

## UNATTAINED CONVERGENCE FOR SAMPLING METHODS IN LARGE-SCALE OPTIMIZATION MODELS AND A REMEDY WITH BATCH MEANS

John R. Birge

Graduate School of Business  
University of Chicago  
Chicago, Illinois 60637, USA

### ABSTRACT

Optimization software enables the solution of problems with millions of variables and associated parameters. These parameters are, however, often uncertain and represented with an analytical description of the parameter's distribution or with some form of sample. With large numbers of such parameters, optimization of the resulting model is often driven by mis-specifications or extreme sample characteristics, resulting in solutions that are far from a true optimum. This paper describes this phenomenon and the difficulties in achieving asymptotic convergence results that are often applied without determining whether the asymptotic regime has been attained. Through the use of a motivating simple example, the paper provides one potential remedy with the use of batch means.

### 1 INTRODUCTION

Advances in software, hardware, and algorithms for optimization have led to orders-of-magnitude decreases in solution times and similar increases in the sizes of problems amenable to solution (see, e.g., Bixby 2002). The ability to solve problems of virtually any scale appears to offer significant promise for operations research methodology, but this development also comes with a price. As problem size increases, so does the opportunity for errors in the parameters used to describe the model. Unfortunately, optimization tends to focus on these errors and can lead to solutions that are significantly inferior to a true optimal solution and, in some cases, even inferior to naïve rule-of-thumb solutions that require no optimization.

Recognizing uncertainty in optimization model parameters naturally leads to stochastic programming formulations, but, as shown below, these formulations also may have difficulties despite asymptotic convergence results that suggest otherwise. In certain examples, such as portfolio optimization, analytical models might even be possible to capture uncertainties in the model, but these models are also prone

to estimation errors that are inevitable in large-scale models. One remedy for this dilemma is to use *robust optimization* (see, e.g., Ben-Tal and Nemirovski 2002), which optimizes against the worst case of a wide range of parameter choices. This approach, however, still requires some assumption on the parameter ranges and loses the form of expected utility that is usually assumed for rational decision making.

In this paper, we consider the expected utility framework directly and assume that the set of parameters can be described by some distribution that may be used for Monte Carlo simulation. We state some of the known asymptotic convergence properties for optimization in this framework. We use a small example to demonstrate the problems that may occur in large-scale models in achieving the regime in which these asymptotic results apply. For this example, we show how the use of batch means can lead to improved convergence results. We conclude with implications for more general models.

### 2 Asymptotic Results and Difficulties in Large-Scale Models

The canonical problem that we consider is to find  $x \in X \subset \mathfrak{R}^n$  to minimize:

$$E[g(x, \xi(\omega))], \quad (1)$$

where  $E$  represents mathematical expectation and  $\omega$  is associated with a probability space  $(\Omega, \Sigma, P)$ . The function  $g$  can be interpreted as the (negative of) the utility resulting from action  $x$  and outcome  $\xi$ . The set of parameters are given by the random vector,  $\xi : \Omega \mapsto \mathfrak{R}^m$ . Our interest is in cases when  $m$  is very large.

In some cases, (1) can be solved directly using a known (or supposed) distribution on  $\xi$ . A typical example is the mean-variance (Markowitz 1959) portfolio optimization to minimize the portfolio variance subject to constraints on the expected return. This case can be expressed using  $X = \{x | e^T x = 1, \bar{r}^T x = r_0\}$ , where  $r_0$  is the target return and  $\bar{r}$  is the vector of expected returns;  $g(x, \xi(\omega)) = (r(\omega)^T x - \bar{r}^T x)^2$ .

The analytical representation is then that  $E[g(x, \xi(\omega))] = x^T V x$ , where  $V$  is the variance-covariance matrix of the returns  $r(\omega)$ .

While the mean-variance problem can be solved analytically if both the mean returns  $\bar{r}$  and variance-covariance matrix  $V$  are known, in practice, these distribution parameters are not known with certainty. In this case, if the number of assets is  $n$ ,  $V$  then includes  $(n(n+1))/2$  distinct elements. Errors in any of these estimates lead to significant deviations in the optimum from the (unknown) true value. In a recent study, [DeMiguel et al. \(2005\)](#) found that the simplistic allocation of  $x = e/n$ , equal allocation to each asset, out-performed every considered optimization procedure based on empirical and simulated data. Even in these small portfolio problems, limited data sets lead to significantly sub-optimal solutions.

In the mean-variance case, the difficulty concerns misestimates of distribution parameters. Deviations due to inexact integration or sampling are, however, avoided. Even when distributions are known exactly, but finding expectations is difficult, the results can be equally troublesome. These difficulties occur despite encouraging asymptotic results.

In the Monte Carlo simulation, we assume that the distribution of  $\xi$  is known and consider a sample  $\{\xi^i, i = 1, \dots, v$  of independent observations of  $\xi$  that lead to the following sample problem:

$$\min_{x \in X} \left( \frac{1}{v} \sum_{i=1}^v g(x, \xi^i) \right). \quad (2)$$

Following the description in [Birge and Louveaux \(1997\)](#), let  $x^v$  be the random vector of solutions to (2) with independent random samples. [King and Rockafellar \(1993\)](#) provided the following result.

**Theorem 1.** *Suppose that  $g(\cdot, \xi)$  is convex and twice continuously differentiable,  $X$  is a convex polyhedron,  $\nabla g : \Xi \times \mathfrak{R}^n \mapsto \mathfrak{R}^n$ :*

1. *is measurable for all  $x \in X$ ;*
2. *satisfies the Lipschitz condition that there exists some  $a : \Xi \mapsto \mathfrak{R}$ ,  $\int_{\Xi} |a(\xi)|^2 P(d\xi) < \infty$ ,  $|\nabla g(x_1, \xi) - \nabla g(x_2, \xi)| \leq a(\xi)|x_1 - x_2|$ , for all  $x_1, x_2 \in X$ ;*
3. *satisfies that there exists  $x \in X$  such that  $\int_{\Xi} |g(x, \xi)|^2 P(d\xi) < \infty$ ; and, for  $G^* = \int \nabla^2 g(x^*, \xi) P(d\xi)$ ,*
4.  $(x_1 - x_2)^T G^* (x_1 - x_2) > 0, \forall x_1 \neq x_2, x_1, x_2 \in X$ .

Then the solution  $x^v$  to (2) satisfies:

$$\sqrt{v}(x^v - x^*) \mapsto u, \quad (3)$$

where  $u$  is the solution to:

$$\begin{aligned} \min \quad & \frac{1}{2} u^T G^* u + c^T u \\ \text{s. t.} \quad & A_i u_i \leq 0, i \in I(x^*), u^T \nabla \bar{g}^* = 0, \end{aligned} \quad (4)$$

$X = \{x | Ax \leq b\}$ ,  $(x^*, \pi^*)$  solve  $\nabla \int_{\Xi} g(x^*, \xi) P(d\xi) + (\pi^*)^T A = 0, \pi^* \geq 0, Ax^* \leq b, I(x^*) = \{i | A_i x^* = b_i\}$ ,  $\nabla \bar{g}^* = \int \nabla g(x^*, \xi) P(d\xi)$ , and  $c$  is distributed normally  $N(0, \Sigma^*)$  with  $\Sigma^* = \int (\nabla g(x^*, \xi) - g^*) (\nabla g(x^*, \xi) - g^*)^T P(d\xi)$ .

The result of this theorem implies that asymptotically, the sample average problem (2) approaches a true optimal solution to (1) quickly. As shown in [Birge and Louveaux \(1997\)](#) and [Shapiro and Homem-de-Mello \(2000\)](#), the convergence is actually often directly to a point, leading to exact convergence in a finite number of samples. These results do not, however, give an iteration number  $v$  at which this asymptotic regime begins to apply. In fact, this regime may only take hold for very large  $v$ .

As an example, consider the following.

$$\min_{\|x\| \leq 1} E[\xi^T x + \varepsilon \|x\|], \quad (5)$$

where  $\|\cdot\|$  is the 1-norm (i.e., making this formulation equivalent to a linear program),  $E[\xi] = 0$ , and  $\varepsilon > 0$  is a constant. The optimal solution to (5) is  $x^* = 0$ , but achieving  $x^v \rightarrow x^*$  can be difficult.

For a sample average version of (5) with  $\xi^v = \sum_{i=1}^v \xi^i / v$ , it can be shown that  $\|x^v - x^*\| = 1$  if there exists  $|\xi^v(j)| > \varepsilon$  for any  $j$ . When  $n$  is large, the chance of  $\xi^v(j) \leq \varepsilon$  for all  $j$  becomes small. When each  $\xi_j$  is  $N(0, 1)$ , in fact, the result is

$$P\{\|x^v - x^*\| < 1\} = (1 - 2\Phi(-\varepsilon v^{0.5}))^n, \quad (6)$$

which means that the asymptotic regime does not apply until  $v$  is very large.

To provide a remedy to this problem, we propose dividing the  $v$  samples into  $k$  batches of  $v/k$  samples each. We let  $\xi^{v,i}$  be the mean of batch  $i = 1, \dots, k$ , and then solve (2) to obtain a solution  $x^{v,i}$ . Now, we let  $x^{v,b} = (1/k) \sum_{i=1}^k x^{v,i}$ . In this case, we can show that

$$P\{\|x^{v,b} - x^*\| < (1/k)\} \approx \quad (7)$$

$$1 - (1 - p(v, k))^k - kp(v, k)^{k-1} (1 - p(v, k)) / n,$$

where  $p(v, k) = P\{\|x^{v,i} - x^*\| < 1\}$ . The result of (7) is that  $x^{v,b}$  is much more likely to approach  $x^*$  than  $x^v$  for large  $n$  and moderate values of  $v$ .

### 3 Conclusions

Large-scale stochastic optimization problems present significant problems for effective solution. Monte Carlo sample estimates have good asymptotic results but the asymptotic regime may only be attained for excessively large sample sizes for large-scale problems. Batch means may achieve improved asymptotic results in these cases. Further research is needed to determine the generality of these results for broader classes of problems.

### REFERENCES

- A. Ben-Tal, A. Nemirovski, Robust optimization - methodology and applications, *Mathematical Programming Series B* 92 (2002), 453–480.
- J.R. Birge and F.V. Louveaux, *Introduction to Stochastic Programming*, Springer, New York, 1997.
- R.E. Bixby, Solving real-world linear programs: a decade and more of progress, *Operations Research* 50 (2002), 3–15.
- V. DeMiguel, L. Garlappi, and R. Uppal, How inefficient is the  $1/N$  asset-allocation strategy?, CEPR Discussion Paper No. 5142, London Business School working paper, 2005.
- A. King and R.T. Rockafellar, Asymptotic theory for solutions in generalized M-estimation and stochastic programming, *Mathematics of Operations Research* 18 (1993), 148–162.
- H.M. Markowitz, *Portfolio Selection; Efficient Diversification of Investments* (John Wiley, Inc., New York, NY, 1959).
- A. Shapiro and T. Homem-de-Mello, On the rate of convergence of Monte Carlo approximations of stochastic programs, *SIAM Journal on Optimization* 11 (2000), 70–86.

### AUTHOR BIOGRAPHY

**JOHN R. BIRGE** is the Jerry W. and Carol Lee Levin Professor of Operations Management at the University of Chicago Graduate School of Business. He was previously dean of engineering at Northwestern University and chair and professor in the Department of Industrial and Operations Engineering at the University of Michigan. He is an editor with multiple journals and has interests in stochastic optimization and its applications in all areas. His web page is [www.chicagogsb.edu/fac/john.birge](http://www.chicagogsb.edu/fac/john.birge).