Daryn Ramsden
Stochastic Programming Presentation
12/11/02

Summary of**: Variance Reduction and Objective Function Evaluation in Stochastic Linear Programs** **–** Julia Higle.

Introduction

Stochastic programs serve as a robust decision-making model for planning under uncertainty. However this robustness is counteracted by the sheer magnitude of the computation involved in precise objective function evaluation. As a consequence it is often necessary for estimates to be used. These estimates are often plagued by inappropriately large variances. Higle[1] aims to provide the reader with a survey of techniques that are currently employed for the purpose of variance reduction in the evaluation of objective functions of stochastic linear programs. Four main variance reduction techniques are introduced with simple non-counterintuitive arguments to explain the motivation behind each of these philosophies that further serve to expose the positive and negative aspects of each method. After all four of these methods is introduced, the paper attempts (quite successfully at that) to provide empirical justification for the effectiveness of each method. This empirical justification is achieved by using each of the methods to solve a selection of two stage stochastic problems from the open literature. The results were then tabulated, compared and contrasted.

1. The need for Variance Reduction methods

Consider the rather innocuous looking program:

$$\begin{aligned} &\min cx + E[h(x,\omega)] \\ &s.t. Ax = b \qquad\qquad \text{(SLP)} \\ &x \geq 0, \end{aligned}$$

where

$$\begin{aligned} &h(x,\omega) = \min qy \\ &s.t. Wy \geq r(\omega) - T(\omega)x \qquad \text{(S)} \\ &y \geq 0 \end{aligned}$$

For the rather plausible scenario in which ω is a ten-dimensional vector with each component having three different possible outcomes, a precise evaluation of the objective function requires the solution of 3^{10} or 59,049 linear programs. The inspiration to use samples of $h(x,\omega)$ comes naturally from the fact that the troublemaking term is an expected value. Herein lies the problem: the variance of the estimator derived from any sample is inversely proportional to the size of the sample i.e. larger samples provide more reliable estimators than smaller samples. Larger samples, inconveniently, defeat the purpose of using samples in the first place. The ideal situation would be to have a small sample (all things being relative) that provides a variance in the estimator that would normally be characteristic of a larger sample size.

2. Four approaches to variance reduction.

2.1 Conditional Sampling

Conditional Sampling recognizes that random sampling results in an unreliable quantity of observations coming from regions of the sample space that have low probability masses. This unreliability expectedly leads to avoidable variability in the estimators that are obtained via random sampling. Conditional Sampling entails the partitioning of the sample space and then guaranteeing that the expected number of observations drawn from a given partition is actually the number of observations drawn from said partition. The effectiveness of this method as a means of variance reduction is reliant on the variances of the conditional expectations within the partition. Thus it is prudent to partition the sample space in such a way that all the constituents of any given partition have comparable effects on the objective value. It is pointed out that the objective value is monotonic in the right hand side of the inequality constraint of (S) and thus when the distributions of $r(\omega)$ and $T(\omega)$ are available these partitions are easy to come about.

2.2 Antithetic Sampling

The second sampling methodology considered is that of antithetic sampling. Antithetic sampling seeks to induce a negative correlation in order to take advantage of the fact that the variance of the estimator obtained by finding the mean of two observations (ω^1 and ω^2) can be expressed as a function of the covariance of the two observations as follows:

$$\text{Var[H]} = \tfrac{1}{2}\text{Var}[\,h(x,\omega) + \tfrac{1}{2}\text{Cov}[\,[h(x,\omega 1\,, h(x,w 2\,)]$$

This means that the variance of the estimator obtained by averaging two negatively correlated observations is less than the variance obtained by averaging two independent observations which is just the first term on the right hand side of the equality. Ideally this negative correlation would be obtained by sampling from opposing tails of the distribution of $h(x,\omega)$, however this distribution isn't directly available so the sampling is done indirectly via the distribution of ω, capitalizing on the fact that the objective is monotonic in the right hand side of the inequality constraint. This makes this approach susceptible to the frailties of the transformation h. The degree to which the correlation is retained through the transformation puts a restriction on the extent to which the variance reduction can be effective. The power of antithetic sampling is also diminished as the number of random elements increase.

2.3 Control Variates

The method is of control variates relies on correlated observations, however they are not hindered by the fact that the distribution of h is sampled indirectly. A random variable Z is introduced that has a correlation (positive or negative) with $h(x,\omega)$ and a new function $Y = h(x,\omega) + cZ$ is created with c being a scalar. Observations of Y can be shown to provide lower variance estimators for $\text{E}[h(x,\omega)]$ if $\text{E}[Z]$=0. A further challenge lies in selecting Z such that obtaining observations of it is not an unnecessarily difficult undertaking.

Three potential types of control are considered, the first of which is norm based control in which Z is taken to be the vector of deviations of the right-hand side from its expected value. By definition $\mathrm{E}[Z]=0$. Another form of the method of control variates is to use $Z = \pi\Delta\chi$ where π is the optimal solution of the dual representation of h and $\Delta\chi$ represents the vector of deviations from the right hand side. This is referred to as convexity-based control. A third method employing the use of control variates is an augmentation of norm-based controls that is employed in situations in which generic norm-based controls would result in oversimplification of the relationship between χ and $h(x,\omega)$. This augmented form is referred to as "multiple controls" and requires a few of the random variables to be set aside and be controlled by other measures. Deciding which elements of $\Delta\chi$ are to controlled separately requires investigation into the sensitivity of χ with respect to changes in each of the random elements.

2.4 Importance Sampling

The last means of variance reduction looked into is that of importance sampling in which a biased sample is used and then "debiased". It is shown that

$$H = \tfrac{1}{n}\sum_{t=1}^{n}\tfrac{h(x,\xi^t)p(\xi^t)}{q(\xi^t)}$$

is an unbiased estimator where q defines a probability mass function of a variable ξ. If ξ is defined in an appropriate manner variance reduction can be achieved, however a poor choice can actually lead to increases in observed variances. It is shown that in order that ξ should be defined such that $q \propto \Gamma p(\omega)$ where Γ is a "good" approximation of $h(x,\omega)$. This approach carries the additional burden of requiring the solution (at least implicitly) of one LP for each marginal outcome of ω.

3. Empirical Evaluation.

The following five test problems were used for the purpose of comparing the performance of the different methods:

APL1P: A small-scale power-planning model
CEP1: A small-scale model of capacity expansion planning
PGP2: Another small-scale power-planning model
SSN: A large-scale telecommunications network-planning model
STORM: A large-scale airfreight scheduling model

Some key characteristics of the problems are summarized below:

Problem	# of rows in (S)	# of columns in (S)	# of random variables	# of outcomes	# of marginal outcomes
APL1P	5	9	5	1280	21
CEP1	7	15	3	216	18
PGP2	7	12	3	576	25
SSN	175	706	86	$>5^{86}$	571
STORM	526	1259	118	5^{118}	590

Optimal Solutions to each of these problems were obtained using the Stochastic Decomposition algorithm (Higle and Sen[2]). Since the different methodologies have varying computational requirements the number of LPs solved was fixed for each run, with two different sets of runs undertaken: one in which 500 LPs were solved and one in which 1000 LPs were solved. Each set consisted of 10 runs. The calculation of the estimates of variances of the estimator differed amongst the different methods and were predicated on the manner in which sampling was undertaken.

The three different types of control mechanisms were compared and contrasted, and it was demonstrated that they all provided lower estimation errors than random sampling with the exception of the case in which the convexity-based control was applied to SSN (in this case the errors were identical to that of a random sample). The convexity based control could be said to outperform the others, with the multiple-control method outperforming norm-based control.

Two types of importance sampling were also considered 1) in which the approximation of $h(x, \omega)$ was a lower bound and 2) in which the approximation was an upper bound. Two of the three problems (APL1P and PGP2) favored the lower bound method decisively while CEP1 favored the upper bound method decisively. Neither of the two was able to create a sampling distribution for the other two problems as SSN and STORM require in excess of 500 LPs just to create the sampling distribution. In fact importance sampling was not able to be utilized in the solution of SSN even when 1000 LPs were solved.

In doing the comparison of all four main methods the lower-bounding case of importance sampling was used and the multiple controls case was used as the representative of control mechanisms. These particular choices were made to avoid the observed worst-case scenarios. For the implementation of conditional sampling the sample space was partitioned by splitting the support of the three random variables that the objective function was most sensitive with respect to.

All the methods in general consistently showed some degree of variance reduction with the exception of conditional sampling. Taking into consideration the number of random variables involved in the larger problems it is unreasonable to expect that partitioning according to any three would incite any recognizable reduction in the variability of the estimator. With regard to the lower-variance problems of the set (STORM, PGP2 and APL1P) there was no marked difference amongst the performance of antithetic sampling, importance sampling and control mechanisms. As the variance increased as was the case in the other two problems the method of control variates superceded the alternatives.

4. CONCLUSION

All of the methods employed were seen to have the effect of producing less variable estimates, though some to greater extents than others. The selection of test problems was such to incorporate a variety of characteristics and as such control variate techniques appear to be especially robust and it is anticipated that they could be applied effectively to a large range of problems.

References

1. Julia L. Higle, 1997. Variance Reduction and Objective Function Evaluation in Stochastic Linear Programs, *INFORMS Journal on Computing 236-247.*
2. J.L. Higle and S. Sen, 1991. Stochastic Decomposition: An Algorithm for Two-Stage Linear Programs with Recourse, *Mathematics of Operations Research 16,* 650-669