From Encyclopedia of OR/MS, S. Gass and C. Harris (eds.)

Stochastic Programming: Computational Issues and Challenges

Suvrajeet Sen
SIE Department
University of Arizona, Tucson, AZ 85721

Introduction

Stochastic programming deals with a class of optimization models and algorithms in which some of the data may be subject to significant uncertainty. Such models are appropriate when data evolve over time and decisions need to be made prior to observing the entire data stream. For instance, investment decisions in portfolio planning problems must be implemented before stock performance can be observed. Similarly, utilities must plan power generation before the demand for electricity is realized. Such inherent uncertainty is amplified by technological innovation and market forces. As an example, consider the electric power industry. Deregulation of the electric power market, and the possibility of personal electricity generators (e.g. gas turbines) are some of the causes of uncertainty in the industry. Under these circumstances it pays to develop models in which plans are evaluated against a variety of future scenarios that represent alternative outcomes of data. Such models yield plans that are better able to hedge against losses and catastrophic failures. Because of these properties, stochastic programming models have been developed for a variety of applications, including electric power generation (Murphy et al [1982]), financial planning (Cariño et al [1994]), telecommunications network planning (Sen et al [1994]), and supply chain management (Fisher et al [1997]), to mention a few. The widespread applicability of stochastic programming models has attracted considerable attention from the OR/MS community, resulting in several recent books (Kall and Wallace [1994], Birge and Louveaux [1997], Prékopa [1995]) and survey articles (Birge [1997], Sen and Higle [1999]). Nevertheless, stochastic programming models remain one of the more challenging optimization problems.

While stochastic programming grew out of the need to incorporate uncertainty in linear and other optimization models (Dantzig [1955], Beale [1955], Charnes and Cooper [1959]), it has close connections with other paradigms for decision making under uncertainty. For instance, decision analysis, dynamic programming and stochastic control, all address similar problems, and each is effective in certain domains. Decision analysis is usually restricted to problems in which discrete choices are evaluated in view of sequential observations of discrete random variables. One of the main strengths of the decision analytic approach is that it allows the decision maker to use very general preference functions in comparing alternative courses of action. Thus, both single and multiple objectives are incorporated in the decision analytic framework. Unfortunately, the need to enumerate all choices (decisions) as well as outcomes (of random variables) limits this approach to decision making problems in which only a few strategic alternatives are considered. These limitations are similar to methods based on dynamic programming, which also require finite action (decision) and state spaces. Under Markovian assumptions the dynamic programming approach can also be used to devise optimal (stationary) policies for infinite horizon problems of stochastic control (see also Neuro-Dynamic Programming by Bertsekas and Tsitsiklis [1996]). However, DP-based control remains wedded to Markovian Decision Problems, whereas path dependence is significant in a variety of emerging applications.

Stochastic programming (SP) provides a general framework to model path dependence of the stochastic process within an optimization model. Furthermore, it permits uncountably many states and actions, together with constraints, time-lags etc. One of the important distinctions that should be highlighted is that unlike DP, SP separates the model formulation activity from the solution algorithm. One advantage of this separation is that it is not necessary for SP models to all obey the same mathematical assumptions. This leads to a rich class of models for which a variety of algorithms can be developed. On the downside of the ledger, SP formulations can lead to very large scale problems, and methods based on approximation and decomposition become paramount. In this article, we will provide a road map for these methods, and point to fruitful research directions along the way.

Mathematical Models and Properties

Consider a model in which the design/decision associated with a system is specified via vector x_1. Under uncertainty, the system operates in an environment in which there are uncontrollable parameters which are modeled using random variables. Consequently, the performance of such a system can also be viewed as a random variable. Accordingly, SP models provide a framework in which designs (x_1) can be chosen to optimize some measure of the performance (random variable). It is therefore natural to consider measures such as the worst case performance, expectation and other moments of performance, or even the probability of attaining a predetermined performance goal. Furthermore, measures of performance must reflect the decision maker's attitudes towards risk. For example in the financial literature, it is common to model risk aversion through the use of a utility function.

The following mathematical model represents a general SP formulation in which the design/decision variable x_1 is restricted to the set X_1, and $\tilde{\omega}_1$ denotes a multi-dimensional random variable.

$$\operatorname*{Min}_{x_1 \in X_1} \quad f_1(x_1) + E[h_2(x_1, \tilde{\omega}_1)] \tag{1a}$$

$$\text{s.t.} \quad P[g_1(x_1, \tilde{\omega}_1) \geq 0] \; \geq p_1 \tag{1b}$$

Here E denotes the expectation with respect to $\tilde{\omega}_1$ and P denote the probability distribution associated with $\tilde{\omega}_1$. The function g_1 is often modeled by a linear function and h_2 is the value function of another optimization problem as follows:

$$h_2(x_1, \omega_1) = \operatorname*{Min}_{x_2 \in X_2(x_1, \omega_1)} \quad f_2(x_2; x_1, \omega_1).$$

In the SP literature, the function h_2 is used to reflect costs associated with adapting to information revealed through an outcome ω_1. In financial applications, this function may reflect the utility associated with costs of rebalancing the portfolio. Because the function $E[h_2]$ is associated with a recourse action, it is referred to as the recourse function. Constraint (1b) is called a probabilistic (or chance) constraint. Such a constraint is used to model system reliability. We should mention that formulation (1) is somewhat more general than one usually finds in the SP literature. Historically, the probabilistic constraint (1b) is treated separately from models using the recourse functions (1a). However, including both types of functions within a model allows us to view the SP problems in a more cohesive manner.

While model (1) appears somewhat static, it is not difficult to glean a dynamic element in the formulation: note that the function h_2 is realized only after the design x_1 is in place. This sequential nature is an essential element of decision making under uncertainty. Indeed, if we define h_2 recursively, problem (1) may be looked upon as the first stage problem of a more extensive multistage formulation. To present the multistage generalization of (1), consider an N stage problem. Let the boundary conditions be given by $h_{N+1} = 0$ and let $\tilde{\omega}_0$ denote a degenerate random variable reflecting the deterministic information available prior to decisions in stage 1. For $t = 1, \ldots, N$, let ξ^t denotes the history prior to stage t (i.e $\xi^t = (\omega_0, \ldots, \omega_{t-1})$). Note that the decision variables in stage t depend on the history of the data process. Hence these variables are functions of random variables, and will be denoted $x_t(\xi^t)$. The entire history of decisions until stage t will then be represented as a superscripted vector $x^t(\xi^t) = (x_1(\xi^1), x_2(\xi^2), \ldots, x_t(\xi^t))$, or simply x^t. For $t = 2, \ldots, N$, we can now define the value functions

$$h_t(x^{t-1}, \xi^t) = \underset{x_t \in X_t(x^{t-1}, \xi^t)}{\text{Min}} \quad f_t(x_t; x^{t-1}, \xi^t) + E[h_{t+1}(x^t, \tilde{\xi}^{t+1} | \xi^t)]$$
$$\text{s.t.} \qquad P[g_t(x^t, \tilde{\xi}^{t+1} | \xi^t) \geq 0] \ \geq p_t,$$

where E denotes the conditional expectation and P denotes the conditional probability associated with the appropriate random variables. Using these functions in (1) yields a multistage SP formulation.

While we have used a DP-type recursion to state the SP problem, it is important to note that all random variables are path dependent, and furthermore, unlike DP, the statement of the problem does not constitute the algorithm. In fact, alternative statements of the multistage problem are also possible. Consider a formulation in which we allow the decisions to depend on the entire realization ξ^N. Let $x^N(\tilde{\xi}^N)$ denote a sequence of random vectors $(x_1(\tilde{\xi}^N), x_2(\tilde{\xi}^N), \ldots, x_N(\tilde{\xi}^N))$. It is important to note that such a policy cannot be implemented since decisions in stage t require the knowledge of the entire realization! Hence, the plans (denoted $x^N(\tilde{\xi}^N)$) cannot be feasible, unless, the decisions are such that x_t depends only on data available until stage $t-1$. As shown below we can incorporate such information constraints explicitly.

Let $\omega^t \equiv (\omega_t, \ldots, \omega_N)$. Since any outcome $\xi^N = (\xi^t, \omega^t)$ for any t, the decisions in stage t can be represented as a random vector denoted $x_t(\tilde{\omega}^t | \xi^t)$. Then the information constraints (also called the nonanticipativity constraints) may be stated as

$$x_t(\tilde{\omega}^t | \xi^t) - E[x_t(\tilde{\omega}^t | \xi^t)] = 0 \quad \text{almost surely.}$$

Since a non-separable objective function can be written as $E[f(x^N(\tilde{\xi}^N), \tilde{\xi}^N)]$, the inclusion of information (nonanticipativity) constraints provides a legitimate multistage model which does not appeal to either separability or recursion.

The formulations presented above impose very few restrictions. Perhaps the most important restriction imposed in a SP formulation arises from the assumption that randomness is exogenous and cannot be affected by decisions. In certain design problems, such an assumption may not be valid, and in these cases, the models outlined above are inadequate. Nevertheless, there is a large class of applications where randomness is exogenous (e.g. weather, loads, prices of financial instruments, market demands etc.), and SP models provide a sound approach.

The main challenge in designing algorithms for stochastic programming problems arises from the need to calculate conditional expectation and/or probability associated with multi-dimensional random variables. For all but the smallest of problems, we resort to approximations. The study of stochastic programming algorithms has therefore led to alternative ways of approximating problems, some of which obey certain asymptotic properties. This reliance on approximations has prompted researchers to study conditions for the convergence of approximations, and/or the convergence of solutions of approximate problems (to a solution of the original). Of course, conditions ensuring the former imply the latter, but the converse does not hold. Issues related to convergence of approximations can be addressed through the theory of epi-convergence (King and Wets [1991], Rockafellar and Wets [1998]) whereas issues pertaining to convergence of solutions of approximations (to a solution of the original) can be addressed through the notion of epigraphical nesting (Higle and Sen [1992], [1995]).

The computational challenges associated with SP problems vary a great deal with the class of problems being addressed. As with any large scale optimization problem, exploiting properties and the structure of problems provides the key to effective algorithms. We discuss properties associated with some important classes of SP problems, and then proceed to discuss the computational issues.

Some Properties of Stochastic Linear Programs with Recourse

For this class of problems, all functions and constraints are defined by linear/affine functions, and the probabilistic constraints are absent. This remains one of the more widely studied models, and most of the applications reported in the literature belong to this category (including the applications mentioned earlier). Problems of this type can be shown to be convex optimization problems, and the full power of convex analysis can be brought to bear on such problems. Notwithstanding such mathematical attractiveness, SLP problems lack one of the more desirable numerical properties, namely, smoothness. Only under very special circumstances (absolute continuity of random variables, Kall [1976]), can one expect (1a) to be differentiable.

Some Properties of Stochastic Mixed Integer Linear Programs

For this class of problems, we continue with the absence of probabilistic constraints. In a stochastic mixed integer linear program (S-MILP), if only the first stage decisions include integer restrictions, then the remaining problem inherits the properties of a SLP. This class of problems (with first stage integer variables) is similar to the problems originally envisioned by Benders in his seminal paper (Benders [1962]). In general though (i.e when integer variables appear in future stages) the S-MILP is much more challenging. For such problems, convexity of the objective function is far too much structure to expect. Indeed, the objective function (1a) can be discontinuous. However, by assuming that any setting of decision variables yields a finite objective value (i.e. complete recourse), and assuming a weak covariance condition (Schultz [1993]) the objective function can be shown to be lower semicontinuous.

Some Properties of Probabilistically Constrained Problems

These models are widely used to reflect grade of service constraints (e.g. Medova [1998]). The early work for this class of problems was restricted to normally distributed random variables. Prékopa [1971] showed that a much larger class of random variables yield the convexity property; he showed that if the function g (see (1b)) is linear/affine in x and randomness only appears additively, and the random variable has a log-concave probability density function, then the resulting feasible region is convex. However, for discrete random variables this is no longer true, and in this case, the set of feasible solutions can be represented as a disjunctive set (Sen [1993]).

Computational Issues and Challenges

The main computational challenges can be attributed to presence of multi-dimensional integration (to calculate either expectation or probability) within an optimization algorithm. Even in cases where the random variables are discrete, the total number of outcomes of a multi-dimensional random vector can be so large that calculations associated with the summations may be far too demanding. Hence even in the case of discrete random variables, one may have to resort to approximations. Discretizations and/or aggregations in multistage problems result in alternative data scenarios or sample paths. These scenarios may be organized in the form of a scenario tree which is a structure representing the evolution of information over the stages. In such a tree, two scenarios that share a common history until stage t are indistinguishable until that stage, and thereafter they are represented by distinct paths. Thus every distinct scenario represents a path from the root node to a leaf node of the scenario tree. In the absence of appropriate approximations, these trees can become extremely large, and the model difficult to manage and solve.

There are essentially two major approaches to generating approximations. One is based on aggregating data points, and another based on selecting data points. The former class of algorithms lead to successive approximation methods in which finer discretizations of the sample space are created based on the solution of an aggregated stochastic program. Methods based on data aggregation and successive refinements have been forwarded by several authors, and a survey for two stage problems can be found in Frauendorfer [1992]. More recently, Edirisinghe and Ziemba [1996] have reported solving two stage problems with approximately 20 random variables. For multistage problems, data-aggregation methods have been proposed in Frauendorfer [1994], but computational results are very limited.

The idea of selecting data points to create approximations arises mainly in the context of sample-based algorithms. If one uses a fixed sample, then it is necessary to perform a statistical analysis of the output, as suggested by the work of Romisch and Schultz [1991] and Shapiro [1991]. To obtain asymptotic results, Shapiro and Homem-de-Mello [1998] (see also sample path optimization, Robinson [1996]) suggest solving a sequence of sampled approximations, with increasing sample sizes. As the approximating problem becomes larger, each iteration may become substantially demanding. In order to speed up computations associated with such a method, it may be advantageous to update approximations generated in earlier iterations. One such method for two stage problems is the stochastic decomposition (SD) algorithm (Higle and Sen [1991]) which incorporates sampling within a decomposition method. This combination allows the SD method to update approximations from one iteration to the next, thus allowing matrix updates and warm starts during re-optimization. Because sampling and decomposition are intimately interwoven in the SD algorithm, it allows the possibility of using empirical data directly within the algorithmic process. A detailed exposition of this work appears in Higle and Sen [1996], and recent results are summarized in Higle and Sen [1999]. As with methods based on data-aggregation, computational results with sample-based algorithms for multistage problems are extremely limited.

Most of the approximation schemes mentioned above are paired with some deterministic algorithm. This genre of methods traces back to the L-shaped method of Van Slyke and Wets [1969] which builds on arguments similar to Benders' decomposition (Benders [1962]) for two stage problems. The method has been extended in several ways, including generalizations to multistage problems (Birge [1985]). When the number of scenarios is small, these methods can be applied directly. Otherwise, they should be used in conjunction with approximation-based methods such as those discussed above.

Another class of deterministic decomposition algorithms is based on relaxing the information (nonanticipativity) constraints. This approach is particularly promising for parallelizing algorithms for multistage problems. Two such methods are the progressive hedging method of Rockafellar and Wets [1991] and diagonal quadratic approximation method of Mulvey and Ruszczyński [1995]. One of the biggest advantages of these methods is that they retain the structure of a deterministic counterpart (e.g. network structure) and are easily parallelizable. Furthermore, each processor can be allocated a collection of scenarios which can be coordinated with minimal oversight. Nielsen and Zenios [1993] report significant speed-ups of their parallel implementation over a serial code. It would be interesting to design sample-based algorithms of this type, and although some preliminary steps have been taken, it is too early to tell how such methods will perform.

One of the more demanding problems in stochastic programming involves the solution of stochastic mixed integer linear programs (S-MILP). In cases where the first stage has binary variables, Laporte and Louveaux [1993] have proposed an extension of the L-shaped method for two stage S-MILP problems. Unfortunately, it requires that the second stage problem (possibly a mixed-integer linear program (MILP)) be solved to optimality. Given the computational difficulties associated with MILPs, this is cumbersome. One possible way to alleviate this difficulty may involve an extension that incorporates the stochastic branch and bound algorithm proposed by Norkin, Ermoliev and Ruszczyński [1998]. Since the latter allows the use of sampled bounds, it may provide a more computationally feasible approach to practical S-MILP problems.

Finally, one area that has not attracted as much attention as it should is the development of software systems which integrate stochastic modeling with stochastic programming algorithms. Such a system would provide software tools to build models, validate them, and experiment with alternative algorithms. With few exceptions (e.g. Kall and Mayer [1996] and Gassmann and Ireland [1996]), there has been relatively little activity in this important area. It is unlikely that stochastic programming will attain its potential without the development of systems which allow easy interactions between stochastic models and stochastic programming algorithms. The development of a high level language or system that allows manipulation and representation of models and data, together with the ability to experiment with alternative solvers, is long overdue.

Before closing this article, we should point the reader to an extensive list of papers maintained by Maarten van der Vlerk at the following web site.

`http://mally.eco.rug.nl/biblio/SPlist.html`

References

Birge, J.R. [1985]. "Decomposition and partitioning methods for multi-stage stochastic linear programs," *Operations Research*, 33, pp. 989-1007.

Birge, J.R. [1997]. "Stochastic Programming Computation and Applications," *INFORMS J. on Computing*, 9, pp. 111-133.

Birge, J.R. and F.V. Louveaux [1997]. *Introduction to Stochastic Programming*, Springer, New York.

Beale, E.M.L [1955]. "On minimizing a convex function subject to linear inequalities," *J. of the Royal Statistical Society, Series B*, 17, pp. 173-184.

Benders, J.F. [1962], Partitioning procedures for solving mixed variables programming problems, *Numerische Mathematik*, **4**, pp. 238-252.

Bertsekas, D. and J. Tsitsiklis [1996] *Neuro-Dynamic Programming*, Athena Scientific, Belmont, MA.

Cariño, D.R., T. Kent, D.H. Meyers, C. Stacy, M. Sylvanus, A.L. Turner, K. Watanabe, and W.T. Ziemba [1994]. "The Russell-Yasuda Kasai Model: An asset/liability model for a Japanese insurance company using multistage stochastic programming," *Interfaces*, 24, pp.29-49.

Charnes, A. and W.W. Cooper [1959]. "Chance-constrained programming," *Management Science*, 5, pp. 73-79.

Dantzig, G.B. [1955]. "Linear programming under uncertainty," *Management Science*, 1, pp. 197-206.

Edirisinghe, N.C.P. and W.T. Ziemba [1996]. "Implementing bounds-based approximations in convex-concave two stage programming," *Mathematical Programming*, 19, pp. 314-340.

Fisher, M., J. Hammond, W. Obermeyer, and A. Raman [1997]. "Configuring a supply chain to reduce the cost of demand uncertainty," *Production and Operations Management*, 6, pp.211-225.

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

Frauendorfer, K. [1994]. "Multistage stochastic programming: error analysis for the convex case," *Zeitschrift fur Operations Research*, 39, pp. 93-122.

Gassmann, H.I. and A.M. Ireland [1996]. "On the formulation of stochastic linear programs using algebraic modelling languages," *Annals of Operations Research*, 64, pp. 83-112.

Higle, J.L. and S. Sen [1991]. "Stochastic Decomposition: An algorithm for two-stage linear programs with recourse," *Mathematics of Operations Research*, 16, pp. 650-669.

Higle, J.L. and S. Sen [1992]. "On the convergence of algorithms with implications for stochastic and nondifferentiable optimization," *Math. of Operations Research*, 17, pp. 112-131.

Higle, J.L. and S. Sen [1995]. "Epigraphical Nesting: a unifying theory for the convergence of algorithms," *Journal of Optimization Theory and Applications*, 84, pp. 339-360.

Higle, J.L. and S. Sen [1996]. *Stochastic Decomposition: A Statistical Method for Large Scale Stochastic Linear Programming*, Kluwer Academic Publishers, Dordrecht.

Higle, J.L. and S. Sen [1999]. "Statistical approximations for stochastic linear programming problems," *Annals of Operations Research*, 85, pp. 173-192.

Kall, P. [1976]. *Stochastic Linear Programming*, Springer-Verlag, Berlin.

Kall, P. and S.W. Wallace [1994]. *Stochastic Programming*, John Wiley & Sons, Chichester, England.

Kall, P. and J. Mayer [1996]. "SLP-IOR: an interactive model management system for stochastic linear programs," *Mathematical Programming, Series B*, 75, 221-240.

King, A.J. and R. J-B. Wets [1991]. "Epi-consistency of convex stochastic programs," *Stochastics*, 34, pp.83-92.

Laporte, G. and F.V. Louveaux [1993]. "The integer L-shaped method for stochastic integer programs with complete recourse," *Operations Research Letters*, 13, pp.133-142.

Medova, E. [1998]. "Chance constrained stochastic programming for integrated services network management," *Annals of Operations Research*, 81, pp. 213-229.

Mulvey, J.M. and A. Ruszczyński [1995]. "A new scenario decomposition method for large scale stochastic optimization," *Operations Research*, 43, pp. 477-490.

Murphy, F.H., S. Sen and A.L. Soyster [1982]. "Electric utility capacity expansion planning with uncertain load forecasts," *AIIE Transaction*, 14, pp. 52-59.

Nielsen, S.S. and S.A. Zenios [1993]. "A massively parallel algorithm for nonlinear stochastic network problems," *Operations Research*, 41, 319-337.

Norkin, V.I., Y.M. Ermoliev and A. Ruszczyński [1998]. "On optimal allocation of indivisibles under uncertainty," *Operations Research*, 46, pp. 381-395.

Prékopa, A. [1971]. " Logarithmic concave measures with application to stochastic programming," *Acta Scientiarium Mathematiarum (Szeged)*, 32, pp. 301-316.

Prékopa, A. [1995]. *Stochastic Programming*, Kluwer Academic Publishers, Dordrecht.

Robinson, S.M. [1996]. "Analysis of sample path optimization," *Math. of Operations Research*, 21, 513-528.

Rockafellar, R.T. and R. J-B. Wets [1991]. "Scenarios and policy aggregation in optimization under uncertainty," *Math. of Operations Research*, 16, pp. 119-147.

Rockafellar, R.T and R.J-B. Wets [1998]. *Variational Analysis*, Springer-Verlag, Berlin.

Romisch, W. and R. Schultz [1991]. "Distribution sensitivity in stochastic programming," *Mathematical Programming*, 50, 197-226.

Schultz, R. [1993]. "Continuity properties of expectation functions in stochastic integer programming," *Math. of Operations Research*, 18, pp. 578-589.

Sen, S. [1992]. "Relaxations for probabilistically constrained programs with discrete random variables," *Operations Research Letters*, 11, pp. 81-86.

Sen, S. R.D. Doverspike and S. Cosares [1994]. "Network Planning with Random Demand," *Telecommunication Systems*, 3, pp. 11-30.

Sen, S. and J.L. Higle [1999]. "An Introductory Tutorial on Stochastic Linear Programming Models," *Interfaces*, 29, pp. 33-61.

Shapiro, A. [1991]. "Asymptotic analysis of stochastic programs," *Annals of Operations Research*, 30, pp. 169-186.

Shapiro, A. and Homem-de-Mello, T. [1998]. "A Simulation-based approach to stochastic programming with recourse," *Mathematical Programming*, 81, pp. 301-325.

Van Slyke, R. and R. J-B. Wets [1969], L-Shaped linear programs with application to optimal control and stochastic programming, *SIAM J. on Appl. Math.*, **17**, pp. 638-663.