Systems Optimization Laboratory

**A Probabilistic Lower Bound
for Two-Stage Stochastic Programs**

by
George B. Dantzig and Gerd Infanger

TECHNICAL REPORT SOL 95-6

November 1995

Department of Operations Research
Stanford University
Stanford, CA 94305

DISCLAIMER

DISCLAIMER

Portions of this document may be illegible electronic image products. Images are produced from the best available original document.

A Probabilistic Lower Bound
for Two-Stage Stochastic Programs[*]

George B. Dantzig[†] and Gerd Infanger[†]

Stanford University

Abstract

In the framework of Benders decomposition for two-stage stochastic linear programs, we estimate the coefficients and right-hand sides of the cutting planes using Monte Carlo sampling. We present a new theory for estimating a lower bound for the optimal objective value and we compare (using various test problems whose true optimal value is known) the predicted versus the observed rate of coverage of the optimal objective by the lower bound confidence interval.

1. Introduction

It has long been recognized that traditional deterministic models do not reflect the true dynamic behavior of real-world applications because they fail to take into account uncertainty. Since Dantzig (1955) [9] and Beale (1955) [1] independently proposed a stochastic model formulation, these models have been studied intensively. In the literature, a number of different algorithmic approaches have been proposed that we can broadly categorize as deterministic methods, approximation schemes, sampling-based algorithms and others.

Deterministic methods attempt to solve the deterministic equivalent problem, either directly or by exploiting structure. Prominent among these are the L-shaped method of Van Slyke and Wets (1969) [48], its multi-cut variant by Birge and Louveaux (1985) [4], the Progressive Hedging algorithm of Rockafellar and Wets (1989) [46], the use of interior-point methods by Lustig *et al.* (1991) [37], Ruszczynski's (1986) [47] regularized decomposition method, and other large-scale techniques implemented on serial and parallel computers. Clearly, even the most sophisticated deterministic techniques can only solve problems with a limited number of scenarios. So far, problems with up to about 100,000 scenarios have been solved using deterministic techniques.

Approximation schemes calculate deterministic lower and upper bounds on the optimal objective of the problem via the inequalities of Jensen (1906) [33] (lower bound) and Edmundson (1956) [17] and Madansky (1959) [39] (upper bound), and successively improve these bounds. Refinements of these bounds have been proposed by many authors, e.g., Kall (1974) [34], Huang, Ziemba and Ben-Tal (1977) [30], Frauendorfer and Kall (1988) [21], Frauendorfer (1988) [20], (1992) [22], Birge and Wallace (1988), Birge and Wets (1987) [6], (1989) [7], Prékopa (1989) [43] and others. Approximation techniques work very well for problems with a small number of stochastic parameters, but seem to encounter difficulties when the number of stochastic parameters is large.

Sampling-based algorithms can be further categorized into methods that pre-sample a number of scenarios to create a tractable deterministic equivalent problem, which is then solved by a suitable deterministic method, and methods that use sampling within the algorithm. In the latter category fall stochastic quasigradient methods (Ermoliev (1983) [18] and Gaivoronski (1988) [23]) that select sequentially random search directions based on a limited number of observations of the random function in each iteration. Others are based on modifications of deterministic decomposition techniques to allow for sampling. Higle and Sen's (1991) [26] Stochastic Decomposition method relies on taking only one observation or a very small number of observations per iteration, while Pereira *et al.* (1989) [44] used control variables as a variance-reduction technique in Monte Carlo sampling in a modified Benders decomposition framework.

Sampling seems to be the best method for practical problems with a large number of stochastic parameters. The approach by Dantzig and Glynn (1990) [10] and Infanger (1992) [31] combines Benders decomposition and Monte Carlo importance sampling for solving stochastic linear programs. Importance sampling serves as a variance-reduction technique and in practice often results in accurate estimates being obtained with only small sample sizes. Infanger (1992) [31] and Dantzig and Infanger (1991) [13] report the solution on personal computers of large-scale problems that seemed to be intractable in the past, even on large mainframes.

2

In this paper we present a rigorous theory for obtaining a probabilistic lower bound for the true optimal objective value when using Benders (1962) [2] decomposition and Monte Carlo sampling for estimating coefficients and right-hand sides of cuts to solve two-stage stochastic linear programs. In Section 2, we state the problem. In Section 3 we review the original Benders decomposition algorithm for two-stage stochastic linear programs (Van Slyke and Wets (1969) [48]). We then derive in Section 4 the theory of a probabilistic lower bound. Finally, in Section 5, we discuss the numerical results obtained from testing the theory on a number of practical problems.

2. The Problem

We consider the following two-stage stochastic linear program:

$$
\begin{array}{rllll}
\min z & = & cx & + & E^{\omega} f^{\omega} y^{\omega} \\
s/t & & Ax & & = b \\
& & -B^{\omega} x & + & D^{\omega} y^{\omega} = d^{\omega}, \quad \omega \in \Omega \\
& & x, & & y^{\omega} \geq 0.
\end{array}
\tag{2.1}
$$

The matrix of constraint coefficients A, the right-hand side vector b, and the objective function coefficients c of the first stage are assumed to be known with certainty. In the second stage, the transition matrix B, the technology (or recourse) matrix D, the right-hand side vector d and the objective function coefficients f are not known with certainty — only their joint distribution of values is assumed to be known.

We denote particular outcomes by $B = B^{\omega}$, $D = D^{\omega}$, $d = d^{\omega}$, $f = f^{\omega}$, $\omega = 1, \ldots, W$ and the known probability of realization ω occurring by $p(\omega)$. The set of all possible realizations of ω is denoted by $\Omega = \{1, \ldots, W\}$, or by $\omega \in \Omega$. Since in many practical applications, D and f don't depend on ω, we have omitted their ω superscript in the rest of the paper to simplify the presentation; the analysis is the same if D and f are replaced by D^{ω} and f^{ω}.

The equivalent deterministic form of (2.1) is (2.2):

$$
\begin{array}{rlllllllll}
\min z = & cx & + & \theta \\
-1: & & & -\theta & + & p^1 f y^1 & + \cdots + & p^{\omega} f y^{\omega} & + \cdots + & p^W f y^W & = 0 \\
\rho: & Ax & & & & & & & & & = b \\
p^1 \pi^1: & -B^1 x & & & + & Dy^1 & & & & & = d^1 \\
& & & & & & \ddots & & & & \vdots \\
p^{\omega} \pi^{\omega}: & -B^{\omega} x & & & & & + & Dy^{\omega} & & & = d^{\omega} \\
& & & & & & & & \ddots & & \vdots \\
p^W \pi^W: & -B^W x & & & & & & & + & Dy^W & = d^W \\
& x, & & & y^1, & \cdots & & y^{\omega}, & \cdots, & y^W & \geq 0,
\end{array}
\tag{2.2}
$$

where θ denotes the expected second-stage costs.

The dual variables corresponding to the primal constraints of (2.2) are displayed in the column to the left of the equations. In particular, ρ is the vector of dual variables associated with the first-stage constraints $Ax = b$, and $p^{\omega} \pi^{\omega}$ is the vector of dual variables associated with the second-stage constraint $-B^{\omega} x + Dy^{\omega} = d^{\omega}$, for each $\omega \in \Omega$.

3

In practical applications the number of possible second-stage realizations W can be very large (e.g., 10^{20} or even 10^{100}), and it is impossible to express system (2.2) explicitly . The different realizations are implicitly generated as needed by a combination of an underlying small set of h independent random parameters $\phi_1, \phi_2, \ldots, \phi_h$, where h might be 100 or so. Sampling is required to solve such problems. The ϕ structure underlying the set Ω makes it possible to use variance-reduction techniques, such as *importance sampling*, to reduce the size of sample required. Since the theory for the lower bound estimate is the same with variance reduction, we omit the discussion of the latter to simplify the presentation.

The stochastic algorithm follows the same steps as the original Benders algorithm, except that the necessary conditions (the *true cuts*) are approximated by *pseudo cuts* obtained by summing over a random sample of ω instead of all ω. After a preassigned number of iterations K the algorithm terminates with a proposed first-stage decision $x = \xi^l$, which yields the lowest approximate expected first-stage and second-stage cost found so far. We review the original algorithm first and then use estimated cuts based on random samples to determine the first-stage decision and estimate bounds on how close its objective is to the true minimum of z.

3. Benders Decomposition

Assumption 1

The problem $\min cx$, $Ax = b$, $x \geq 0$ has a finite optimal solution $x = \xi^1$. It is used to initiate iteration 1 of the Benders algorithm for solving (2.2).

Assumption 2

For any feasible first-stage decision $x = \xi$, each second-stage subproblem ω,

$$
\begin{aligned}
\min \ \theta_\xi^\omega &= f y^\omega \\
s/t \ \pi^\omega : \quad D y^\omega &= d^\omega + B^\omega \xi \\
y^\omega &\geq 0, \ \omega \in \Omega,
\end{aligned}
\tag{3.1}
$$

has optimal primal and dual feasible solutions $y^\omega = y_\xi^\omega$, $\pi^\omega = \pi_\xi^\omega$ that depend on ξ. By the duality theorem, these satisfy

$$
f - \pi_\xi^\omega D \geq 0, \quad D y_\xi^\omega = d^\omega + B^\omega \xi, \quad f y_\xi^\omega = \pi_\xi^\omega (d^\omega + B^\omega \xi), \quad (f - \pi_\xi^\omega D) y_\xi^\omega = 0.
\tag{3.2}
$$

Assumptions 1 and 2 imply that the convex set C of feasible solutions to (2.2) is non-empty and that there exists an optimal solution to (2.2): $z^*, x^*, \theta^*, y_*^\omega$, for $\omega \in \Omega$.

Generating True Cuts

Let $x, \theta, y^1, \ldots, y^W$ be any feasible solution to (2.2). It satisfies

$$
-B^\omega x + D y^\omega = d^\omega, \quad \text{for each } \omega \in \Omega.
\tag{3.3}
$$

For any given first-stage solution ξ and optimal dual feasible π_ξ^ω, multiplying by π_ξ^ω and then adding and subtracting $f y^\omega$ we obtain

$$
-\pi_\xi^\omega B^\omega x + (\pi_\xi^\omega D y^\omega - f y^\omega) + f y^\omega = \pi_\xi^\omega d^\omega.
\tag{3.4}
$$

4

Note that

$$\pi_\xi^\omega D y^\omega - f y^\omega \le 0 \qquad (3.5)$$

because $f - \pi_\xi^\omega D \ge 0$ by (3.2) and $y^\omega \ge 0$. Dropping this term in parentheses in (3.4) we obtain

$$-\pi_\xi^\omega B^\omega x + f y^\omega \ge \pi_\xi^\omega d^\omega, \quad \omega \in \Omega. \qquad (3.6)$$

Multiplying by $p^\omega \ge 0$ and taking the sum over all $\omega \in \Omega$ yields

$$-\sum_{\omega \in \Omega} p^\omega \pi_\xi^\omega B^\omega x + \sum_{\omega \in \Omega} p^\omega f y^\omega \ge \sum_{\omega \in \Omega} p^\omega \pi_\xi^\omega d^\omega. \qquad (3.7)$$

In iteration k the value of ξ is denoted by $\xi = \xi^k$, and we denote

$$G^k = \sum_{\omega \in \Omega} p^\omega \pi_\xi^\omega B^\omega, \quad g^k = \sum_{\omega \in \Omega} p^\omega \pi_\xi^\omega d^\omega, \quad \xi = \xi^k. \qquad (3.8)$$

Substituting these into (3.7), and noting from (2.2), the second term is θ by definition, we obtain the true cut (k), which is valid for all feasible solutions to (2.2):

True Cut (k)

$$-G^k x + \theta \ge g^k. \qquad (3.9)$$

Starting with the first-stage solution $x = \xi^1$ (see Assumption 1), cut (1) is generated and used to form master program (1). The optimal solution x of master problem (1) generates the first-stage solution $x = \xi^2$. After k iterations of the Benders decomposition algorithm, we arrive at the master problem (k), see (3.10).

Master Problem (k)

$$
\begin{array}{rlcll}
\min v = & cx & + & \theta & \\
& Ax & & = & b \\
& -G^1 x & + & \theta \ge & g^1 \\
& \vdots & & \vdots & \vdots \\
& -G^k x & + & \theta \ge & g^k \\
& x, & & \ge & 0.
\end{array} \qquad (3.10)
$$

Master problem (k) is optimized and its optimum solution is denoted $x = \xi^{k+1}$, $\theta = \theta^{k+1}$, $v = v^{k+1}$. ξ^{k+1} is used to initialize iteration $k+1$. Cut $(k+1)$ is then determined by (3.8) and adjoined to master (k) to generate master $(k+1)$, and so on.

Lower bound

In particular, an optimal solution to (2.2), x^*, θ^*, y_*^ω, $\omega \in \Omega$, satisfies (3.10). Therefore

$$z^* = cx^* + \theta^* \qquad (3.11)$$

and

$$-G^k x^* + \theta^* \ge g^k, \quad k = 1, 2, \dots. \qquad (3.12)$$

5

Since master problem (k) is a set of necessary conditions derived from (2.2), clearly its optimum objective $v = v^{k+1}$ is a lower bound for z^*, i.e.,

$$v^{k+1} \leq z^*, \tag{3.13}$$

because master (k) may be viewed as a relaxation of its conditions evaluated at $x = x^*$, $\theta = \theta^*$ as shown here:

$$
\begin{array}{rcccc}
cx^* & + & \theta^* & = & z^* \\
Ax^* & & & = & b \\
-G^1 x^* & + & \theta^* & \geq & g^1 \\
& & \vdots & \vdots & \vdots \\
-G^k x^* & + & \theta^* & \geq & g^k \\
x^* & & & \geq & 0.
\end{array}
\tag{3.14}
$$

Upper bound

Given the first-stage solution $x = \xi^{k+1}$, the corresponding minimum second-stage cost is

$$\theta_\xi^{k+1} = \sum_{\omega \in \Omega} p^\omega f y_\xi^\omega, \quad \xi = \xi^{k+1}. \tag{3.15}$$

The minimum expected cost, given the first-stage decision $\xi = \xi^{k+1}$, is $z_\xi^{k+1} = c\xi^{k+1} + \theta_\xi^{k+1}$. Therefore,

$$z^* \leq z_\xi^{k+1}. \tag{3.16}$$

Termination

If $v^{k+1} = z_\xi^l$ for some $l \leq k+1$, then $x = \xi^l$ is an optimal first-stage decision and $z_\xi^l = z^*$. The condition $v^{k+1} = z_\xi^l$ will be reached in a finite number of iterations; see Benders (1962) [2]. In practice, however, the iterative process is terminated if

$$z_\xi^l - v^{k+1} \leq TOL, \quad l \leq k+1, \tag{3.17}$$

where TOL is a preassigned "close enough" criterion, implying

$$z_\xi^l - z^* \leq TOL. \tag{3.18}$$

The objective value of the first-stage decision $x = \xi^l$ is then deemed to be "close enough" to z^*. Otherwise, the iterative process continues with increasing k until k reaches a preassigned maximum number K and the solution $x = \xi^l$, where $l = \arg\min z_\xi^k$, $k = 1, \ldots, K$, is then chosen as the first-stage decision.

4. Probabilistic Lower Bound

4.1. Cut Generation Using Sampling

The need for sampling arises if W, the number of possible realizations, is large. We then use Monte Carlo sampling to estimate the cutting planes. If a cut is estimated by a sample of ω, we will call it a *pseudo cut*. Given a first-stage solution $x = \xi$, to generate a pseudo

cut, a random sample S of size $|S|$ of the necessary conditions (3.3) with replacement is taken according to the distribution $p^1, p^2, \ldots, p^W$ and averaged. Thus, summing (3.6) for $\omega \in S$ and averaging, we obtain a true cut, called a *stochastic cut*. Adding $\theta = \sum_{\omega \in \Omega} p^\omega f y^\omega$ to both sides and rearranging terms, we obtain

$$-\frac{1}{|S|} \sum_{\omega \in S} \pi^\omega_\xi B^\omega x + \theta \geq \frac{1}{|S|} \sum_{\omega \in S} \pi^\omega_\xi d^\omega - \left(\frac{1}{|S|} \sum_{\omega \in S} f y^\omega - \theta\right). \tag{4.1}$$

On iteration k, we denote ξ and S by ξ^k and S^k and we denote the various terms of (4.1) by

$$\bar{G}^k = \frac{1}{|S^k|} \sum_{\omega \in S^k} \pi^\omega_\xi B^\omega, \quad \bar{g}^k = \frac{1}{|S^k|} \sum_{\omega \in S^k} \pi^\omega_\xi d^\omega, \quad \xi = \xi^k \tag{4.2}$$

and define

$$\bar{\epsilon}^k = \frac{1}{|S^k|} \sum_{\omega \in S^k} f y^\omega - \theta \tag{4.3}$$

as the *sampling error* of estimating $\theta = \sum_{\omega \in \Omega} p^\omega f y^\omega$ by a random sample S^k of size $|S^k|$ from the distribution of $f y^\omega$ with density p^ω. Substituting $\bar{G}^k$, $\bar{g}^k$ and $\bar{\epsilon}^k$ for the expressions in (4.1) we obtain the stochastic cut (k) as follows:

Stochastic Cut (k)

$$-\bar{G}^k x + \theta \geq \bar{g}^k - \bar{\epsilon}^k, \quad \xi = \xi^k, \quad S = S^k. \tag{4.4}$$

We note that this is a true statement for any feasible solution of (2.2), x, θ, y^ω, for all $\omega \in \Omega$. The corresponding *pseudo cut* (k) is obtained by dropping the error term:

Pseudo Cut (k)

$$-\bar{G}^k x + \theta \geq \bar{g}^k, \quad \xi = \xi^k, \quad S = S^k. \tag{4.5}$$

4.2. Stochastic Benders Algorithm

Starting with any feasible solution $x = \xi^1$ to the first-stage problem (such as the optimal solution of $\min cx$, $Ax = b$, $x \geq 0$) each subsequent iteration k is initialized using as its feasible first-stage solution $x = \xi^k$. A random sample S^k of size $|S^k|$ is drawn and the parameters $\bar{G}^k$ and $\bar{g}^k$ of the pseudo cut (k) are computed according to (4.2). The pseudo cut (k) is then adjoined to pseudo master $(k-1)$ to generate the pseudo master problem — (k), see (4.6).

Pseudo Master (k)

$$
\begin{array}{rlrcccl}
\min \bar{v} & = & cx & + & \theta & & \\
\bar{\rho}: & & Ax & & & = & b \\
\bar{\lambda}^1: & & -\bar{G}^1 x & + & \theta & \geq & \bar{g}^1 \\
\bar{\lambda}^2: & & -\bar{G}^2 x & + & \theta & \geq & \bar{g}^2 \\
& & \vdots & & \vdots & & \vdots \\
\bar{\lambda}^k: & & -\bar{G}^k x & + & \theta & \geq & \bar{g}^k \\
& & x & & & \geq & 0.
\end{array}
\tag{4.6}
$$

Pseudo cuts are iteratively adjoined in this manner until the maximum iteration $k = K$ is reached.

4.3. First-Stage Decision $x = \xi^l$, Upper Confidence Interval for z^*

An unbiased estimate of the *minimum expected second-stage costs* θ_ξ^k, calculated on iteration k, is

$$\tilde{\theta}_\xi^k = \frac{1}{|S^k|} \sum_{\omega \in S^k} f y_\xi^\omega, \quad \xi = \xi^k, \tag{4.7}$$

where y_ξ^ω minimizes subproblem ω given $x = \xi^k$; see (3.1). An unbiased estimate of the variance of $\tilde{\theta}_\xi^k$ about its true mean,

$$\theta_\xi^k = \sum_{\omega \in \Omega} p^\omega f y_\xi^\omega \tag{4.8}$$

is also calculated on iteration k:

$$\tilde{\sigma}_k^2 = \frac{1}{|S^k|(|S^k| - 1)} \sum_{\omega \in S^k} (f y_\xi^\omega - \tilde{\theta}^k)^2, \quad \xi = \xi^k. \tag{4.9}$$

Therefore an unbiased estimate of the first-stage and second-stage costs (given $\xi = \xi^k$) is

$$\tilde{z}_\xi^k = c\xi^k + \tilde{\theta}_\xi^k, \quad \xi = \xi^k \tag{4.10}$$

and $\tilde{\sigma}_k^2$ is an unbiased estimate of the variance of $\tilde{z}_\xi^k$.

At the termination of the stochastic Benders algorithm, we choose $x = \xi^l$ as the first-stage decision, where $l = \arg\min \tilde{z}^k$. However, the minimum of several minima $\tilde{z}^k$ is no longer an unbiased estimate of z^l. To obtain an unbiased estimate of z^l, we re-estimate $\tilde{\theta}_\xi^l$, $\tilde{z}_\xi^k$ and $\tilde{\sigma}_l^2$ using a new independently drawn sample $\hat{S}^l$. In our applications, sample sizes are $|S| = 100$ or more, so it is reasonable (by the central limit theorem) to assume $\tilde{z}_\xi^l$ is normally distributed. Therefore an α upper confidence interval for z^* is

$$\text{Prob}(\tilde{z}_\xi^l + t\sigma_l \geq z^*) \geq \alpha, \tag{4.11}$$

where t is defined as

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{t^2}{2}} dt = \alpha. \tag{4.12}$$

4.4. Lower Confidence Interval for z^*

In order to know how close the objective associated with the first-stage decision $x = \xi^l$ is to the true $\min z = z^*$, we estimate a lower confidence level for z^*. All feasible solutions to (2.2) satisfy the stochastic cuts (4.4), $-\tilde{G}^k x + \theta \geq \tilde{g}^k - \tilde{\epsilon}^k$, where $\tilde{\epsilon}^k = \frac{1}{|S^k|} \sum_{\omega \in S^k} f y^\omega - \theta$.

In particular, the optimum z^*, x^*, θ^*, ϵ_*^k, $k = 1, \ldots, K$ satisfies the following system of inequalities:

$$
\begin{array}{rcccl}
cx^* & + & \theta^* & = & z^* \\
Ax^* & & & = & b \\
-\tilde{G}^1 x^* & + & \theta^* & \geq & \tilde{g}^1 - \tilde{\epsilon}_*^1 \\
-\tilde{G}^2 x^* & + & \theta^* & \geq & \tilde{g}^2 - \tilde{\epsilon}_*^2 \\
& & \vdots & \vdots & \vdots \\
-\tilde{G}^K x^* & + & \theta^* & \geq & \tilde{g}^K - \tilde{\epsilon}_*^K \\
x^* & & & \geq & 0,
\end{array}
\tag{4.13}
$$

where for each k,

$$
\tilde{\epsilon}_*^k = \frac{1}{|S^k|} \sum_{\omega \in S^k} f y_*^\omega - \theta^*
\tag{4.14}
$$

is a sample average about the true mean θ^* drawn from the same distribution of $f y_*^\omega$. We now show how the optimal dual multipliers of the pseudo master (K), depicted in (4.15) below, can be applied to (4.13) to derive a lower bound for the optimal solution z^*.

Pseudo Master Problem (K)

$$
\begin{array}{rlrcccl}
\min \tilde{v} & = & & cx & + & \theta & \\
\tilde{\rho}: & & & Ax & & & = b \\
\tilde{\lambda}^1: & & & -\tilde{G}^1 x & + & \theta & \geq \tilde{g}^1 \\
\tilde{\lambda}^2: & & & -\tilde{G}^2 x & + & \theta & \geq \tilde{g}^2 \\
& & & \vdots & & \vdots & \vdots \\
\tilde{\lambda}^K: & & & -\tilde{G}^K x & + & \theta & \geq \tilde{g}^K \\
& & & x & & & \geq 0.
\end{array}
\tag{4.15}
$$

Let $\tilde{\rho}, \tilde{\lambda}^1, \ldots, \tilde{\lambda}^K$ be the optimal dual variables of (4.15) and let $\tilde{v}^* = \min \tilde{v}$. These satisfy

$$
\tilde{v}^* = \tilde{\rho} b + \sum_{k=1}^{K} \tilde{\lambda}^k \tilde{g}^k
\tag{4.16}
$$

$$
\tilde{\rho} A - \sum_{k=1}^{K} \tilde{\lambda}^k \tilde{G}^k + \gamma = c, \quad \gamma \geq 0
\tag{4.17}
$$

$$
\sum_{k=1}^{K} \tilde{\lambda}^k = 1
\tag{4.18}
$$

$$
\tilde{\lambda}^k \geq 0, \quad k = 1, \ldots, K.
\tag{4.19}
$$

Applying these same multipliers to the corresponding relations in (4.13) and subtracting from the first relation of (4.13), we obtain

$$
0 \leq \gamma x^* \leq z^* - (\tilde{\rho} b + \sum_{k=1}^{K} \tilde{\lambda}^k \tilde{g}^k) + \sum_{k=1}^{K} \tilde{\lambda}^k \tilde{\epsilon}_*^k.
\tag{4.20}
$$

9

Dropping $\gamma x^* \geq 0$, substituting $\bar{v}^*$ from (4.16) for the middle term in parentheses, and rearranging terms, we obtain a lower bound for the optimal objective z^* of (2.2):

$$\bar{v}^* - \Delta \leq z^*, \quad \Delta = \sum_{k=1}^{K} \bar{\lambda}^k \bar{\epsilon}_*^k, \tag{4.21}$$

where $\bar{\lambda}^k$ are optimal dual multipliers of the pseudo master (K), $\bar{\lambda}^k \geq 0$, $\sum_{k=1}^{K} \bar{\lambda}^k = 1$, and $\bar{\epsilon}_*^k$ for $k = 1, \ldots, K$ are the deviations from their true mean θ^* of the sample means of samples S^k of size $|S^k|$ of terms fy_*^ω.

4.5. The Distribution of the Error Term Δ

Our goal is to derive an approximation of the probability distribution for the random variable $\Delta = \sum_{k=1}^{K} \bar{\lambda}^k \bar{\epsilon}_*^k$. In fact we develop two such distributions upper bounding Δ; the first we call "worst-case bound" and the second we call "conservative bound". For this analysis we need the distributions of the error terms $\bar{\epsilon}_*^k$.

By definition, σ_*^2, the variance of the population of the optimal second-stage costs fy_*^ω, $\omega \in \Omega$ corresponding to the optimal first-stage solution x^*, is given by

$$\sigma_*^2 = \sum_{\omega \in \Omega} p^\omega (fy_*^\omega - \theta^*)^2. \tag{4.22}$$

We assume that all sample sizes used in the various iterations are equal to N, i.e.,

$$|S^k| = N, \quad k = 1, \ldots, K. \tag{4.23}$$

Since we have fixed $x = x^*$ at the optimum, the random samples S^k of size N for computing the means $\frac{1}{|S^k|} \sum_{\omega \in S^k} fy_*^\omega$ are *all drawn from the same distribution of values* fy_*^ω. Therefore each variance of $\bar{\epsilon}_*^k$, the expected value of $(\frac{1}{|S^k|} \sum_{\omega \in S^k} fy_*^\omega - \theta^*)^2$, is the same:

$$\mathrm{var}(\bar{\epsilon}_*^k) = \frac{\sigma_*^2}{N}, \quad k = 1, \ldots, K. \tag{4.24}$$

Note that this is a theoretical result because we do not know x^* and hence cannot generate sampled values of fy_*^ω to estimate σ_*^2. We estimate σ_*^2 by setting it equal to the estimated variance of fy_ξ^ω, where $\xi = \xi^l$:

$$\text{estimated } \sigma_*^2 = \bar{\sigma}_l^2 = \frac{1}{N-1} \sum_{\omega \in S} (fy_\xi^\omega - \frac{1}{N} fy_\xi^\omega)^2, \quad \xi = \xi^l, \quad S = S^l. \tag{4.25}$$

Each error term $\bar{\epsilon}_*^k$ is the difference of the average of N independently drawn observations— with repacements *from the same distribution* of fy_*^ω minus its true mean θ^*. Because sample sizes in our applications typically satisfy $N \geq 100$ and are often several hundred, it is reasonable (by the central limit theorem) to assume $\bar{\epsilon}_*^k$ are normally distributed:

$$\bar{\epsilon}_*^k \sim N(0, \frac{\sigma_*}{\sqrt{N}}) = \frac{\sigma_*}{\sqrt{N}} N(0,1). \tag{4.26}$$

Our goal now is to determine an upper bound for $\Delta = \sum_{k=1}^{K} \bar{\lambda}^k \bar{\epsilon}_*^k$, say Δ^α, such that $\mathrm{Prob}(\Delta \leq \Delta^\alpha) \geq \alpha$. Since we do not know the distribution of Δ, we cannot compute Δ^α directly. Instead, we determine two distributions, a worst-case distribution Δ_W and a conservative distribution Δ_C, each of which dominates the Δ distribution, and find an α-point for each distribution.

4.6. Worst-Case Lower Bound for z^*

Since $\bar{\lambda}^k \geq 0$ and $\sum_{k=1}^{K} \bar{\lambda}^k = 1$, the worst-case upper bound for Δ is Δ_W,

$$\Delta_W = \max_k \bar{\epsilon}_*^k \geq \sum_{k=1}^{K} \bar{\lambda}^k \bar{\epsilon}_*^k = \Delta. \tag{4.27}$$

We know from (4.26) that $\Delta_W = \max \bar{\epsilon}_*^k$ is a random variable distributed as the maximum of K independent normal $(\sigma_*/\sqrt{N})N(0,1)$ variates. We determine the point Δ_W^α on this distribution such that

$$\text{Prob}(\Delta_W \leq \Delta_W^\alpha) \geq \alpha, \tag{4.28}$$

where

$$\Delta_W^\alpha = \bar{v}^* - t_K \frac{\sigma_*}{\sqrt{N}}, \tag{4.29}$$

and t_K is defined according to

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t_K} e^{-\frac{t^2}{2}} dt = \alpha^{\frac{1}{K}}. \tag{4.30}$$

Since $\Delta \leq \Delta_W$ by (4.27),

$$\text{Prob}(\Delta \leq \Delta_W^\alpha) \geq \alpha, \tag{4.31}$$

and

$$\text{Prob}(\bar{v}^* - \Delta \geq \bar{v}^* - \Delta_W^\alpha) \geq \alpha. \tag{4.32}$$

Since $z^* \geq \bar{v}^* - \Delta$ by (4.21),

$$\text{Prob}(\bar{v}^* - \Delta_W^\alpha \leq \bar{v}^* - \Delta \leq z^*) \geq \alpha. \tag{4.33}$$

We call the lower bound $(\bar{v}^* - \Delta_W^\alpha)$ the probabilistic worst-case α lower bound for z^*.

4.7. Conservative Lower Bound for z^*

The conservative lower bound is based on the observation that the $\bar{\lambda}^k$ and $\bar{\epsilon}_*^k$ in the expression $\Delta = \sum_{k=1}^{K} \bar{\lambda}^k \bar{\epsilon}_*^k$ tend to be positively correlated. This can be seen intuitively. It is reasonable to expect, in forming the pseudo master problem, that dropping the term $\bar{\epsilon}_*^k$ when it is large and positive will give rise to a tighter constraint k, and dropping it when negative will slacken the constraint. This positive correlation of $\bar{\lambda}^k$ with $\bar{\epsilon}_*^k$ is very evident in practical problems; see the empirical evidence presented in Figures 1 and 2.

In order to obtain Δ_C (the conservative upper bound for Δ) and hence $\bar{v}^* - \Delta_C$ (the conservative lower bound for z^*), we reorder the $\bar{\lambda}^k$ from high to low and relabel them

$$\bar{\mu}^1 \geq \bar{\mu}^2 \geq \cdots \geq \bar{\mu}^K \geq 0. \tag{4.34}$$

We also order the $\bar{\epsilon}_*^k$ from high to low and relabel them

$$\bar{\delta}^1 \geq \bar{\delta}^2 \geq \cdots \geq \bar{\delta}^K. \tag{4.35}$$

It is then obvious (and straightforward to prove) that

$$\Delta = \sum_{k=1}^{K} \bar{\lambda}^k \bar{\epsilon}_*^k \leq \sum_{k=1}^{K} \bar{\mu}^k \bar{\delta}^k. \tag{4.36}$$

The difference $\sum_{k=1}^{K} \bar{\mu}^k \bar{\delta}^k - \Delta$ depends on how correlated $\bar{\lambda}^k$ and $\bar{\epsilon}_*^k$ are; the higher the correlation, the smaller the difference. We have observed empirically that after $\bar{\lambda}^k$ and $\bar{\epsilon}_*^k$ are reordered, $\bar{\mu}^k$ and $\bar{\delta}^k$ are no longer correlated; accordingly, for the development of the conservative lower bound, *we make the assumption that $\bar{\mu}^k$ and $\bar{\delta}^k$ are independent.*

Notice that the $\bar{\delta}^k$'s are the distribution of instances of K normal $(\sigma_*/\sqrt{N})N(0,1)$ variates ordered from high to low. We do not know which particular instance of the K ordered normal deviates $\bar{\delta}^k$ formed the products with the ordered $\bar{\mu}^k$ and were then summed. We can, however, by our assumption of independence, view

$$\Delta_C = \sum_{k=1}^{K} \bar{\mu}^k \bar{\delta}^k, \quad \text{for the observed value of } \bar{\mu}^k, \tag{4.37}$$

as a random variable, generated by all possible instances of ordered sets of K ordered normal deviates $\bar{\delta}^k$. The distribution of Δ_C can be determined, either by numerical integration or by Monte Carlo approximation. We approximate it by generating many K independent $(\sigma_*/\sqrt{N})N(0,1)$ normal deviates, and reordering them into many ordered instances of $\bar{\delta}^1, \ldots, \bar{\delta}^K$. Substituting each of these ordered instances into (4.37) we obtain an approximate distribution that we call the *probabilistic* upper bound distribution for Δ_C. We determine Δ_C^α, the α percent point of this distribution, for which

$$\text{Prob}(\Delta_C \leq \Delta_C^\alpha) \geq \alpha. \tag{4.38}$$

Since $\Delta \leq \sum_{k=1}^{K} \bar{\mu}^k \bar{\delta}^k = \Delta_C$ by (4.36) and (4.37),

$$\text{Prob}(\Delta \leq \Delta_C^\alpha) \geq \alpha, \tag{4.39}$$

and

$$\text{Prob}(\bar{v}^* - \Delta \geq \bar{v}^* - \Delta_C^\alpha) \geq \alpha. \tag{4.40}$$

Since $z^* \geq \bar{v}^* - \Delta$ by (4.21),

$$\text{Prob}(\bar{v}^* - \Delta_C^\alpha \leq \bar{v}^* - \Delta \leq z^*) \geq \alpha. \tag{4.41}$$

We call the lower bound $(\bar{v}^* - \Delta_C^\alpha)$ the probabilistic α lower bound for z^*.

To summarize, the main steps are as follows:

- Obtain a realization of $\bar{\rho}$, $\bar{\lambda}^k$, $k = 1, \ldots, K$ and $\bar{v}^* = \bar{\rho}b + \sum_{k=1}^{K} \bar{\lambda}^k \bar{g}^k$ by solving the *pseudo master problem* (4.15) for $\bar{v}^* = \min \bar{v}$.

- Order $\bar{\lambda}^k$ from high to low to obtain $\bar{\mu}^k$, $k = 1, \ldots, K$.

- Obtain the distribution Δ_C by generating many sets of K independent observations of the distribution $(\sigma_*/\sqrt{N})N(0,1)$, reordering them from high to low to form the $\bar{\delta}^1, \ldots, \bar{\delta}^K$ observations, and substituting into $\Delta_C = \sum_{k=1}^{K} \bar{\mu}^k \bar{\delta}^k$ to obtain instances of Δ_C.

12

- Determine the Δ_C^α point of the distribution Δ_C by Monte Carlo and construct the confidence interval $\text{Prob}(\bar{v}^* - \Delta_C^\alpha \leq \bar{v}^* - \Delta \leq z^*) \geq \alpha$, where the estimate of the lower bound for z^* is $\bar{v}^* - \Delta_C^\alpha$.

Comment

When generating the distribution Δ_C for approximating $\sum_{k=1}^K \bar{\mu}^k \bar{\delta}^k$, we are assuming that $\bar{\mu}^k$ and $\bar{\delta}^k$ are uncorrelated. Empirically, once we reorder, that seems to be the case. Even if there is some correlation remaining, it cannot have much effect because the $\bar{\delta}^k$ are confined to a narrow range of values.

What deserves more study, and is a goal for future research, is how best to estimate σ_*. If ξ^l is a nearly optimal solution, $\bar{\sigma}_l$ should be a good approximation to the true value of σ_*. Alternatively, $\bar{\sigma}_{K+1}$, the estimated variance of $\bar{z}_\xi^{K+1}$ corresponding to $\xi = \xi^{K+1}$, the optimal solution of the pseudo master problem (K), could be used as a good estimate for σ_*. As a third way of estimating σ_*, we suggest computing the solution

$$\hat{\xi} = \sum_{k=1}^K \bar{\lambda}^k \xi^k \tag{4.42}$$

and estimating the variance of $\bar{\theta}_\xi = \frac{1}{|S|} \sum_{\omega \in S} f y_\xi^\omega$, $\xi = \hat{\xi}$ around its true mean $\theta_\xi = \sum_{\omega \in \Omega} f p^\omega y_\xi^\omega$, $\xi = \hat{\xi}$, as

$$\bar{\sigma}_\xi^2 = \frac{1}{|S|(|S|-1)} \sum_{\omega \in S} (f y_\xi^\omega - \bar{\theta}_\xi)^2, \quad \xi = \hat{\xi}, \tag{4.43}$$

using an independent sample $S = \hat{S}$. We recommend in practice using $\bar{\sigma}_l$ as an estimator for σ_*. Nevertheless, to be on the safe side, it may be a good idea to inflate the estimated σ_*, say by 5%.

5. Test

In order to test the theory, we used a number of problems discussed in the literature (see, e.g., [29]), and also problems we designed in-house. For each of these, the number W of scenarios ω was small enough to allow us to compute the optimal solution exactly. For the test we solved the universe problem to obtain the optimal universe solution z^*, x^*, y_*^ω, and the true population variance σ_*^2. We also recorded the sequence of solutions $x = \xi^k$, $k = 1, \ldots, K$ that led to the optimal solution. Actually, we reserved room for only a fixed number of cuts, say $K = K_{cut}$. If the number of iterations for solving the universe problem exceeded K_{cut}, we replaced previously generated cuts. The solutions ξ^k, $k = 1, \ldots, K$ then represented the solutions corresponding to the cutting planes stored in the master problem when it terminated at the optimal solution.

To carry out the test we used the recorded solutions ξ^k, $k = 1, \ldots, K$ to generate the stochastic cuts. For each of these, we computed the true correction terms $\bar{\epsilon}_*^k$. Based on the true correction terms we computed the true value of the lower bound $\bar{v}^* - \Delta$, where $\bar{v}^* = \bar{\rho} b + \sum_{k=1}^K \bar{\lambda}^k \bar{g}^k$ is the optimal objective of the pseudo master problem (K), and we

13

computed the less tight but true lower bound $\bar{v}^* - \Delta_C$, based on the reordering of the $\bar{\lambda}^k$ and $\bar{\epsilon}_*^k$.

We then computed empirically an upper bound distribution Δ_C by repeating R times the sampling of K observations from the normal distribution $(\sigma_*/\sqrt{N})N(0,1)$, and ordering the resulting values to obtain R observations of $\Delta_C = \sum_{k=1}^K \bar{\mu}^k \bar{\delta}^k$. We used this empirical distribution of Δ_C to calculate Δ_C^α and used Δ_C^α to calculate $\bar{v}^* - \Delta_C^\alpha$, a one-sided α lower confidence interval for z^*.

For illustration purposes we describe in detail the results of applying the methodology to the test problem APL1P, which is a small electric power expansion planning problem with uncertainty in three demands and in the availability of two generators; see Infanger (1992) [31]. The master problem has 3 rows and 3 columns and each second-stage scenario has 6 rows and 9 columns. The total number of scenarios is $W = 1280$. The optimal values of x^* and y_*^ω resulted in $z^* = 24642.3$, $\theta^* = 13513.7$ and $\sigma_* = 4808.8$. For the experiment we estimated $K = 20$ cuts using a sample size of $N = 100$. The optimal dual multipliers for the pseudo master problem $\bar{\lambda}^k$, and the values of $\bar{\epsilon}_*^k$, as well as the ordered values $\bar{\mu}^k$ and $\bar{\delta}^k$ for $k = 1, \ldots, 20$, are displayed in Table 1. The lower bound for z^* before reordering the $\bar{\lambda}^k$ and $\bar{\epsilon}_*^k$ is

$$z^* \geq \bar{v}^* - \sum_{k=1}^K \bar{\lambda}^k \bar{\epsilon}^k = 25188.6 - 546.3 = 24642.3, \tag{5.1}$$

which verifies that $\bar{v}^* - \sum_{k=1}^K \bar{\lambda}^k \bar{\epsilon}^k$ is, as stated earlier, a true lower bound (24642.3) for z^*. It happened in this case that the 20 corrected cuts were sufficient to determine the true minimum $z^* = 24642.3$. We then reordered the $\bar{\lambda}^k$ to obtain $\bar{\mu}^k$, and reordered the $\bar{\epsilon}_*^k$ to obtain $\bar{\delta}^k$. Clearly,

$$z^* \geq \bar{v}^* - \sum_{k=1}^K \bar{\mu}^k \bar{\delta}^k = 25188.6 - 559.1 = 24629.5, \tag{5.2}$$

which verifies that $\bar{v}^* - \sum_{k=1}^K \bar{\mu}^k \bar{\delta}^k$ is a smaller true lower bound for z^*. Using the largest observed value $\max_k \bar{\epsilon}_*^k$, the worst-case lower bound for z^* is

$$z^* \geq \bar{v}^* - \max_k \bar{\epsilon}_*^k = 25188.6 - 636.6 = 24552.0. \tag{5.3}$$

So far, all the evaluations for the example were exact; there were no estimates. We now discuss how we estimated the term $\sum_{k=1}^K \bar{\mu}^k \bar{\delta}^k$. We know that the $\bar{\delta}^k$ are a reordering of $\bar{\epsilon}_*^k$, and that these $\bar{\delta}^k$, $k = 1, \ldots, K$ are randomly drawn (approximately) $(\sigma_*/\sqrt{N})N(0,1)$ deviates. What we did was draw $R = 500$ samples of $K = 20$ independent normally distributed $(\sigma_*/\sqrt{N})N(0,1)$ deviates. We ordered the $\bar{\epsilon}_*^k$ in each of the sets of twenty from high to low, and for each reordered set i of the twenty $\bar{\delta}_k$ we computed $\sum_{k=1}^{20} \bar{\mu}^k \delta^k = \Delta_C^i$. Next we generated the distribution of Δ_C^i, $i = 1, \ldots, 500$ cases of Δ_C and computed the point $\Delta_C^{0.95}$ as the upper bound estimate of Δ_C, where $\Delta_C^{0.95}$ is the value of Δ_C that was exceeded in 95% of the cases. Hence

$$z^* \geq \bar{v}^* - \Delta_C^{0.95} = 25188.6 - 1143.7 = 24044.9, \tag{5.4}$$

where $\sigma_* = 4808.90$. In practice, of course, we do not have the true value of σ_*. Using $\bar{\sigma}_{K-2} = 4449.8$ as an estimate of σ_* (to simulate a situation where the optimum x^* has not

been reached yet), we obtain

$$z^* \geq \bar{v}^* - \Delta_C^{0.95}|_{\bar{\sigma}_{K-2}} = 25188.6 - 1058.3 = 24130.3, \tag{5.5}$$

which under-bounds the true value of $z^* = 24642.3$ by 2.1%. We calculated the worst-case lower bound using the true value of σ^* as

$$z^* \geq \bar{v}^* - \Delta_W^{0.95} = 25188.6 - 1345.9 = 23842.7, \tag{5.6}$$

and using the estimate $\bar{\sigma}_{K-2} = 4449.8$ as

$$z^* \geq \bar{v}^* - \Delta_W^{0.95}|_{\bar{\sigma}_{K-2}} = 25188.6 - 1245.4 = 23943.2. \tag{5.7}$$

Table 1: Test results for problem APL1P

cut(k)	λ^k	$\bar{\epsilon}_*^k$	$\bar{\mu}^k$	δ^k
1	0.558	636.757	0.558	636.757
2	0.000	-656.181	0.276	504.965
3	0.276	504.965	0.166	388.265
4	0.000	-500.738	0.000	339.834
5	0.000	- 71.361	0.000	316.619
6	0.000	-519.535	0.000	311.268
7	0.000	-457.835	0.000	114.471
8	0.000	-181.275	0.000	96.202
9	0.000	388.265	0.000	- 48.166
10	0.000	96.202	0.000	- 71.361
11	0.000	-362.272	0.000	-117.495
12	0.000	-117.495	0.000	-181.275
13	0.000	-929.552	0.000	-331.046
14	0.000	339.834	0.000	-362.272
15	0.000	114.471	0.000	-457.835
16	0.166	311.268	0.000	-500.738
17	0.000	-601.078	0.000	-519.535
18	0.000	- 48.166	0.000	-601.078
19	0.000	316.619	0.000	-656.181
20	0.000	-331.046	0.000	-929.552
	$\sum \lambda^k =$ 1	$\sum \lambda^k \bar{\epsilon}_*^k =$ 546.309	$\sum \bar{\mu}^k =$ 1	$\sum \bar{\mu}^k \delta^k =$ 559.086

To test the coverage of the computed lower bound, we repeatedly (say 100 times) ran the test described above with different seeds. That is, for each replication we used the sequence of solutions ξ^k, $k = 1, \ldots, K$ to compute cutting planes using independent samples S^k, and solved the corresponding pseudo master problem (K) to obtain the optimal $\bar{p}$ and $\bar{\lambda}^1, \ldots, \bar{\lambda}^K$. We computed for each of these replications the minimum objective function

15

value of the pseudo master problem $\tilde{v}^* = \bar{\rho}b + \sum_{k=1}^{K} \tilde{\lambda}^k \tilde{g}^k$; a true lower bound $\tilde{v}^* - \sum_{k=1}^{K} \bar{\mu}^k \tilde{\delta}^k$, based on the true ordered values of the correction terms $\tilde{\epsilon}_*^k$; and the 95% point $\tilde{v}^* - \Delta_C^{0.95}$, based on the reordering procedure. We also computed and recorded $\tilde{v}^* - \Delta_W^{0.95}$, the 95% point of the worst-case lower bound distribution $\tilde{v}^* - \Delta_W$. Based on the 100 replications of the experiment we estimated the coverage of the probabilistic lower bounds by computing the percent of the 100 cases in which the probabilistic lower bound was actually less than or equal to the true optimal objective value z^*. The results for the test problem APL1P are represented in Figure 1, which displays the values of $\tilde{v}^*$, $\tilde{v}^* - \sum_{k=1}^{K} \bar{\mu}^k \tilde{\delta}^k$, $\tilde{v}^* - \Delta_C^{0.95}$, and $\tilde{v}^* - \Delta_W^{0.95}$ respectively as a histogram of the 100 values from the replications. Instead of the actual values, we report the quantities as percent deviation from the true objective function value z^*. The true objective function value in the graph is labeled as 0.

The curve labeled "v-pseudo" represents the histogram of $\tilde{v}^*$, the minimum objective of the pseudo master problem. One can see that most of the 100 replications had $\tilde{v}^*$ values larger than the true optimal objective z^*. It clearly reveals the bias of the optimal objective of the pseudo master $\tilde{v}^*$ as an under-estimator for z^*. The curve labeled "true" represents the histogram of the true lower bound $\tilde{v}^* - \sum_{k=1}^{K} \bar{\mu}^k \tilde{\delta}^k$ based on the reordering of $\tilde{\lambda}^k$ and $\tilde{\epsilon}_*^k$. There was (as predicted by the theory) no instance in which this true lower bound exceeded the true optimal objective. The two observations at the zero point of the curve are two observations in the interval between -1 and 0. The curve labeled "conservative" represents the histogram (constructed from the 100 replications) of $\tilde{v}^* - \Delta_C^{0.95}$, the 95% point of the $\tilde{v}^* - \Delta_C^{0.95}$ distribution. The curve shows that $\Delta_C^{0.95}$ is a conservative estimate of Δ. The coverage of the $\tilde{v}^* - \Delta_C^{0.95}$ turned out to be 96%. Finally, the curve labeled "worst-case" represents the histogram of $\tilde{v}^* - \Delta_W^{0.95}$, the 95% point of the probabilistic worst-case lower bound. As expected, the $\tilde{v}^* - \Delta_W^{0.95}$ values turned out to be smaller than the $\tilde{v}^* - \Delta_C^{0.95}$ values, which makes the probabilistic worst-case lower bound $\tilde{v}^* - \Delta_W^{0.95}$ an even smaller but nevertheless tight lower bound for z^*. Its coverage turned out to be 96%.

With the other test problems, we obtained very similar results. As a representative example, we show the results for the test problem STORM described in Mulvey and Ruszczynski (1992) [41]. The problem is a freight-scheduling problem with uncertainty in demands. The version we used had a total of 40 universe scenarios. The size of the master problem was 126 rows and 289 columns, and each of the 40 subproblems had 347 rows and 769 columns. The optimal objective of the universe problem was $z^* = 15.569 \, 10^6$ and the variance σ_* was 89159.5. For the experiment we used $K = 30$ cutting planes estimated with a sample size of $N = 20$. Figure 2 gives the results. The Figure looks very similar to the one for APL1P, except the pseudo cutting planes are better estimates of the true cutting planes. Looking at the distribution labeled "v-pseudo" one can see the bias of the optimal solution of the pseudo master problem $\tilde{v}^*$ as an estimator for z^*. There was no instance where the true conservative lower bound $\tilde{v}^* - \sum_{k=1}^{K} \bar{\mu}^k \tilde{\delta}^k$ (labeled "true") exceeded the true objective z^*. The 95% point $\tilde{v}^* - \Delta_C^{0.95}$ of the estimated conservative lower bound distribution (labeled "conservative") gave a conservative estimate but nevertheless an excellent lower bound for z^*. The coverage of $\tilde{v}^* - \Delta_C^{0.95}$ turned out to be 100%. The point $\Delta_W^{0.95}$ of the worst-case lower bound distribution (labeled "worst-case") gave a smaller but nevertheless tight lower bound for z^*. The coverage of the worst-case lower bound $\tilde{v}^* - \Delta_W^{0.95}$ proved to be 100%.

We tested further with the following problems: PGP2, CEP1, and SCTAP1, all described in [29]. PGP2 (Louveaux and Smeers (1988) [38]) is a small electric power capacity expansion

planning test problem (master: 2 rows and 4 columns, sub: 7 rows and 16 columns) with uncertain parameters in three demands. The number of universe scenarios was $W = 1280$. CEP1 (Higle and Sen (1990) [25]) is a small machine capacity expansion planning problem (master: 9 rows and 8 columns, sub: 7 rows and 15 columns) with uncertain parameters in the right-hand side. The number of universe scenarios was $W = 1000$. SCTAP1 (Ho (1980) [28]) is a traffic assignment problem (master: 30 rows and 48 columns, sub: 60 rows and 96 columns) with stochastic right-hand sides. The number of universe scenarios was $W = 864$. The values of z^* and σ_* for the various test problems, as well as the sample sizes N and the number of cutting planes K used for the experiments, are given in Table 2. The coverage results of the test problems are summarized in Table 3 based on using the true value of σ_*, and in Table 4 based on using the estimate $\tilde{\sigma}_{K-2}$ as approximation for σ_* (to simulate a situation where the optimum x^* has not been reached yet).

Table 2: Test data

problem	z^*	σ_*	N	K
APL1P	24642.3	4808.8	100	20
PGP2	447.3	77.60	100	20
CEP1	355159.5	420458.4	100	6
STORM	$15.569 \cdot 10^6$	89159.5	20	30
SCTAP1	248.5	24.01	100	3

Table 3: Coverage test results (lower bounds calculated using σ_*)

problem	$\bar{v}^* - \Delta_C^{0.95}$	coverage	$\bar{v}^* - \Delta_W^{0.95}$	coverage
	%	%	%	%
APL1P	-1.78	95	-2.39	94
PGP2	-2.94	97	-3.51	98
CEP1	-12.12	95	-12.27	94
STORM	-0.40	100	-0.58	100
SCTAP1	-0.9	99	-0.91	99

Table 4: Coverage test results (lower bounds calculated using $\bar{\sigma}_{K-2}$)

problem	$\bar{v}^* - \Delta_C^{0.95}$	coverage	$\bar{v}^* - \Delta_W^{0.95}$	coverage
	%	%	%	%
APL1P	-1.80	96	-2.41	96
PGP2	-2.35	91	-2.84	94
CEP1	-12.07	92	-12.23	92
STORM	-0.40	98	-0.58	99
SCTAP1	-0.9	98	-0.91	98

References

[1] Beale E.M.L. (1955): On Minimizing a Convex Function Subject to Linear Inequalities, *J. Roy. Stat. Soc.* 17b, 173–184.

[2] Benders, J.F. (1962): Partitioning Procedures for Solving Mixed-Variable Programming Problems, *Numerische Mathematic* 4, 238–252.

[3] Birge, J.R. (1985): Decomposition and Partitioning Methods for Multi-Stage Stochastic Linear Programming, *Operations Research* 33, 989–1007.

[4] Birge, J.R. and Louveaux F.V. (1985): A Multicut Algorithm for Two-Stage Linear Programs, Technical Report, Department of IOE, University of Michigan, Ann Arbor, MI.

[5] Birge, J.R. and Wets, R.J. (1986): Designing Approximation Schemes for Stochastic Optimization Problems, in Particular For Stochastic Programs with Recourse, *Math. Progr. Study* 27, 54–102.

[6] Birge, J.R. and Wets, R.J. (1987): Computing Bounds for Stochastic Programming Problems by Means of a Generalized Moment Problem, *Mathematics of Operations Research* 12, 149–162.

[7] Birge, J.R. and Wets, R.J. (1989): Sublinear Upper Bounds For Stochastic Programs with Recourse, *Mathematical Programming* 43, 131–149.

[8] Birge, J.R. and Wallace, S.W. (1988): A Separable Piecewise Linear Upper Bound for Stochastic Linear Programs, *SIAM J. Control and Optimization*, 26, 3.

[9] Dantzig, G.B. (1955): Linear Programming Under Uncertainty, *Management Science* 1, 197–206.

[10] Dantzig, G.B. and Glynn, P.W. (1990): Parallel Processors for Planning Under Uncertainty, *Annals of Operations Research* 22, 1–21.

[11] Dantzig, G.B., Glynn, P.W., Avriel, M., Stone, J., Entriken, R., Nakayama, M. (1989): Decomposition Techniques for Multi-Area Generation and Transmission Planning under Uncertainty, EPRI Report 2940-1.

[12] Dantzig, G.B., Ho, J.K. and Infanger, G. (1991): Solving Stochastic Linear Programs on a Hypercube Multicomputer, Report SOL 91-10, Department of Operations Research, Stanford University, Stanford, CA.

[13] Dantzig, G.B. and Infanger, G.(1991): Large-Scale Stochastic Linear Programs: Importance Sampling and Benders Decomposition, Report SOL 91-4, Department of Operations Research, Stanford University, Stanford, CA.

[14] Dantzig, G.B., and Infanger, G. (1993): Multi-Stage Stochastic Linear Programs for Portfolio Optimization, *Annals of Operations Research* 45, 59–76.

[15] Dantzig, G.B. and Madansky M. (1961): On the Solution of Two-Staged Linear Programs under Uncertainty, *Proc 4th Berkeley Symp. on Mathematical Statistics and Probability* I, ed. J. Neyman, 165–176.

[16] Davis, P.J., and Rabinowitz, P. (1984): *Methods of Numerical Integration*, Academic Press, London.

[17] Edmundson, H.P. (1956): Bounds on the Expectation of a Convex Function of a Random Variable, Paper 982, Rand Corporation, Santa Monica, CA.

[18] Ermoliev, Y. (1983): Stochastic Quasi-gradient Methods and Their Applications to Systems Optimization, *Stochastics* 9, 1–36.

[19] Ermoliev,Y. and R.J. Wets Eds. (1988): *Numerical Techniques for Stochastic Optimization*, Springer Verlag.

[20] Frauendorfer, K. (1988): Solving SLP Recourse Problems with Arbitrary Multivariate Distributions – The Dependent Case, *Mathematics of Operations Research*, Vol. 13, No. 3, 377–394.

[21] Frauendorfer K. and Kall, P. (1988): Solving SLP Recourse Problems with Arbitrary Multivariate Distributions – The Independent Case, *Problems of Control and Information Theory*, Vol 17 (4), 177–205.

[22] Frauendorfer, K. (1992): *Stochastic Two-Stage Programming*, Lecture Notes in Economics and Mathematical Systems 392, Springer Verlag, Berlin, Germany.

[23] Gaivoronski, A. (1988): Implementation of Stochastic Quasigradient Methods, in Ermoliev, Y. and Wets, R.J-B. (eds.): *Numerical Techniques for Stochastic Optimization*, Springer Verlag, Berlin, Germany.

[24] Glynn, P.W. and Iglehart, D.L. (1989): Importance Sampling for Stochastic Simulation, *Management Science*, Vol 35, 1367–1392.

[25] Higle, J.L. and Sen, S. (1990): Finite Master Programs in Regularized Stochastic Decomposition, Report, Department of Systems and Industrial Engineering, University of Arizona, Tucson, AZ.

[26] Higle, J.L. and Sen, S. (1991): Stochastic Decomposition: An Algorithm for Two-Stage Linear Programs with Recourse, *Mathematics of Operations Research* 16/3, 650–669.

[27] Ho, J.K. (1975): Optimal Design of Multi-Stage Structures: A Nested Decomposition Approach, *Computers and Structures* Vol 5, 249–255.

[28] Ho, J.K. (1980): A Successive Linear Optimization Approach to the Dynamic Traffic Assignment Problem, *Transportation Science* 14 (4), 295–305.

[29] Holmes, D. (1994): A Collection of Stochastic Programming Problems, Report 94-11, Department of Industrial and Operations Engineering, The University of Michigan, Ann Arbor, MI.

[30] Huang, C.C., Ziemba, W.T. and Ben-Tal, A. (1977): Bounds on the Expectation of a Convex Function with a Random Variable with Applications to Stochastic Programming, *Operations Research* 25, 315–325.

[31] Infanger, G. (1992): Monte Carlo (Importance) Sampling within a Benders Decomposition Algorithm for Stochastic Linear Programs, *Annals of Operations Research 39, 69-95*.

[32] Infanger, G. (1994): *Planning Under Uncertainty – Solving Large-Scale Stochastic Linear Programs*, Boyd and Fraser, Ferncroft Village, Danvers, MA.

[33] Jensen, J.L. (1906): Sur les fonctions convexes et les inegalites entres les valeurs moyennes, *Acta Mathematica* 30, 175–193.

[34] Kall, P. (1974): Approximations to Stochastic Programs with Complete Fixed Recourse, *Numer. Math.* 22, 333–339.

[35] Kall, P. (1979): Computational Methods for Two Stage Stochastic Linear Programming Problems, *Z. angew. Math. Phys.* 30, 261–271.

[36] Krishna, A.S. (1993): Importance Sampling Techniques for Stochastic Linear Programming, Dissertation, Systems Optimization Laboratory, Department of Operations Research, Stanford University, Stanford, CA.

[37] Lustig, I.J, Mulvey, J.M. and Carpenter, T.J. (1991): Formulating Two-Stage Stochastic Programs for Interior Point Methods, *Operations Research 39*, 757–770.

[38] Louveaux, F. and Smeers, Y. (1988): Optimal Investments for Electricity Generation, a Stochastic Model and a Test Problem, in Ermoliev, Y. and Wets, R.J. (Eds.): Numerical Techniques for Stochastic Optimization, Springer Verlag, Berlin, 33–64.

[39] Madansky, A. (1959): Bounds on the Expectation of a Convex Function of a Multivariate Random Variable, *Annals of Mathematical Statistics* 30, 743–746.

[40] Morton, D.P. (1993): Algorithmic Advances in Multi-stage Stochastic Programming, Dissertation, Systems Optimization Laboratory, Department of Operations Research, Stanford University, Stanford, CA.

[41] Mulvey, J.M. and Ruszczynski, A. (1992): A New Scenario Decomposition Method for Large-Scale Stochastic Optimization, Report SOR-91-19, Department of Civil Engineering and Operations Research, Princeton University, Princeton, NJ.

[42] Murtagh, B.A. and Saunders, M.A. (1983): MINOS User's Guide, SOL 83-20, Department of Operations Research, Stanford University, Stanford CA.

[43] Prékopa, A. (1989): Sharp Bounds on Probabilities Using Linear Programming, *Operations Research.*

[44] Pereira, M.V., Pinto, L.M.V.G., Oliveira, G.C. and Cunha, S.H.F. (1989): A Technique for Solving LP-Problems with Stochastic Right-Hand Sides, CEPEL, Centro del Pesquisas de Energia Electria, Rio de Janeiro, Brazil.

[45] Pereira, M.V., Pinto, L.M.V.G. (1989): Stochastic Dual Dynamic Programming, Technical Note, DEE-PUC/RJ – Catholic University of Rio de Janeiro, Caixa Postal 38063 Gávea, Rio de Janeiro RJ CEP 22452 Brazil.

[46] Rockafellar, R.T. and Wets, R.J-B. (1989): Scenario and Policy Aggregation in Optimization Under Uncertainty, *Mathematics of Operations Research* 16, 119–147.

[47] Ruszczynski, A. (1986): A Regularized Decomposition method for Minimizing a Sum of Polyhedral Functions, *Mathematical Programming* 35, 309–333.

[48] Van Slyke and Wets, R.J. (1969): L-Shaped Linear Programs with Applications to Optimal Control and Stochastic Programming, *SIAM Journal of Applied Mathematics*, Vol 17, 638–663.

[49] Wets R.J. (1984): Programming under Uncertainty: The Equivalent Convex Program, *SIAM J. on Appl. Math.* 14, 89–105.
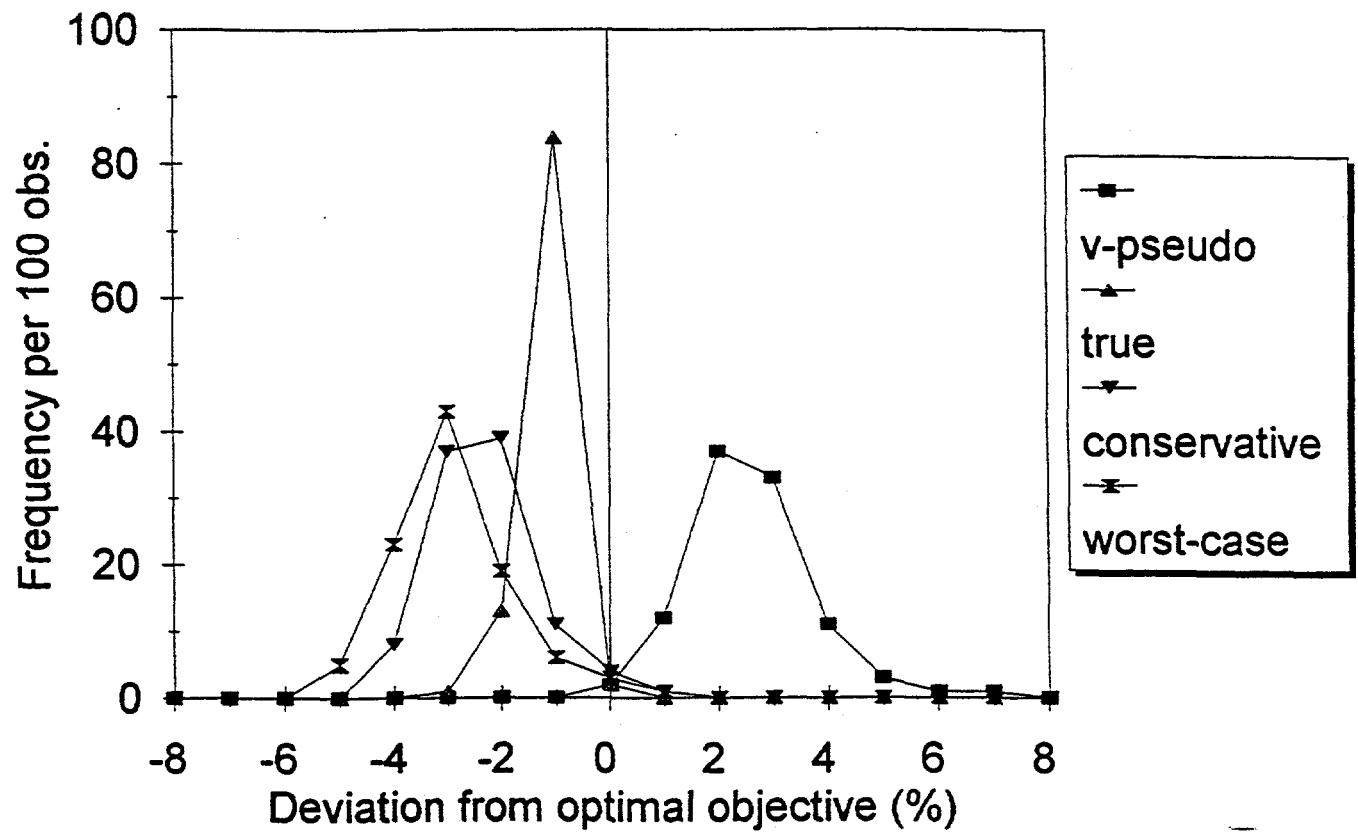
Figure 1: APL1P: lower bound distributions
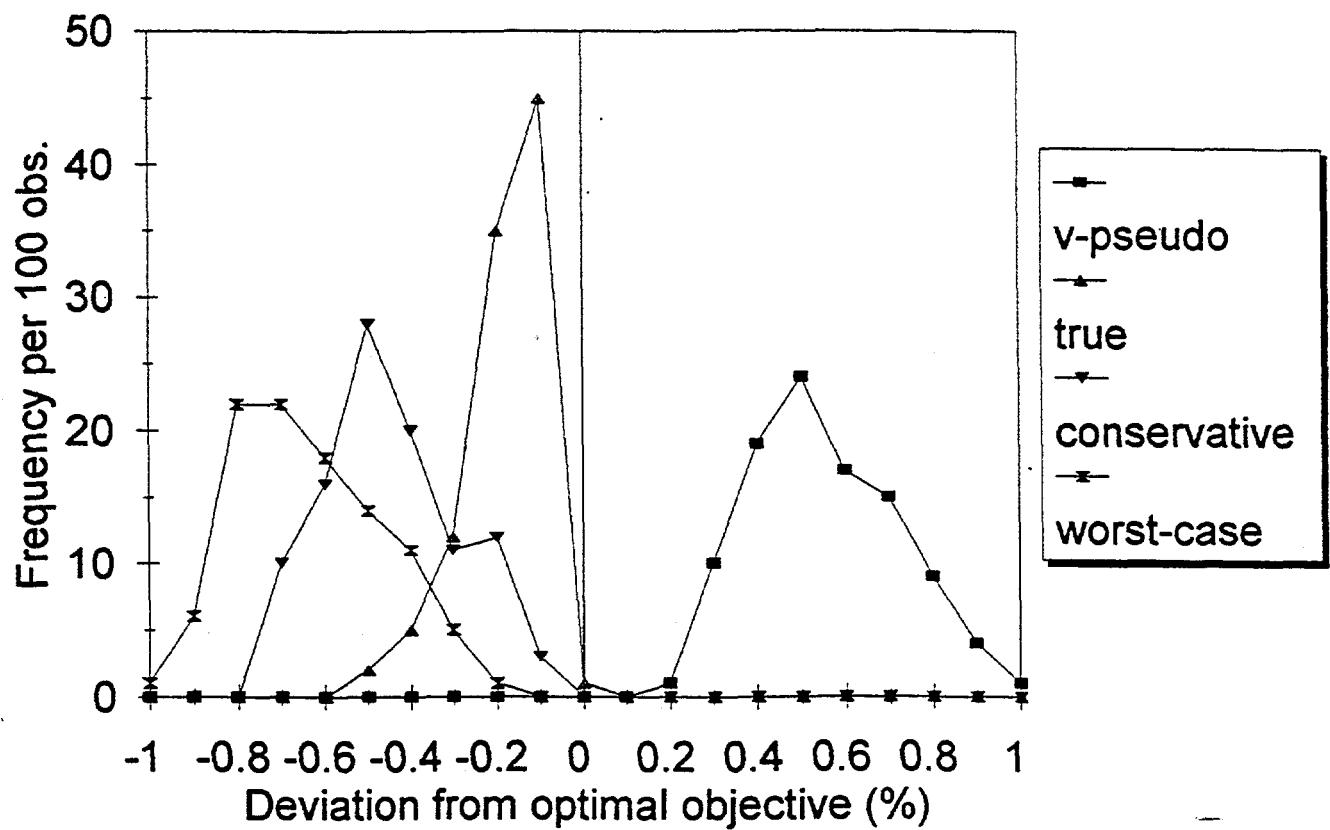
Figure 2: STORM: lower bound distributions