y^\omega,\ \forall\omega,$$

where $\bar{N} = N(\log N - 1) - \log\gamma'$. Combining the two minimizations gives LRLP-2 in extensive form

$$\begin{aligned} \min_{\mathbf{x},\lambda,\mu,y^\omega} \ & \mathbf{cx} + \mu + \bar{N}\lambda + \sum_\omega N_\omega \lambda(\log\lambda - \log(\mu - \mathbf{q}^\omega y^\omega)) \\ \text{s.t. } & A\mathbf{x} = \mathbf{b}, \quad \mathbf{x} \geq 0 \\ & -B^\omega \mathbf{x} + D^\omega y^\omega = \mathbf{d}^\omega, \ \forall\omega \\ & \mu \geq \mathbf{q}^\omega y^\omega, \ \forall\omega \\ & \lambda \geq 0, \quad y^\omega \geq 0, \ \forall\omega. \end{aligned} \tag{8}$$

Finally, we wish to return the LRLP-2 to two-stage formulation. All terms inside the sum over ω will be put into the second stage. To make the formulation as similar to SLP-2 as possible, we choose to express the second stage as an expected value using the maximum likelihood distribution $\frac{N_\omega}{N}$. The formulation becomes

$$\begin{aligned} \min_{\mathbf{x},\lambda,\mu} \ & \mathbf{cx} + \mu + \bar{N}\lambda + \sum_\omega \frac{N_\omega}{N} h_\omega(\mathbf{x},\lambda,\mu) \\ \text{s.t. } & A\mathbf{x} = \mathbf{b}, \quad \mathbf{x},\lambda \geq 0 \end{aligned} \tag{9}$$

where

$$h_\omega(\mathbf{x},\lambda,\mu) = \min_{y^\omega} N\lambda(\log\lambda - \log(\mu - \mathbf{q}^\omega y^\omega)) \tag{10}$$

$$\begin{aligned} \text{s.t. } & -B^\omega \mathbf{x} + D^\omega y^\omega = \mathbf{d}^\omega \\ & \mu - \mathbf{q}^\omega y^\omega \geq 0 \\ & y^\omega \geq 0. \end{aligned} \tag{11}$$

3 BASIC PROPERTIES OF LRLP-2

In this section, we list some basic properties of LRLP-2. Many of these have also been noted in Wang, Glynn, and Ye (2010) in the newsvendor setting and Ben-Tal et al. (2013) in the phi-divergence setting and in (Hu and Hong 2013) in KL-divergence setting, but we list them here for completeness. We also provide slight additions and point to when these properties help with our special solution method.

3.1 Coherent Risk Measure and Convexity

As noted in Wang, Glynn, and Ye (2010), the LRLP-2 problem can be viewed as minimizing a coherent risk measure. It is well known that coherent measures of risk can be interpreted as worst-case expectations from a set of probability measures and LRLP-2 is one such example. A coherent risk measure (in the basic sense), as defined in Rockafellar (2007), is a functional $\mathscr{R} : L^2 \to (-\infty,\infty]$ defined on random variables such that

1. $\mathscr{R}(C) = C$ for all constants C,
2. $\mathscr{R}((1-\lambda)X + \lambda X') \leq (1-\lambda)\mathscr{R}(X) + \lambda\mathscr{R}(X')$, i.e., $\mathscr{R}$ is convex,
3. $\mathscr{R}(X) \leq \mathscr{R}(X')$ when $X \leq X'$, i.e., $\mathscr{R}$ is monotonic,
4. $\mathscr{R}(X) \leq 0$ when $||X^k - X||_2 \to 0$ with $\mathscr{R}(X^k) \leq 0$, i.e., $\mathscr{R}$ is closed,
5. $\mathscr{R}(\lambda X) = \lambda\mathscr{R}(X)$ for $\lambda > 0$, i.e., $\mathscr{R}$ is positively homogeneous.

Proposition 1 LRLP-2 is equivalent to minimizing a coherent risk measure.

Proof. Rockafellar (2007) shows that $\mathscr{R}$ is a coherent risk measure if and only if it can be written using a risk envelope. We will show that LRLP-2 can be written in the form of a risk envelope in the primal form

(3) with the change of variables $Q_\omega = \frac{p_\omega}{1/n}$. Throughout the proof, all expectations are taken with respect to the discrete uniform distribution. First, probability constraint (5) can be written as $\mathbb{E}[Q] = 1$, where Q is the random variable taking values Q_ω with equal probability. Then the likelihood constraint (4) becomes $\sum_{\omega=1}^{n} N_\omega \log Q_\omega \geq \gamma - N \log n$. Combining these yields the set $\mathscr{Q} = \{Q : \quad \mathbb{E}[Q] = 1, \sum_{\omega=1}^{n} N_\omega \log Q_\omega \geq \gamma - N \log n\}$, a closed and convex risk envelope. Finally, we can rewrite the inner maximization of (3) as $\max_{Q \in \mathscr{Q}} \mathbb{E}\left[Q h^\dagger(x)\right]$, which defines a coherent risk measure, where $h^\dagger(x)$ denotes the random variables defined by $\{h^\dagger_\omega(x)\}_{\omega=1}^{n}$. Thus we see that LRLP-2 can be written to minimize a coherent risk measure. $\square$

LRLP-2 is a convex problem. This immediately follows from minimizing a coherent risk measure over a polyhedron.

3.2 Time Structure

LRLP-2 preserves the same time structure as SLP-2. Since log is uniformly increasing, we can rewrite the second stage problem (10) as $h_\omega(\mathbf{x}, \lambda, \mu) = (-N\lambda)\log(\mu - \min_{y^\omega \in Y^\omega} \mathbf{q}^\omega y^\omega)$ with $Y^\omega = \{y^\omega : \quad -B^\omega \mathbf{x} + D^\omega y^\omega = \mathbf{d}^\omega, \mu - \mathbf{q}^\omega y^\omega \geq 0, y^\omega \geq 0\}$. Thus we can state the second stage of LRLP-2 in terms of the second stage of SLP-2, $h_\omega(\mathbf{x}, \lambda, \mu) = N\lambda \left[\log \lambda - \log(\mu - h^\dagger_\omega(\mathbf{x}))\right]$ with $\mu \geq h^\dagger_\omega(\mathbf{x}), \forall \omega$. This preservation of the time structure allows us to easily convert (sub-)derivatives of $h^\dagger_\omega(\mathbf{x})$ to (sub-)derivatives of $h_\omega(\mathbf{x}, \lambda, \mu)$. We will use this in the decomposition method provided in Section 6.

3.3 KKT Conditions

Since LRLP-2 is a convex optimization problem, the KKT conditions for (8) give the relation between primal and dual variables

$$p_\omega = \frac{\lambda N_\omega}{\mu - h^\dagger_\omega(\mathbf{x})}. \tag{12}$$

This is essentially the same condition identified in Wang, Glynn, and Ye (2010) for the newsvendor problem but for the more general case we consider here.

3.4 Relation to KL Divergence and Level of Robustness

The likelihood (4) and relative likelihood (7) constraints can be written in the form of a Kullback-Leibler divergence condition, i.e., (4) can be rewritten as

$$\sum_{\omega=1}^{n} \frac{N_\omega}{N} \log\left(\frac{N_\omega / N}{p_\omega}\right) \leq \frac{1}{N} \log \frac{1}{\gamma'},$$

where $\sum_{\omega=1}^{n} q_\omega \log \frac{q_\omega}{p_\omega}$ is the KL divergence $D_{KL}(q,p)$ of distributions $p = \{p_\omega\}_{\omega=1}^{n}$ and $q = \{q_\omega\}_{\omega=1}^{n}$. So the LRO constraint (4) essentially restricts the set of distributions to be sufficiently close to the empirical (or maximum likelihood) distribution in terms of the KL divergence.

We have introduced the relative likelihood parameter γ', but have not yet made any recommendations on how to choose it. Remark 3.1 of Pardo (2005) shows that $2ND_{KL}(\hat{p}^N, p^{\text{true}}) \Rightarrow \chi^2_{n-1}$, where p^{true} is the true distribution from which the empirical distribution $\hat{p}^N$ is sampled, $\Rightarrow$ indicates convergence in distribution, and χ^2_{n-1} is a χ^2 distribution with $n-1$ degrees of freedom. Thus, to choose γ' to generate a $100(1-\alpha)\%$ confidence region on the true distribution, select the asymptotic value

$$\gamma' = \exp\left(-\frac{\chi^2_{n-1,1-\alpha}}{2}\right). \tag{13}$$

Note that keeping γ' constant asymptotically produces constant-size confidence regions.

Wang, Glynn, and Ye (2010) suggest a Bayesian interpretation of the likelihood constraint (4) which yields a Monte Carlo method for establishing a value for γ'. Our computational test indicate that this Monte Carlo estimate converges to γ' given in (13) as more data is collected.

4 THE VALUE OF DATA

With a data driven formulation such as LRLP-2, it is natural to ask how the behavior changes as more data is gathered. In particular, for robust formulations like LRLP-2 one might be concerned about being overly conservative in the problem formulation and thus missing the opportunity to find a better solution to the true distribution. For LRLP-2, this means that the initial model is likely to be more conservative in an effort to be robust, while the new information could make the model less conservative because new information removes the current worst case distribution from the ambiguity set. In this section, we present a simple condition to determine if taking an additional sample will eliminate the old worst-case distribution and allow for better optimization; i.e., a lower-cost solution. We also provide a computationally simple way to estimate a lower bound on the probability of sampling such an observation.

Proposition 2 An additional sample of scenario $\hat{\omega}$ will result in a decrease in the worst-case expected cost of the LRLP-2 if the following condition is satisfied

$$\frac{N_{\hat{\omega}}}{N} > \left(\frac{N+1}{N}\right) p_{\hat{\omega}}, \tag{14}$$

where $p_{\hat{\omega}}$ is the probability of scenario $\hat{\omega}$ in the worst-case distribution found by solving LRLP-2 using N total observations.

Proof. Consider again the deterministic equivalent formulation of LRLP-2 in (8). Let $f_N(\mathbf{x},\mu,\lambda) = \mathbf{cx}+\mu+\bar{N}\lambda+\sum_\omega N_\omega\lambda\,(\log\lambda\, - \log(\mu - h^\dagger(\mathbf{x})))$ be the objective function, and $z_N = \min_{\mathbf{x},\mu,\lambda} f_N(\mathbf{x},\mu,\lambda)$. We wish to find a simple estimate of the decrease in the optimal cost associated with taking an additional sample, $z_N - z_{N+1}$, looking in particular for a condition under which $z_N - z_{N+1} > 0$.

Let $\mathbf{x}^*_N, \mu^*_N, \lambda^*_N \in \operatorname{argmin} f_N(\mathbf{x},\mu,\lambda)$ be optimal solutions to the N-sample problem. Then $z_N - f_{N+1}(\mathbf{x}^*_N, \mu^*_N, \lambda^*_N)$ is a lower bound on the decrease in optimal cost $z_N - z_{N+1}$. Let $\hat{\omega}$ be the scenario that is selected with the additional sample, then

$$z_N - f_{N+1}(\mathbf{x}^*_N, \mu^*_N, \lambda^*_N) = \left[\overline{N} - \overline{N+1} - \log\lambda^*_N + \log(\mu^*_N - h^\dagger_{\hat{\omega}}(\mathbf{x}^*_N))\right]\lambda^*_N.$$

We can bound $\overline{N} - \overline{N+1} = N\log N - (N+1)\log(N+1)+1$ by using the tangent lines $\log\mathbf{x}+1 \le (\mathbf{x}+1)\log(\mathbf{x}+1) - \mathbf{x}\log\mathbf{x} \le \log(\mathbf{x}+1)+1$ to get $\overline{N} - \overline{N+1} \ge -\log(N+1)$. Combining these results gives the condition

$$z_N - f_{N+1}(\mathbf{x}^*_N, \mu^*_N, \lambda^*_N) \ge \left[-\log(N+1) - \log\lambda^*_N + \log(\mu^*_N - h^\dagger(\mathbf{x}^*_N))\right]\lambda^*_N > 0.$$

Note that $\lambda^*_N > 0$, so to guarantee a drop in optimal cost we must show that the first term is positive. This then simplifies to

$$\frac{\mu^*_N - h^\dagger(\mathbf{x}^*_N)}{\lambda^*_N(N+1)} > 1.$$

Using the KKT condition (12), this can be rewritten as (14). □

We can interpret (14) as follows. If an additional sample is taken from the unknown distribution and the resulting observed scenario $\hat{\omega}$ satisfies (14), then the $(N+1)$-sample problem will have a lower cost than the N-sample problem that was already solved. This is equivalent to saying that an additional observation of $\hat{\omega}$ will rule out the computed worst-case distribution given by $\{p_\omega\}_{\omega=1}^n$ given in (12).

Next, we would like a lower bound on the probability that the next sample will decrease the optimal cost. Let $L = \left\{\omega : \frac{N_\omega}{N} > \left(\frac{N+1}{N}\right) p_\omega\right\}$, where p_ω is the worst-case distribution discussed above and $N_L = \sum_{\omega\in L} N_\omega$.

That is, L gives the set of scenarios that, if sampled one more observation, would result in a decrease in the optimal cost in LRLP-2 and N_L gives the current number of observations in set L. While we would like to estimate a probability of sampling from set L, we do not know the distribution. However, we can find an approximate lower bound on this probability by using the ambiguity set; that is, by solving $\min\{\sum_{\omega \in L} q_\omega : \quad \sum_\omega N_\omega \log q_\omega \geq \gamma, \sum_\omega q_\omega = 1, q_\omega \geq 0, \ \forall \omega\}$, where we introduced q_ω to distinguish from the worst-case distribution calculated in LRLP-2. Note that N_L/N provides an upper bound to this problem as the maximum likelihood estimator is always in the ambiguity set. More details on this probability estimation can be found in Love and Bayraksan (2013).

5 ASYMPTOTIC ANALYSIS OF LRLP-2

We now wish to discuss conditions under which the optimal value and solution of LRLP-2 converges to the optimal value and solution of the corresponding SLP-2 with the (unknown) true distribution p^{true}. In the discussion below, $\hat{p}^N_\omega = \frac{N_\omega}{N}$ denotes the empirical distribution and p_ω denotes the worst-case distribution found by solving the LRLP-2.

We begin by showing that the worst-case distribution converges weakly to the true distribution as $N \to \infty$. Let γ' be fixed and probability space $(\Xi, \mathscr{F}, \mathbb{P}^\infty)$ be the space associated with taking infinitely many random samples from the distribution p^{true}. We know from the Strong Law of Large Numbers (SLLN) that $p^N_\omega \to p^{\text{true}}_\omega$ with probability one (wp1) for all $\omega = 1, \ldots, n$. Let $\Xi' \subseteq \Xi$ be a measure 1 set such that $\hat{p}^N_\omega(\xi) \to p^{\text{true}}_\omega$ for all $\omega = 1, \ldots, n$.

Proposition 3 For all $\varepsilon > 0$ and $\xi \in \Xi'$ there exists N' such that for all $N \geq N'$ $D_{KL}(\hat{p}^N, p) \leq \frac{1}{N} \log\left(\frac{1}{\gamma'}\right)$ implies $\max_\omega |p_\omega - p^{\text{true}}_\omega| \leq \varepsilon$.

Proof. Without loss of generality, we assume that $p^{\text{true}}_\omega > 0$ for all ω. This is valid because $p^{\text{true}}_\omega = 0$ implies $\hat{p}^N_\omega = 0$, which implies $p_\omega = 0$ in the worst-case distribution by (12). For simplicity, we additionally assume ε is chosen so that $p^{\text{true}}_\omega > \frac{\varepsilon}{2}$ for all ω. Note that KL divergence can be written as

$$D_{KL}(\hat{p}^N, p) = \sum_{\omega=1}^n \hat{p}^N_\omega \left(-\log \frac{p_\omega}{\hat{p}^N_\omega} + \frac{p_\omega}{\hat{p}^N_\omega} - 1 \right),$$

where each term in parentheses $\phi(t) = -\log t + t - 1$ is convex, nonnegative for $t \geq 0$, and attains its minimum at $\phi(1) = 0$.

First, note that $\max_\omega |p_\omega - p^{\text{true}}_\omega| \leq \max_\omega |p_\omega - \hat{p}^N_\omega| + \max_\omega |\hat{p}^N_\omega - p^{\text{true}}_\omega|$. For each $\xi \in \Xi'$, let N'' be such that $\max_\omega |\hat{p}^N_\omega - p^{\text{true}}_\omega| \leq \frac{\varepsilon}{2}$ for all $N \geq N''$. Suppose $\max_\omega |p_\omega - p^{\text{true}}_\omega| > \varepsilon$. This implies that $\max_\omega |p_\omega - \hat{p}^N_\omega| > \frac{\varepsilon}{2}$. To complete the proof, we will show that for each $\xi \in \Xi'$ one can choose $N' \geq N''$ such that $\forall N \geq N'$, $\max_\omega |p_\omega - \hat{p}^N_\omega| > \frac{\varepsilon}{2}$ implies $D_{KL}(\hat{p}^N, p) > \frac{1}{N} \log\left(\frac{1}{\gamma'}\right)$. First, bound the KL divergence by

$$\begin{aligned}
D_{KL}(\hat{p}^N, p) &= \sum_{\omega=1}^n \hat{p}^N_\omega \phi\left(\frac{p_\omega}{\hat{p}^N_\omega}\right) \\
&\geq \min_\omega\{\hat{p}^N_\omega\} \cdot \max_\omega \left\{ \phi\left(\frac{p_\omega}{\hat{p}^N_\omega}\right) \right\} \\
&\geq \min_\omega\{\hat{p}^N_\omega\} \cdot \min\left\{ \phi\left(1 + \frac{\varepsilon}{2}\right), \phi\left(1 - \frac{\varepsilon}{2}\right) \right\} \\
&\geq \min_\omega\left\{ p^{\text{true}}_\omega - \frac{\varepsilon}{2} \right\} \cdot \min\left\{ \phi\left(1 + \frac{\varepsilon}{2}\right), \phi\left(1 - \frac{\varepsilon}{2}\right) \right\},
\end{aligned}$$

where the second inequality is true because $\phi\left(\frac{p_\omega}{\hat{p}^N_\omega}\right) \geq \min\left\{ \phi\left(\frac{\hat{p}^N_\omega + \frac{\varepsilon}{2}}{\hat{p}^N_\omega}\right), \phi\left(\frac{\hat{p}^N_\omega - \frac{\varepsilon}{2}}{\hat{p}^N_\omega}\right) \right\}$ for at least one ω, and applying the inequalities $\frac{a+\eta}{a} \geq 1 + \eta$ and $\frac{a-\eta}{a} \leq 1 - \eta$ for $0 < a \leq 1$. This follows from the fact that $\phi(t)$ is a convex function that attains its minimum at 1 over $t > 0$. Then choose $N' \geq N''$ to satisfy $\min_\omega \left\{p_\omega^{\text{true}} - \frac{\varepsilon}{2}\right\} \cdot \min\left\{\phi\left(1+\frac{\varepsilon}{2}\right), \phi\left(1-\frac{\varepsilon}{2}\right)\right\} \geq \frac{1}{N'} \log\left(\frac{1}{\gamma'}\right)$. □

Proposition 3 implies that the worst-case distributions of (3) converge weakly to p^{true}. Let $f(\mathbf{x},\omega) = c\mathbf{x} + h_\omega^\dagger(\mathbf{x})$ and let $\mathbb{E}_{p^{\text{true}}}[f(\mathbf{x},\omega)]$ and $\mathbb{E}_p[f(\mathbf{x},\omega)]$ denote the expectation of $f(\mathbf{x},\omega)$ under the true and worst-case distributions, respectively. Then, using Proposition 3, we can establish epiconverenge of $\mathbb{E}_p[f(\mathbf{x},\omega)]$ to $\mathbb{E}_{p^{\text{true}}}[f(\mathbf{x},\omega)]$ under the conditions that the objective function (under the worst-case distribution) is continuous with respect to ω and lower semicontinuous and locally lower Lipschitz with respect to x by using Theorem 3.7 of Dupacová and Wets (1988). The type of problems that satisfy these conditions can be found in (Ruszczyński and Shapiro 2003). This can then be used to establish for a class of problems, as $N \to \infty$, all limit points of optimal solutions of LRLP-2 solve SLP-2 that uses p^{true} and the optimal value of LRLP-2 converges to that SLP-2 with p^{true}.

6 DECOMPOSITION-BASED SOLUTION METHOD

As the model gets larger, as in our water application presented in Section 7, a direct solution of LRLP-2 becomes computationally expensive. Decomposition-based methods could significantly reduce the solution time and allow for larger problems to be solved efficiently. In this section, we briefly discuss a specialized Bender's decomposition-based method for solving LRLP-2. The algorithm removes constraint (11) from the second-stage problem (10) and exchanges it with a series of linear feasibility constraints (or cuts) in the first-stage problem. Making this change ensures that the second-stage problems are solved using the formulation of $h_\omega^\dagger(\mathbf{x})$ for SLP-2 given in (2), and is more efficient. As a result, both the master and subproblems solved are linear.

The master problem is given by

$$\begin{aligned} \min_{\mathbf{x},\lambda,\mu} \ & \mathbf{cx} + \mu + \bar{N}\lambda + \theta \qquad (15) \\ \text{s.t. } & A\mathbf{x} = \mathbf{b} \\ & \theta \geq T_j(\mathbf{x},\lambda,\mu)^T + t_j, \ j \in J \\ & \mu \geq M_k\mathbf{x} + m_k, \ k \in K \\ & \mathbf{x}, \lambda \geq 0, \end{aligned}$$

where $T_j(\mathbf{x},\lambda,\mu)^T + t_j$ are the objective cuts, $M_k\mathbf{x} + m_k$ are the feasibility cuts on constraint (11), and J and K are the sets of objective and feasibility cuts, respectively. The proposed algorithm is shown in Figure 1. For details on the derivation of optimality and feasibility cuts, see Love and Bayraksan (2013).

In order to enhance the performance of the above decomposition-based algorithm, we made some adjustments. First, we included an L_∞-norm trust region which is scaled up (by a factor of 3) or down (by a factor of $\frac{1}{4}$) when the trust region inhibits finding the optimal solution or when the polyhedral lower approximation is far from the second-stage expected cost, respectively. The trust region is an implementation of Algorithm 4.1 in (Nocedal and Wright 1999). Because we are also interested in the worst-case LRO probabilities given in the primal variables and not computed directly, we include a second tolerance as a stopping condition, ensuring that $|1 - \sum_{i=1}^n p_\omega| < \texttt{TOL2}$ when the algorithm is completed. This must be satisfied in addition to the original condition $z_u - z_l < \texttt{TOL} \min\{|z_u|, |z_l|\}$.

7 APPLICATION TO WATER ALLOCATION PROBLEM AND COMPUTATIONAL RESULTS

7.1 Generalized Network Water Model

We applied LRLP-2 to a multi-period generalized network flow model of Colorado River water allocation in Tucson, AZ, defined by a set of nodes and directed arcs (N, A). The nodes represent available water

---

Algorithm 1 Specialized Bender's Decomposition for LRLP-2

Initialize $z_l = -\infty, z_u = \infty$
Solve first-stage problem given in (15) with $\theta = 0$ to generate $\mathbf{x}$
Solve all second stage scenario sub-problems $h^{\dagger}_{\omega}(\mathbf{x})$ given in (2)
Initialize $\lambda \leftarrow 1$, μ that minimizes $\mu - \sum_{\omega} N_{\omega} \hat{\lambda} \log(\mu - h^{\dagger}_{\omega}(\hat{\mathbf{x}}))$
Generate initial objective cut
while $z_u - z_l \geq \texttt{TOL} \min\{|z_u|, |z_l|\}$ **do**
  Solve master problem (15), get $\mathbf{x}, \lambda, \mu, \theta_M$
  Solve sub-problems $h^{\dagger}_{\omega}(\mathbf{x})$ given in (2)
  $\theta_{\text{true}} \leftarrow \sum_{\omega=1}^{n} \frac{N_{\omega}}{N} h_{\omega}(\mathbf{x}, \lambda, \mu)$
  if $\mu < \max_{\omega} h^{\dagger}_{\omega}(\mathbf{x})$ **then**
    Generate feasibility cut
    Find μ that minimizes $\mu - \sum_{\omega} N_{\omega} \hat{\lambda} \log(\mu - h^{\dagger}_{\omega}(\hat{\mathbf{x}}))$
  else
    $z_l \leftarrow$ master optimal cost $\mathbf{cx} + \mu + \bar{N}\lambda + \theta_{\text{true}}$
  end if
  Generate objective cut
  if $\mathbf{cx} + \mu + \bar{N}\lambda + \theta_{\text{true}} < z_u$ **then**
    $z_u \leftarrow \mathbf{cx} + \mu + \bar{N}\lambda + \theta_{\text{true}}$
    $\mathbf{x}_{\text{best}} \leftarrow \mathbf{x}, \lambda_{\text{best}} \leftarrow \lambda, \mu_{\text{best}} \leftarrow \mu$
    $p_{\omega} \leftarrow \frac{\lambda_{\text{best}} N_{\omega}}{\mu_{\text{best}} - h^{\dagger}_{\omega}(\mathbf{x}_{\text{best}})}$ for $i = 1, \ldots, n$
  end if
end while

---

supply from the Colorado River, water treatment plants, reservoirs, and water demand sites. The arcs represent the conveyance system (pipes, etc.) that carry water between the nodes. Water can be stored in between time periods in reservoirs to meet future demands. The model aims to find the minimal cost water flows considering energy, treatment, storage, and transportation costs over the planning period. Generalized network water allocation models have been used to find water allocations and delivery reliabilities and to assess values of different water use operations; see, e.g., Draper et al. (2003).

Water flows on arc $(i, j) \in A$ during time period $t = 1, \ldots, P$ are represented by decisions x_{ijt}. Each arc $(i, j) \in A$ and time period t has a unit cost c^x_{ijt}, loss coefficient $0 \leq a_{ijt} \leq 1$ to account for evaporation, leakage from the pipes, etc., and bounds on the flow $l^x_{ijt} \leq x_{ijt} \leq u^x_{ijt}$. Each node $j \in N$ has a supply/demand for time period t, denoted b_{jt}. Nodes representing reservoirs are able to store water between time periods. Stored water available at node j at the beginning of time period t is s_{jt}, with associated cost c^s_{jt} and bounds $l^s_{jt} \leq s_{jt} \leq u^s_{jt}$. Finally, water released into the environment from node j in period t is given by r_{jt}, with bounds $l^r_{jt} \leq r_{jt} \leq u^r_{jt}$. The deterministic model is a multi-period generalized network flow model. The model is converted to a two-stage stochastic model with P_1 periods in the first stage and $P - P_1$ stages in the second stage.

$$\min_{(x,s,r)\in L^1} \sum_{(i,j)\in A} \sum_{t=1}^{P_1} c^x_{ijt} x_{ijt} + \sum_{j\in N} \sum_{t=1}^{P_1} c^s_{jt} s_{jt} + \sum_{\omega=1}^{n} p_{\omega} h^{\dagger}_{\omega}(s) \tag{16}$$

$$\text{s.t.} \sum_{i\in N} x_{jit} + s_{j,t+1} + r_{jt} = \sum_{i\in N} a_{ijt} x_{ijt} + s_{jt} + b_{jt}, \quad \forall j, 1 \leq t \leq P_1,$$

where the second stage problems

$$h^{\dagger}_{\omega}(s) = \min_{(x,s,r)\in L^2_{\omega}} \sum_{(i,j)\in A} \sum_{t=P_1+1}^{P} c^x_{ijt} x_{ijt} + \sum_{j\in N} \sum_{t=P_1+1}^{P} c^s_{jt} s_{jt} \tag{17}$$
$$\text{s.t.} \sum_{i\in N} x_{jit} + s_{j,t+1} + r_{jt} = \sum_{i\in N} a_{ijt} x_{ijt} + s_{jt} + b^{\omega}_{jt}, \quad \forall j, P_1+1 \le t \le P,$$

and L^1 and L^2_{ω} represent the feasible regions defined by the lower and upper variable bounds. Note that we assume that the supplies and demands are uncertain, as well as the bounds on the decision variables.

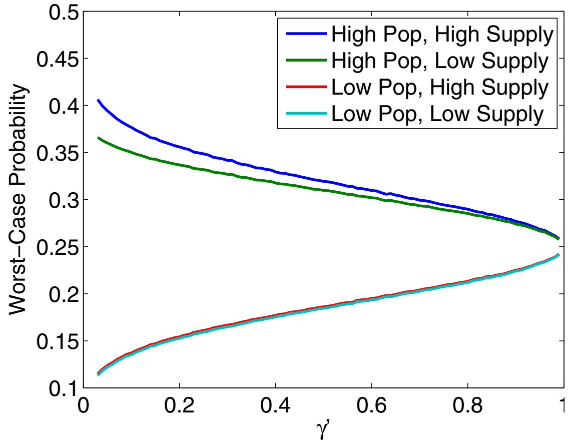
From this point on, in the first stage, decision variables $\{x_{ijt}\}$, $\{s_{jt}\}$ and $\{r_{jt}\}$ become the vector $\mathbf{x}$, costs $\{c^x_{ijt}\}$ and $\{c^s_{jt}\}$ are written as the row vector $\mathbf{c}$, the supply/demand parameters b_{jt} become the vector $\mathbf{b}$ and the constraint matrix is written as A. In the second stage, we denote the decisions as $\mathbf{y}^{\omega}$, the costs as $\mathbf{q}^{\omega}$, the supply/demands as $\mathbf{d}^{\omega}$, and the constraint matrices multiplying $\mathbf{y}^{\omega}$ and $\mathbf{x}$ as D^{ω} and B^{ω}, respectively. Then, we are back to SLP-2 given in (1)–(2) and we turn it to LRLP-2 in the way discussed.

In our application, the model has a total of $P = 41$ time periods, representing years 2010–2050. For each time period, the network has 62 nodes representing demand for potable and nonpotable (reclaimed) water, pumps, water treatment plants, and the available water supply from the Colorado River. The network in each time period has 102 arcs, representing the pipe network carrying the water between the nodes physically and connecting the network to the five reservoirs that connect the time stages in the model. We use $P_1 = 10$ time periods for the first stage. Uncertainty in the second stage takes the form of uncertain population (thus, demand for water) and supply of water. There are a total of 4 scenarios considered in this test instance: (i) high population, high supply, (ii) high population, low supply, (iii) low population, high supply, and (iv) low population, low supply. Each scenario is assumed to have five observations. The high population scenarios are more costly as the system needs to meet demand or pay for unmet demand. The low population scenarios, on the other hand, are not as costly. We applied the decomposition-based solution algorithm presented in Section 6 to solve this model.

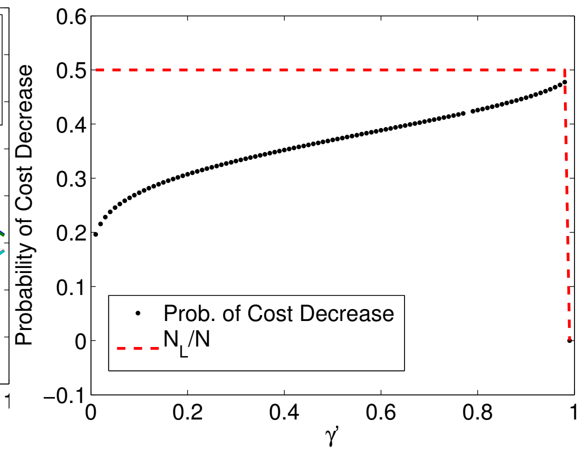
7.2 Computational Results

Figure 1a shows how the worst-case distribution changes with γ'. When γ' is close to 1, we use the maximum likelihood distribution, which has equal $\frac{1}{4}$ probabilities on each of the four scenarios. As γ' is decreased, the ambiguity set increases, and the worst-case distribution used by LRLP-2 changes. It gives higher than $\frac{1}{4}$ probability to the two high-population scenarios and lower than $\frac{1}{4}$ probability to the two low-population scenarios, making the solution more robust to costly scenarios. Note that the scenarios fall into two similar pairs because the cost of each scenario depends strongly on the projected demand but only weakly on the projected supply of Colorado River water. A closer look at the optimal solutions reveals that as γ' is decreased (or as robustness is increased), the solution uses more and more reclaimed water (treated wastewater that is reused for nonpotable purposes such as irrigation) in an effort to meet demands in a least-costly way.

The results of the water model were then analyzed with the value of data techniques from Section 4. Figure 1b shows the estimated probability that an additional sample will remove the worst-case distribution from the likelihood region, resulting in a lower-cost solution. The dashed line in Figure 1b depicts the computed values of $\frac{N_L}{N}$, which provide an upper bound on the estimated probabilities. Because the low-population scenarios have lower costs, an additional sample of either low-population scenario will result in a lower expected cost. This is what we see through most of the computed values of γ', with $\frac{N_L}{N} = 0.5$, indicating that the sufficient condition (14) was satisfied for both low-population scenarios. For extremely large values of γ'—above 0.97—we see the ratio $\frac{N_L}{N}$ quickly drops to zero. As γ' increases and the ambiguity set shrinks, the worst-case probabilities become so close the empirical probabilities that (14) can no longer be satisfied, resulting in $\frac{N_L}{N}$ decreasing to zero.



(a) Worst case distributions



(b) Probability of cost decrease

Figure 1: (a) Worst-case distribution for the likelihood robust water allocation problem. (b) Probability that an additional sample causes a decrease in worst-case expected cost for the likelihood robust water allocation problem. The red line shows the upper bound probability $\frac{N_L}{N}$.

8 CONCLUSION AND FUTURE WORK

In this paper, we proposed an extension of the Likelihood Robust Optimization (LRO) method of Wang, Glynn, and Ye (2010) to general two-stage stochastic programs with recourse, creating a two-stage likelihood robust program with recourse, denoted LRLP-2. The LRO models use the empirical likelihood function to define an ambiguity set of probability distributions using observed data and optimize the worst-case expected cost with respect to this likelihood ambiguity set. We provided a simple condition to determine if an additional sample will produce a likelihood ambiguity set that does not contain the current worst-case distribution and will result in a lower-cost solution and a computationally efficient way to estimate the probability that this will happen. We have also provided a Bender's decomposition-based solution algorithm for the LRLP-2 and applied this method to planning future water distribution in Tucson, Arizona.

Our future work includes the following. On the methodological side, we plan to generalize our results to the more general ϕ-divergence case of Ben-Tal et al. (2013). Multistage extensions also constitute future research. In our application, we plan to augment the existing model first with a richer set of second-stage scenarios. In addition to more varied estimates for future population, we will integrate climate change predictions into the model to generate scenarios for future water supply from the Colorado River. This model is intended to include a facility location problem to determine the best places for an additional waste water treatment plants to increase the use of reclaimed water in the most cost-efficient manner.

ACKNOWLEDGEMENTS

We gratefully acknowledge the support provided by a Water Sustainability Program Fellowship through the Technology and Research Initiative Fund at the University of Arizona. This work has also been partially supported by the National Science Foundation through grant CMMI-1151226. Parts of this paper also appeared in the *Proceedings of the 2013 Industrial and Systems Engineering Research Conference* (Love and Bayraksan 2013).

REFERENCES

Barnett, T., and D. Pierce. 2009. "Sustainable water deliveries from the Colorado River in a changing climate". *Proceedings of the National Academy of Sciences* 106 (18): 7334–7338.

Ben-Tal, A., D. Den Hertog, A. De Waegenaere, B. Melenberg, and G. Rennen. 2013. "Robust solutions of optimization problems affected by uncertain probabilities". *Management Science* 59:341–357.

Calafiore, G., and M. C. Campi. 2005. "Uncertain convex programs: randomized solutions and confidence levels". *Mathematical Programming* 102 (1): 25–46.

Calafiore, G. C., and L. El Ghaoui. 2006. "On distributionally robust chance-constrained linear programs". *Journal of Optimization Theory and Applications* 130 (1): 1–22.

Christensen, N., A. Wood, N. Voisin, D. Lettenmaier, and R. Palmer. 2004. "The effects of climate change on the hydrology and water resources of the Colorado River Basin". *Climatic Change* 62 (1): 337–363.

Delage, E., and Y. Ye. 2010. "Distributionally Robust Optimization Under Moment Uncertainty with Application to Data-Driven Problems". *Operations Research* 58:595–612. Forthcoming. Published online. DOI: 10.1287/opre.1090.0741.

Draper, A., M. Jenkins, K. Kirby, J. Lund, and R. Howitt. 2003. "Economic-engineering optimization for California water management". *Journal of Water Resources Planning and Management* 129:155–164.

Dupačová, J. 1987. "The Minimax Approach to Stochastic Programming and an Illustrative Application". *Stochastics* 20:73–88.

Dupacová, J., and R. Wets. 1988. "Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems". *The Annals of Statistics* 16 (4): 1517–1549.

Erdoğan, E., and G. Iyengar. 2006. "Ambiguous chance constrained problems and robust optimization". *Mathematical Programming* 107 (1-2): 37–61.

Hu, Z., and L. J. Hong. 2013. "Kullback-Leibler Divergence Constrained Distributionally Robust Optimization". Technical report, The Hong Kong University of Science and Technology. Available at: Optimization Online www.optimization-online.org.

Jiang, R., and Y. Guan. 2013. "Data-driven Chance Constrained Stochastic Program". Technical report, University of Florida. Available at: Optimization Online www.optimization-online.org.

Love, D., and G. Bayraksan. 2013. "A Likelihood Robust Method for Water Allocation under Uncertainty". In *Proceedings of the 2013 Industrial and Systems Engineering Research Conference*.

Nocedal, J., and S. Wright. 1999. *Numerical Optimization*. New York, NY: Springer Verlag.

Pardo, L. 2005. *Statistical Inference Based On Divergence Measures*, Volume 185. Chapman and Hall/CRC.

Rockafellar, R. T. 2007. "Coherent approaches to risk in optimization under uncertainty". In *Tutorials in Operations Research*, edited by T. Klastorin, Volume 3, 38–61. Hanover, MD: INFORMS.

Ruszczyński, A., and A. Shapiro. (Eds.) 2003. *Handbooks in OR & MS*, Volume 10: Stochastic Programming. Elsevier Science.

Shapiro, A., and S. Ahmed. 2004. "On a class of minimax stochastic programs". *SIAM Journal on Optimization* 14 (4): 1237–1249.

Shapiro, A., and A. Kleywegt. 2002. "Minimax analysis of stochastic problems". *Optimization Methods and Software* 17 (3): 523–542.

Wang, Z., P. Glynn, and Y. Ye. 2010. "Likelihood robust optimization for data-driven newsvendor problems". Technical report, Department of Management Science and Engineering, Stanford University, USA.

AUTHOR BIOGRAPHIES

DAVID LOVE is a graduate student in the Graduate Interdisciplinary Program in Applied Mathematics at the University of Arizona. His research interests include distributionally robust stochastic programming and water resources management. His email address is dlove@math.arizona.edu and his web page is http://math.arizona.edu/~dlove.

GÜZİN BAYRAKSAN is an Associate Professor of Integrated Systems Engineering at the Ohio State University. Her research interests include Monte Carlo sampling-based and data-driven methods for stochastic programming and applications to water resources management. Her email address is bayraksan.1@osu.edu and her web page is http://go.osu.edu/bayraksan.