Chapter 9
Monte Carlo Methods

Each function value in a stochastic program can involve a multidimensional integral in extremely high dimensions. Because Monte Carlo simulation appears to offer the best possibilities for higher dimensions (see, e.g., Deák [1988] and Asmussen and Glynn [2007]), it seems to be the natural choice for use in stochastic programs. In this chapter, we describe some of the basic approaches built on sampling methods. The key feature is the use of statistical estimates to obtain confidence intervals on results. Some of the material uses probability measure theory which is necessary to develop the analytical results.

To build on our earlier emphasis on decomposition algorithms, Section 9.1 begins this discussion with a description of the basic sampling approximation, the sample-average approximation, and then approaches uses of this system with the L-shaped method. We first consider possibilities for estimating the cuts in this method using a large number of samples for each cut. Section 9.2 then considers the stochastic decomposition method (Higle and Sen [1991b]) that forms many cuts with few additional samples on each iteration. Section 9.3 considers methods based on the stochastic quasi-gradient, which can be viewed as a generalization of the steepest descent method. These approaches have a wide variety of applications that extend beyond stochastic programming. In Section 9.4, we consider extensions of Monte Carlo methods to include analytical evaluations exploiting problem structure in probabilistic constraint estimation and empirical sample information for methods that may use updated information in dynamic problems. Section 9.5 describes basic theoretical results for the statistical analysis of stochastic programs and, in particular, for the sample-average approximation. We describe asymptotic properties and large-deviation bounds for optimal values and solutions to those problems.
9.1 Sample Average Approximation and Importance Sampling in the \(L\)-Shaped Method

The most direct sampling approach to the two-stage stochastic program is to replace the recourse function, \(Q(x)\), by a Monte Carlo estimate,

\[
Q^\nu(x) = \sum_{k=1}^{\nu} \frac{Q(x, \xi^k)}{\nu},
\]

(1.1)

where \(\xi^1, \ldots, \xi^\nu\) are random samples of the random vector \(\xi\). This then yields the sample average approximation (SAA) problem for the general two-stage problem as:

\[
\min_{x \in X} f^1(x) + \sum_{k=1}^{\nu} \frac{Q(x, \xi^k)}{\nu},
\]

(1.2)

where \(X\) represents the feasibility set as, for example, in the nonlinear program in (3.4.1). For a stochastic linear program, we can then write (1.2) as:

\[
\min c^T x + \frac{1}{\nu} \sum_{k=1}^{\nu} q_k^T y_k
\]

s. t. \(Ax = b, \quad T_k x + W y_k = h_k, \quad x \geq 0, y_k \geq 0.\)

(1.3)

As we show in Section 9.5, by increasing the sample size \(\nu\), solutions to (1.3) converge to an optimal solution of the two-stage stochastic program (3.1.2). A disadvantage of solving (1.3) completely for each \(\nu\) using any algorithm is that some effort might be wasted on optimizing when the approximation is not accurate. An approach to avoid these problems is to use sampling within another algorithm without complete optimization. In this section, we describe this process for the \(L\)-shaped method, which often works well for discrete distributions. To ensure that the process makes efficient use of the sample information, we first describe a version using importance sampling to reduce variance in deriving each cut based on a large sample (see Dantzig and Glynn [1990]). In the following section, we consider an approach that uses a single sample stream to derive many cuts that eventually drop away as iteration numbers increase (Higle and Sen [1991b]).

The general approach is to sample \(Q\) to construct cuts in the \(L\)-shaped method to obtain an approximate solution to (3.1.2). Using a crude Monte Carlo sample of \(\xi\), however, may result in high variance for the sample values \(Q(x, \xi^k)\), slowing convergence or leading to biased results. Instead, to reduce the variance of the sample values, we use the importance sampling (see, e.g., Rubinstein [1981] and Deák [1990]) variance-reduction technique to concentrate samples where they provide the most information.