*Proceedings of the 1994 Winter Simulation Conference*
*ed. J. D. Tew, S. Manivannan, D. A. Sadowski, and A. F. Seila*

# IDENTIFYING POLICIES THAT ARE MOST LIKELY TO PRODUCE A DESIRABLE OUTCOME

Kevin J. Healy

School of Industrial Engineering
Purdue University
1287 Grissom Hall
West Lafayette, IN 47907-1287

## ABSTRACT

For some types of stochastic optimization problems it is possible to directly estimate the location of an optimum on a response surface using an experimental technique based on the notion of *simulated hindsight*. The idea is to *retrospectively* solve a related deterministic optimization problem with respect to realizations of the stochastic effects as if the outcomes of all uncertainties were known in advance. We explore issues involved in designing, conducting and analyzing simulation experiments using a variant of the technique that addresses situations where it makes sense to define the *best* solution as one that is most likely to produce a desirable outcome.

## 1 INTRODUCTION

Generally speaking, the problem of optimizing a stochastic system is to find the values for a set of controllable factors such that some characteristic of a random system response is optimized. Specifically, let $\theta \in \Theta$ be the vector of decision parameters whose values are constrained to lie in a set of feasible designs, $\Theta$. The random system response, $f(\theta,\omega)$, is some function of $\theta$ and of the sample path $\omega$ which represents the outcome of stochastic effects in the system.

The goal is to find $\theta^*$ such that

$$\theta^* = \underset{\theta \in \Theta}{argopt}\ \alpha(\theta) \qquad (1)$$

where $\alpha(\theta)$ is the (non-random) real-valued function of $f(\theta,\omega)$ we wish to optimize. The symbol *argopt* indicates that the optimization is performed over feasible values of the argument $\theta$ and that the set of corresponding optimizing values is returned.

Typically, the characteristic of $f(\theta,\omega)$ we wish to optimize is its expectation with respect to a probability measure $P$ over some measurable space $(\Omega, \mathcal{F})$ in which case

$$\alpha(\theta) = \int_\Omega f(\theta,\omega) dP(\omega) \qquad (2)$$

If the relationships that govern the operation of the system are simple enough, it may be possible to use analytic techniques to determine $\theta^*$ exactly. In many situations, a closed form expression for $\theta^*$ or even $\alpha(\theta)$ is unobtainable. In such cases, an empirical methodology is often employed. In simulation, outcomes of $f(\theta,\omega)$ are computed from artificially generated realizations the of stochastic effects, $\omega$, and the resultant data are used to estimate the location of the optimum.

In this paper we discuss how a class of simulation-based techniques designed to solve the problem as defined by (1) and (2) can be used when the goal is instead to find the policy that has the highest probability of yielding the most desirable result.

## 2 Retrospective Optimization

Most conventional simulation-based optimization methods are what might be characterized as *prospective* approaches in the sense that they involve selecting values of $\theta$ according to some design or algorithm and then replicating the experiment (usually many times) at each of the selected values to estimate $\alpha(\theta)$ and, indirectly, the location of the optimum. In contrast, there is another class of what we call *retrospective* techniques (Healy 1992, Schruben 1991) that can in some cases be used to directly estimate the location of the optimum. Whereas prospective simulation experiments are designed to address the question - *"What would happen at a fixed value of $\theta$?"* - retrospective experiments are designed to answer the more pertinent question - *What value of $\theta$ should have been used?"*. Instead of fixing a value of $\theta$, the idea is to first generate realizations of the stochastic effects and then optimize with respect to those realizations as if

the outcomes of all uncertainties had been known in advance.

More specifically, the realizations are used to construct a functional estimate of $\alpha(\theta)$ over all values of $\theta \in \Theta$. An estimate of $\theta^*$ is obtained by directly optimizing over this approximating functional. If the functional approximation to $\alpha(\theta)$ is good, it seems reasonable that its corresponding optimizer would be, in some sense, close to $\theta^*$.

The functional approximation is constructed by replacing the the driving measure $P$ in (2) with the empirical measure, $P_n$, which assigns a mass of $1/n$ to each of $n$ independent realizations of the stochastic effects $\{\omega_1, \omega_2, ..., \omega_n\}$ so that

$$\begin{aligned} \alpha_n(\theta) &= \int_\Omega f(\theta,\omega) dP_n(\omega) \\ &= \frac{1}{n}\sum_{i=1}^{n} f(\theta,\omega_i). \end{aligned} \tag{3}$$

Furthermore, define

$$\theta_n^* = \underset{\theta\in\Theta}{argopt}\ \alpha_n(\theta) \tag{4}$$

to be the (random) value of $\theta$ that optimizes the sample performance function $\alpha_n(\theta)$. Once the $\omega_i$'s are fixed, the evolution of the system is deterministic so a realization of $\theta_n^*$ is obtained by simulating the stochastic effects, $\omega_1, \omega_2, ..., \omega_n$, and retrospectively solving the corresponding deterministic optimization problem.

In one sense, this functional estimate can be viewed as simply an efficient implementation of the method of common random numbers where the same set of realizations is employed in the estimation of the response for each value of $\theta \in \Theta$. The distinction is that we want to simultaneously estimate $\alpha(\theta)$ for all $\theta \in \Theta$ without re-simulating. The validity of the approach depends on the limiting properties of the solution to the sample path optimization problem. As a practical matter, one must also be concerned about the computational effort involved in obtaining a solution to the sample path problem.

The notion of optimizing over sample paths arises in many different contexts. Statistical parameter estimation techniques like least squares regression and the method of maximum likelihood are just two examples. The approach presented here has been employed in obtaining numerical solutions to more general *stochastic programming problems* where the feasible region $\Theta$ is defined by constraints that, like the objective function, take the form of (2) (see for instance, Wets 1989, Kall 1987, or Birge and Wets 1986.)

The technique requires that there can be no dependence on $\theta$ in the representation of the sample path, $\omega$, or its driving distribution, $P$. Otherwise, generating outcomes of $\omega$ would require that we fix values of the parameters before conducting the experiment. The dependence can be always be eliminated, at least in theory, by redefining the sample space $\Omega$ and, accordingly, the driving distribution $P$. For instance, it is common in simulation to view the basic random element $\omega$ as a sequence of independent $U(0,1)$ variates in which case all $\theta$ dependence resides in the sample performance, $f(\theta,\omega)$.

In most situations though it is difficult, if not impossible, to derive a computationally useful closed-form expression for $f(\theta,\omega)$ for all $\theta \in \Theta$ and $\omega \in \Omega$. Indeed, this is usually the primary reason for resorting to simulation where the transformation $f(\theta,\omega)$ is expressed in algorithmic form. Even in cases where this can be done, the formulation and solution of the sample path problem must be performed on a case-by-case basis being heavily dependent on the characteristics of the system, the nature of the performance measure, and parameters of interest. Nevertheless, there are many meaningful situations where this can be accomplished (see for instance, Healy and Xu 1994, Fu and Healy 1992, Healy 1992, and Plambeck, et al. 1994).

A dependence on $\theta$ in the driving distribution can also be eliminated without re-defining the underlying sample space by instead employing a change-of-measure. This approach, based on the idea of importance sampling (Goyal, Heidelberger and Shahabaddin 1987), is the distinguishing characteristic of the so-called *likelihood ratio* (Glynn 1987, Reiman and Weiss 1989) and *score function* (Rubinstein and Shapiro 1993, Rubinstein 1991, Rubinstein 1989, Rubinstein 1986) methods which we also classify as retrospective techniques. The result in all cases is a representation of $\alpha(\theta)$ as in (2) where the driving distribution is independent of $\theta$; however, there exist fundamental differences among the properties of the resulting re-structured sample performance measures (L'Ecuyer 1990).

In any case, it seems plausible that an estimate of $\theta^*$ can be obtained by optimizing over a single realization of $f(\theta,\omega)$

$$\theta_1^* = \underset{\theta\in\Theta}{argopt}\ f(\theta,\omega) \tag{5}$$

While $f(\theta,\omega)$ is unbiased for $\alpha(\theta)$, this property is not generally invariant to functional transformations like maximization and minimization. To see this, assume the problem involves the minimization of $\alpha(\theta)$. Since $\min\{\cdot\}$ is a concave transformation, it follows from

Jensen's inequality (Billingsley 1968) that

$$\begin{aligned} \min\{\ E[f(\theta,\omega)]\ \} &\geq E[\min\{\ f(\theta,\omega)\ \}] \\ \alpha(\theta^*) &\geq E[f(\theta_1^*,\omega)] \end{aligned}$$

with equality holding only in degenerate or trivial instances. The same argument can be be used to show that $E[f(\theta_1^*,\omega)]$ is an upper bound on $\alpha(\theta^*)$ when the problem involves maximization. This does not imply that $\theta_1^*$, the optimizer over a single realization of $\omega$, is necessarily biased for $\theta^*$. Examples where $\theta_1^*$ is unbiased for $\theta^*$ are not difficult to contrive, but they represent the exception rather than the norm.

The problem, of course, is that a reasonable approximation to the driving distribution $P$ cannot be constructed from a single realization, $\omega$. If the estimate of $\alpha(\theta)$ is instead based on $n$ independent realizations of $\omega$, we still have

$$\begin{aligned} \min\{\ E[\alpha_n(\theta)]\ \} &\geq E[\min\{\ \alpha_n(\theta)\ \}] \\ \alpha(\theta^*) &\geq E[\alpha_n(\theta_n^*)] \end{aligned}$$

where $\alpha_n(\theta)$ and $\theta_n^*$ are defined by (3) and (4), respectively. However, under certain conditions (see for instance, Healy and Xu 1994, Robinson 1994, King and Wets 1991, or Dupacova and Wets 1988)

$$\alpha_n(\theta_n^*) \longrightarrow \alpha(\theta^*) \quad w.p.\ 1.$$

Without making claims regarding actual policy convergence, the result asserts that solutions obtained from the procedure are good in the sense that the sequence of optimal sample path performance values is strongly consistent. From a practical standpoint, this might be all that matters.

## 3 An Alternate Optimality Criterion

It is interesting to note that while

$$\theta^* = \underset{\theta\in\Theta}{argopt}\ E[f(\theta,\omega)]$$

(where the characteristic of $f(\theta,\omega)$ to optimize is its expectation) might be regarded as "*the*" problem, $\theta_1^*$, the solution to the single-realization problem defined by (5) has some virtues beyond the fact that it is sometimes easier to solve. An expectation might be appropriate for evaluating the performance of a policy that is implemented on a recurrent basis but for those that aren't, it might involve significant risk.

For example, consider the choice between two strategies both of which have the same expected payoff. The first strategy results in a large payoff with small probability and no payoff with high probability. Conversely, the second strategy produces a small payoff with high probability and no payoff with low probability. Although the expected payoff is the same, the two strategies entail different elements of risk. In this situation, it might make sense to define the *best* policy to be the one that is most likely to produce a good outcome in which case $\theta^*$ can be interpreted as the solution with the maximum likelihood of being optimal for the next outcome. In many real-world settings, this *next* outcome may occur only once. That is, the real system is exercised under just a single realization of $\omega$.

When the parameter space $\Theta$ is discrete, say,

$$\Theta = \{\theta_1,\theta_2,...,\theta_k\}$$

let

$$p_i = P\{\omega : \theta_i = \underset{\theta\in\Theta}{argopt}\ f(\theta,\omega)\}$$

represent the probability that $\theta_i$, i=1,2,...,k is optimal. Then

$$\theta^* = \theta_{[k]}$$

where $[k]$ denotes the index of the largest of the $p_i$'s, i.e.

$$p_{[1]} \leq p_{[2]} \leq \cdots \leq p_{[k]}.$$

An unbiased retrospective estimate of $\theta^*$ is obtained by determining the most frequently occurring result among a number of observations of the solution to the single realization sample path optimization problem defined by (5).

In the prospective approach to the problem (Goldsman 1984), every trial requires that a separate simulation be conducted for each value of $\theta$ and the best chosen from among the results. If a common realization $\omega$ is employed in all $k$ simulations, the prospective observation is the same as the retrospective estimate. In this case, the retrospective distinction applies only if there exists some means to evaluate $f(\theta,\omega)$, either algebraically or algorithmically, at different values of $\theta$ making separate simulations unnecessary.

If $N_i$, i=1,2,...,k, represents the number of times $\theta_i$ is optimal in $n$ iid observations of $\theta_1^*$ then the joint mass function of $(N_1, N_2, \ldots, N_k)$ is

$$f(n_1,\ldots,n_k) = \binom{n}{n_1\ldots n_k} p_1^{n_1}\cdots p_k^{n_k}$$

and the problem can be modeled as a multinomial sampling procedure with the goal of identifying the

cell with the largest underlying probability of occurrence. We estimate each $p_i$ by

$$\hat{p}_i = N_i/n$$

and their ordering by

$$\hat{p}_{[1]} \leq \hat{p}_{[2]} \leq \cdots \leq \hat{p}_{[k]}$$

Guidelines for conducting such experiments can be found in the copious literature on statistical ranking and selection procedures. The most commonly used techniques fall generally into one of two categories, the so-called *indifference zone* (Bechhofer and Elmaghraby 1959) and *subset* (Gupta and Panchapakesan 1979) approaches. Indifference zone approaches prescribe the sample size, $n$, necessary to guarantee the probability of making a correct selection is no less than some prespecified value. In the process, they allow for the analyst's indifference to errors when the probability associated with a chosen cell is within a specified amount of the optimum. In contrast, subset procedures are designed to identify a small but nonempty subset of the $k$ cells such that the probability that the optimal cell is among those chosen is no smaller than some prespecified value. In this case, there is no indifference zone. The sample size is fixed by the analyst and the size of the selected subset is random.

## REFERENCES


R.E. Bechhofer and S. Elmaghraby. 1959. *A single-sample multiple-decision procedure for selecting the multinomial event which has the highest probability.* Annals of Mathematical Statistics 30:102–119.

P. Billingsley. 1968. *Convergence of Probability Measures* . John Wiley & Sons, Inc.

J.R. Birge and R.J-B. Wets. 1986. Designing approximation schemes for stochastic optimization problems, in particular, for stochastic programs with recourse. *Mathematical Programming Study* 27:54-102.

J. Dupacova and R.J-B. Wets. 1988. Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems. *Annals of Statistics* 16:1517–1549.

M. Fu and K. Healy. 1992. Simulation optimization of (S,s) inventory systems. In *Proceedings of the 1992 Winter Simulation Conference*, ed. J.J. Swain, D. Goldsman, R.C. Crain, and J.R. Wilson, 506–514.

P. Glynn. 1987. Likelihood ratio gradient estimation: an overview. In *Proceedings of the 1987 Winter Simulation Conference*, 366–375.

D. Goldsman. 1984. A multinomial ranking and selection procedure: simulation and applications. In *Proceedings of the 1984 Winter Simulation Conference.* ed. S. Sheppard, U. Pooch. D. Pedgen, 259–264.

A. Goyal, Heidelberger and P. Shahabaddin. 1987. Measure specific dynamic importance sampling for availability simulations. In *Proceedings of 1987 Winter Simulation Conference.*

S. S. Gupta and S. Panchapakesan. 1979. *Multiple Decision Procedures.* John Wiley, NY.

K. Healy. 1992. Optimizing stochastic systems: a retrospect/deterministic approach. Ph.D. dissertation, Department of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY.

K. Healy and Y. Xu. 1994. Simulation-based retrospective approaches to stochastic systems optimization. Research Memorandum 94-?, School of Industrial Engineering, Purdue University, West Lafayette, IN.

P. Kall.1987. On approximations and stability in stochastic programming. In *Parametric Optimization and Related Topics*, ed. J. Guddat, et al., Akademie-Verlag, Berlin, 387–407.

A.J. King and R.J-B. Wets. 1991. Epi-consistency of convex stochastic programs. *Stochastics and Stochastics Reports* 34:83–92.

P. L'Ecuyer. 1990. A unified view of of the ipa, sf, and lr gradient estimation techniques. *Management Science* 36:1364–1383.

E. Plambeck, B.-R. Fu, S. Robinson and R. Suri. 1994. *Sample Path Optimization of Convex Stochastic Performance Functions.* Submitted for Publication.

M.I. Reiman and A. Weiss. 1989. Sensitivity analysis for simulations via likelihood ratios. *Operations Research* 37:830–844.

S. Robinson. 1994. *Analysis of sample-path optimization.* Submitted for publication.

R. Y. Rubinstein. 1991. How to optimize discrete-event systems from a single sample path by the score function method. *Annals of Operations Research* 27:175-212.

R. Y. Rubinstein. 1989 Sensitivity analysis of computer simulation models via the efficient score. *Operations Research*, 37:72–81.

R. Y. Rubinstein. 1986. The score function approach for sensitivity analysis of computer simulation models. *Mathematics and Computers in Simulation* 28:251–379.

R. Y. Rubinstein and A. Shapiro. 1993. *Discrete event systems: sensitivity analysis and stochastic optimization by the score function method.* John

Wiley & Sons, Inc..

L. W. Schruben. 1991. Path-wise decomposition and retrospective optimization of some simulation experiments. Technical Report NPSOR-9119, Naval Post Graduate School, Monterrey, CA.

R. J-B. Wets. 1989. Stochastic programming. In *Handbooks in Operations Research and Management Science, Volume 1: Optimization*, ed. G.L. Nemhauser, A.H.G. Rinnooy Kan, and M.J. Todd, 573–630. North Holland, Amsterdam.

## AUTHOR BIOGRAPHY

KEVIN J. HEALY is an assistant professor in the School of Industrial Engineering at Purdue University. He earned B.S. and M.S. degrees in Industrial Engineering from The Pennsylvania State University and a Ph.D. in Operations Research from Cornell University. His research interests include stochastic optimization, simulation output analysis, and modeling formalisms.