Proceedings of the 1986 Winter Simulation Conference
J. Wilson, J. Henriksen, S. Roberts (eds.)

Sensitivity Analysis Via Likelihood Ratios

Martin I. Reiman
and
Alan Weiss

AT&T Bell Laboratories
Murray Hill, New Jersey 07974

ABSTRACT

We present a new method of obtaining derivatives of expectations with respect to various parameters. For example, if λ is the rate of a Poisson process, N_T is the number of Poisson events in $(0,T)$, and ψ is nearly any function of the sample path (e.g. a performance measure in a queuing network), then we show that

$$\frac{d}{d\lambda} E_\lambda(\psi) = E_\lambda\left[\left(\frac{N_T}{\lambda} - T\right)\psi\right],$$

which yields an obvious algorithm. We have proven that the method works for a wide class of parameters and performance measures in regenerative simulation. We also report on the method's limitations and on some numerical experiments.

1. Introduction

This paper outlines a new method of sensitivity analysis that applies to transient quantities and to regenerative simulation of steady-state quantities. We interpret sensitivity as derivatives of expectations with respect to parameters. To emphasize the dependence of the expectation of a random variable ψ on a parameter λ, we will write $E_\lambda(\psi)$ for the mean of ψ. The probability measure of the stochastic system under study will be written P_λ, so that

$$E_\lambda(\psi) = \int_\Omega \psi(\omega)\, dP_\lambda(\omega).$$

The most important assumption we make about P_λ is that it is composed of a product of two measures, R_λ and Q ($P_\lambda(A) = R_\lambda(A)\cdot Q(A)$), where Q does not depend on λ, and R depends on λ only in the simplest manner. For example, if λ represents the rate of a Poisson process associated with a queuing network, then $P_\lambda = R_\lambda \cdot Q$, where R_λ is a standard Poisson process, and Q represents all the rest of the dynamics of the system. This requirement will become clear in the course of the paper; in practice, it is not a restriction on the systems we can analyze.

The random variables we consider, denoted ψ, are perfectly general. For example, in a queuing network, we might have $\psi = T^{-1}\int_0^T q(s)\,ds$, where T is the time we observe the system, and $q(s)$ is the queue length at a particular station. We might have $\psi = T^n$ where T is a stopping time; we could have $\psi = f(X(T))$ where f is an arbitrary function and $X(t)$ is the state of a system; in short, ψ is any $\mathcal{F}_T$-measurable function. (We have in mind a probability space $(\Omega, P_\lambda, \{\mathcal{F}_t\})$ where $\mathcal{F}_t$ is a natural increasing set of σ-fields associated with a stopping time T.)

Our method is reminiscent of importance sampling [Siegmund (1976)], and special cases of our theorems are well known to statisticians [Bickel-Doksum (1977) eq. 4.3.10] and workers in statistical mechanics [e.g. Tolman (1979) Sec. 93]. Rubinstein (1986) has come up with a method that may be regarded as a special case of ours. Our contribution appears to be mainly the synthesis of existing techniques into a coherent method; however, our theorem 4 seems to be new.

2. The Basic Theorems

In this section we will outline the main theorems that underlie our method. The applications to sensitivity analysis will be discussed subsequently. Our first theorem concerns Poisson rates.

Theorem 1: Let λ denote the rate of a Poisson process in a stochastic system, and let T be a stopping time, ψ an $\mathcal{F}_T$-measurable random variable, and N the number of Poisson events in the interval $(0,T]$. If $E_\lambda(\psi)$ is the expectation of ψ over time $(0,T]$, then under A1*

$$\frac{d}{d\lambda} E_\lambda(\psi) = E_\lambda\left[\left(\frac{N}{\lambda} - T\right)\psi\right]. \tag{2.1}$$

Proof: (outline)

$$E_{\lambda+\delta}(\psi) = \int\psi(\omega)\, dP_{\lambda+\delta}(\omega) = \int\psi(\omega)\,\frac{dP_{\lambda+\delta}}{dP_\lambda}(\omega)\, dP_\lambda(\omega)$$

$$= \int\psi(\omega)\,\frac{dR_{\lambda+\delta}}{dR_\lambda}(\omega)\, dP_\lambda(\omega),$$

where of course $\frac{dP_{\lambda+\delta}}{dP_\lambda}(\omega)$ and $\frac{dR_{\lambda+\delta}}{dR_\lambda}(\omega)$ are Radon-Nikodym derivatives. Now if limits and integration are interchangeable,

$$\frac{d}{d\lambda}E_\lambda(\psi) = \lim_{\delta\to 0}\frac{E_{\lambda+\delta}(\psi) - E_\lambda(\psi)}{\delta}$$

$$= \int\psi(\omega)\lim_{\delta\to 0}\left[\frac{\frac{dR_{\lambda+\delta}}{dR_\lambda}(\omega) - 1}{\delta}\right] dP_\lambda(\omega).$$

We claim that (see below for an explanation)

* The technical conditions under which our theorems are proved appear in the appendix.

$$\lim_{\delta\to 0}\frac{\frac{dR_{\lambda+\delta}}{dR_\lambda}(\omega)-1}{\delta}=\frac{N(\omega)}{\lambda}-T(\omega), \tag{2.2}$$

so that

$$\frac{d}{d\lambda}E_\lambda(\psi)=\int\psi(\omega)\left[\frac{N(\omega)}{\lambda}-T(\omega)\right]dP_\lambda(\omega)$$
$$=E_\lambda\left[\left[\frac{N}{\lambda}-T\right]\psi\right].$$

This more or less completes the proof. For more details, including the proof that the interchange is valid, see Reiman and Weiss (1986).

In order to make (2.2) plausible, consider the likelihood of a Poisson path ω on $(0,T)$ that has events at times $t_1, t_2, \ldots, t_N$. The probability that there is no event in $(0,t_1)$ is $e^{-\lambda t_1}$; the probability that there is an event in (t_1, t_1+dt) is λdt; the probability that there is no event in (t_1, t_2) is $e^{-\lambda(t_2-t_1)}$; the probability that there is an event in (t_2, t_2+dt) is λdt; etc. That is,

$$\text{Likelihood}_\lambda(\omega)=$$
$$e^{-\lambda t_1}\lambda dt\, e^{-\lambda(t_2-t_1)}\lambda dt\cdots e^{-\lambda(t_N-t_{N-1})}\lambda dt\, e^{-\lambda(T-t_N)}$$
$$=e^{-\lambda T}\lambda^N dt^N \text{ ``}=\text{''} dR_\lambda(\omega).$$

If we consider the likelihood of the same path under a Poisson process with rate $\lambda+\delta$, then

$$\text{Likelihood}_{\lambda+\delta}(\text{path})=e^{-(\lambda+\delta)T}(\lambda+\delta)^N dt^N \text{ ``}=\text{''} dR_{\lambda+\delta}(\omega).$$

Hence we obtain

$$\frac{dR_{\lambda+\delta}}{dR_\lambda}=e^{-\delta T}\left(1+\frac{\delta}{\lambda}\right)^N,$$

so that (2.2) follows.

Corollary 1: Under the assumptions of Theorem 1, for $K\geqslant 1$,

$$\frac{d^K}{d\lambda^K}E_\lambda(\psi)=E_\lambda\left[\left[\frac{N}{\lambda}-T\right]_0^K\psi\right]$$

where

$$\left[\frac{N}{\lambda}-T\right]_0^K\equiv\sum_{j=0}^K\binom{K}{j}\frac{N_1^j}{\lambda^j}(-T)^{K-j}$$

and

$$N_1^j\equiv N(N-1)\cdots(N-j+1).$$

Corollary 2: If $(\lambda_1,\ldots,\lambda_K)$ denote the rates of K independent Poisson processes, and N_T^i denotes the number of type i events in $(0,T]$, then

$$\nabla_\lambda E_\lambda(\psi)=E_\lambda\left[\psi\left[\frac{N_T^1}{\lambda_1}-T,\frac{N_T^2}{\lambda_2}-T,\ldots,\frac{N_T^K}{\lambda_K}-T\right]\right]$$

where ∇_λ means the gradient with respect to λ.

We now turn our attention to simple random variables. We start with a

Definition: X is a *simple* random variable if it has a finite number of possible outcomes $(x_1,\ldots,x_K)$ with associated probabilities $p_1,\ldots,p_K$.

We wish to estimate the effect of changing the weights $p_1,\ldots,p_K$ to $p_1+\delta v_1,\ldots,p_K+\delta v_K$. Clearly $\sum_{j=1}^K v_j=0$ in order for the probabilities to sum to one. We let N_i denote the number of instances of X that turn out to be x_i during the period of observation.

Theorem 2: Under conditions A2,

$$\lim_{\delta\to 0}\frac{E_{p+\delta v}(\psi)-E_p(\psi)}{\delta}=E_p\left[\psi\left[\sum_{j=1}^K\frac{N_j}{p_j}v_j\right]\right] \tag{2.3}$$

Proof:

$$\frac{dP_{p+\delta v}}{dP_p}=\left(1+\frac{\delta v_1}{p_1}\right)^{N_1}\cdots\left(1+\frac{\delta v_K}{p_K}\right)^{N_K}$$

and argue as in Theorem 1.

We now consider continuous random variables. Suppose that we have a family of densities f_δ, $-\epsilon<\delta<\epsilon$, with

$$\lim_{\delta\to 0}f_\delta(x)=f_0(x)$$

$$\lim_{\delta\to 0}\frac{f_\delta(x)-f_0(x)}{\delta}=v(x).$$

We denote the outcomes of the random variable during the period of observation as $x_1,\ldots,x_N$.

Theorem 3: Under conditions A3,

$$\frac{d}{d\delta}E_\delta(\psi)|_{\delta=0}=E_0\left[\psi\left[\sum_{i=1}^N\frac{v}{f_0}(x_i)\right]\right]. \tag{2.4}$$

Remark: The three theorems have many generalizations and corollaries. For example, state- and time-dependent rates and probabilities can be handled, as can higher and mixed derivatives. For more details see Reiman and Weiss (1986).

3. Discussion

The preceding section contains mathematical theorems rather than algorithms for simulation. There are three topics we will discuss in this section: heuristics for understanding the theorems, difficulties in implementing algorithms based on the theorems, and some resolutions of those difficulties. The following section will address sensitivity analysis in regenerative simulation, where the difficulties disappear.

First we want to discuss why the theorems are true — after all, it is not clear *a priori* that one can obtain derivatives without changing something. The proofs provide one form of understanding, and the following discussion should provide a more intuitive one. Consider equation (2.1), where λ is a Poisson parameter

$$\frac{d}{d\lambda}E_\lambda(\psi) = E_\lambda\left[\left(\frac{N}{\lambda} - T\right)\psi\right] \tag{3.1}$$
$$= E_\lambda((N/T - \lambda)(\lambda T)\psi).$$

The first factor on the right hand side of (3.1) is the difference between the observed rate N/T and the theoretical rate λ. When, by random fluctuations, N/T is larger than λ, then the system behaves as if the rate is higher than λ, and we multiply the response ψ by a positive quantity. When N/T is less than λ, the system acts as if the rate is less than λ, and ψ is multiplied by a proportionally negative quantity. Now the other factor λT represents the confidence we have in our observation ψ; if λT is small, then we do not have many observations and so the estimate of ψ will have a large relative error, but if λT is large then we have a good deal of faith in the observation, so put a large weight on it. In summary, it is the correlation between $\left[\frac{N}{\lambda} - T\right]$ and ψ that gives us the estimate of $\frac{d}{d\lambda}E_\lambda(\psi)$, and we hope that the reason for this is clearer now.

At first it would seem trivial to implement sensitivity analysis using formulae such as (2.1) or (2.3). It appears that keeping an extra statistic during the course of simulation will provide an unbiased estimate. However, this will not always work: the variance of the estimator can be prohibitively large. For example, even if ψ settles down as $T \to \infty$, then $\left[\frac{N}{\lambda} - T\right]\psi$ will have variance proportional to T, so a long run will give a poor estimate! This is not a problem in regenerative simulation, as will be seen in the next section, because there we take T to be the length of a regeneration cycle. Also, the formulae can be used quite freely for transient analysis. There one takes many replications of runs of length T. Hence the problem is in the standard simulation of steady-state quantities.

Unfortunately, we do not have a satisfactory solution as yet to this problem. The most obvious method of lowering the variance of the estimator is to take T small, as in the method of batched means, by dividing time into many intervals. However, this leads to biased estimates of steady-state quantities, as the following simple calculation shows:

$$\frac{d}{d\lambda}E^{\mu(\lambda)}_{\lambda,T}(\psi) = \tag{3.2}$$
$$\frac{\partial}{\partial\lambda}E^{\mu(\lambda)}_{\lambda,T}(\psi) + \left\langle\frac{\partial}{\partial\mu}E^{\mu(\lambda)}_{\lambda,T}(\psi),\ \frac{\partial\mu(\lambda)}{\partial\lambda}\right\rangle$$

where $E^{\mu}_{\lambda,T}(\psi)$ is the mean value of ψ over an interval of length T with initial distribution μ, and $\mu(\lambda)$ is the steady-state distribution. Our theorems give estimates of the first term in the right hand side of (3.2); that is, the change of initial distribution with λ is *not* modeled. It is usually true that

$$\lim_{T\to\infty}\left\langle\frac{\partial}{\partial\mu}E^{\mu}_{\lambda,T}(\psi),\ \frac{\partial\mu}{\partial\lambda}\right\rangle = 0$$

when $\lim_{T\to\infty}E^{\mu(\lambda)}_{\lambda,T}(\psi)$ exists. Now we see the tradeoff: as T becomes large, our estimator has higher variance, and as T becomes small, the bias in the estimator increases. [Under some reasonable assumptions one can estimate that the optimal value of T is proportional to $N^{1/4}$ for a simulation of length N, so that the standard error of the estimate of $\frac{d}{d\lambda}E_\lambda(\psi)$ is $O(N^{-1/4})$, making a fairly poor estimator.] All this means that more work needs to be done before the method can be easily applied to the non-regenerative simulation of steady-state quantities.

4. Regenerative Simulation

We give a very brief introduction to regenerative simulation, the main intent of which is to introduce our notation. For a more thorough and detailed treatment the reader is referred to Iglehart and Shedler (1980).

It is well known from the theory of regenerative processes that many steady state quantities (for a regenerative process) can be expressed as the ratio of two expected values, e.g.

$$r = E[X] = \frac{E[Y]}{E[\alpha]}, \tag{4.1}$$

where (Y,α) depends only on one regenerative cycle. The basic idea of regenerative simulation is to simulate the system for n cycles, generating iid random vectors (Y_i,α_i), $1 \leqslant i \leqslant n$. We can then form the estimates

$$\bar{Y}_n = n^{-1}\sum_{i=1}^{n} Y_i,$$
$$\bar{\alpha}_n = n^{-1}\sum_{i=1}^{n} \alpha_i,$$

and

$$\hat{r}_n = \bar{Y}_n/\bar{\alpha}_n,$$

for $E[Y]$, $E[\alpha]$, and $E[X]$, respectively.

To show how to apply our method to obtain derivative estimates we will consider a system having a Poisson arrival process with rate λ and differentiate with respect to λ. Note that the quotient rule applied to (4.1) yields

$$d \equiv \frac{d}{d\lambda}E_\lambda[X] = \frac{\frac{d}{d\lambda}E_\lambda[Y]}{E_\lambda[\alpha]} - \frac{\frac{d}{d\lambda}E_\lambda[\alpha]}{E_\lambda[\alpha]}\frac{E_\lambda[Y]}{E_\lambda[\alpha]}. \tag{4.2}$$

We define

$$Y^{(1)} = Y, \tag{4.3a}$$
$$Y^{(2)} = \left[\frac{N}{\lambda} - T\right]Y, \tag{4.3b}$$
and
$$Y^{(3)} = \left[\frac{N}{\lambda} - T\right]\alpha. \tag{4.3c}$$

Using equations (2.1) and (4.3) we can rewrite (4.2) as

$$\frac{d}{d\lambda}E_\lambda[X] = \frac{E_\lambda[Y^{(2)}]}{E_\lambda[\alpha]} - \frac{E_\lambda[Y^{(3)}]}{E_\lambda[\alpha]}\frac{E_\lambda[Y^{(1)}]}{E_\lambda[\alpha]}. \tag{4.4}$$

If we wanted to differentiate with respect to a different parameter, we would (by equation (2.3) or (2.4)) end up with different expressions for $Y^{(2)}$ and $Y^{(3)}$. The important point is that the form of (4.2) and hence (4.4) would be unchanged. Higher derivatives can be obtained by repeatedly differentiating (4.1).

To obtain a point estimate for d, we simulate for n cycles, generating iid random vectors $(Y_i^{(1)}, Y_i^{(2)}, Y_i^{(3)}, \alpha_i)$, $1 \leqslant i \leqslant n$, and form the estimates

$$\bar{Y}_n^{(k)} = n^{-1} \sum_{i=1}^{n} Y_i^{(k)}, \quad 1 \leqslant k \leqslant 3,$$

$$\bar{\alpha}_n = n^{-1} \sum_{i=1}^{n} \alpha_i,$$

$$\hat{r}_n^{(k)} = \bar{Y}_n^{(k)}/\bar{\alpha}_n, \quad 1 \leqslant k \leqslant 3,$$

and

$$\hat{d}_n = \hat{r}_n^{(2)} - \hat{r}_n^{(1)} \hat{r}_n^{(3)}.$$

Confidence Intervals

An appealing feature of regenerative simulation is that the iid random vectors used to form point estimates also yield confidence intervals, via the central limit theorem. We will simply present expressions for the covariance matrix associated with first derivative estimates; for a derivation, see Reiman and Weiss (1986).

For $n \geqslant 1$ and $1 \leqslant k \leqslant 3$ let

$$Z_i^{(k)} = Y_i^{(k)} - r^{(i)}\alpha_i, \quad 1 \leqslant i \leqslant n,$$

$$\tilde{r}_n^{(k)} = \sqrt{n}\left[\hat{r}_n^{(k)} - r^{(k)}\right],$$

and $$\tilde{d}_n = \sqrt{n}\left[\hat{d}_n - d\right],$$

where $r^{(k)} = E[Y^{(k)}]/E[\alpha^{(k)}]$.

Theorem 4: $(\tilde{r}_n^{(1)}, \tilde{d}_n) \Rightarrow N(0, C)$, where

$$C_{11} = B_{11},$$

$$C_{22} = B_{22} + (r^{(1)})^2 B_{33} + (r^{(3)})^2 B_{11} - 2r^{(1)}B_{23} - 2r^{(3)}B_{12} + 2r^{(1)}r^{(3)}B_{13},$$

and $$C_{12} = C_{21} = B_{12} - r^{(1)}B_{13} - r^{(3)}B_{11},$$

with

$$B_{jk} = cov\left[Z_1^{(j)}, Z_1^{(k)}\right]/(E[\alpha])^2.$$

The covariance matrix B can be estimated in the simulation, leading to an estimate of C. Confidence intervals can then be constructed in the standard manner.

Experimental Results

We applied the method described above in the regenerative simulation of an $M/G/1$ queue. We chose a hyperexponential distribution with two phases, so that with probability p the service time is exponential with parameter 2/3 and with probability $1-p$ the service time is exponential with parameter 2. This gives us two parameters to differentiate with respect to, λ and p.

In the simulation we estimate the mean sojourn time as well as its derivative with respect to λ and p. We also obtained the covariance matrix associated with each derivative. Five trials of 20,000 busy periods were run at $\lambda = 1/2$, $p = 1/2$. The results are presented in Table 1.

The derivative calculation added about 20 lines of code to the simulation program (written in FORTRAN), most of which are assignment statements. With covariance calculations the derivatives required 25% extra cpu time; without the covariance calculation the overhead was under 3%.

	$E[w]$	$\sigma(w)$	$\frac{d}{d\lambda}E[w]$	$\sigma\left(\frac{d}{d\lambda}E[w]\right)$	corr $\left(E[w], \frac{d}{d\lambda}E[w]\right)$	$\frac{d}{dp}E[w]$	$\sigma\left(\frac{d}{dp}E[w]\right)$	corr $\left(E[w], \frac{d}{dp}E[w]\right)$
Theory	2.250	—	5.000	—	—	−4.250	—	—
Trial 1	2.246	.040	5.032	.522	.673	−3.754	.448	−.350
Trial 2	2.236	.040	4.326	.402	.649	−4.719	.437	−.579
Trial 3	2.247	.042	4.336	.452	.624	−4.960	.473	−.623
Trial 4	2.239	.039	4.415	.480	.655	−4.660	.440	−.512
Trial 5	2.233	.045	5.521	.764	.754	−4.864	.644	−.654

Table 1. Mean sojourn time

APPENDIX

Definition: An F_T-measurable random variable ψ is *amiable* with respect to (f,c,β,s) if there exists a nonnegative F_T-measurable random variable f and positive numbers c,β, and s with $s > \beta$ and

$$|\psi| \leqslant ce^{\beta f}$$

and

$$E(e^{sf}) < \infty.$$

Assumption A1: ψ is amiable with respect to the same (f,c,β,s) for all λ in an open neighborhood, and $f \geqslant T+N_T$.

Assumption A2: $v_j = 0$ whenever $p_j = 0$, and ψ is amiable with respect to the same (f,c,β,s) for all vectors $(p+\delta v)$, where δ is in an open neighborhood of 0, and $f \geqslant N_1+\cdots+N_K$.

Assumption A3: ψ is amiable with respect to the same (g,c,β,s) for all δ in an open neighborhood of 0, and

$$g \geqslant \frac{1}{\delta}\sum_{i=1}^{N}\left|\log\frac{f_\delta}{f_0}(x_i)\right|$$

for all δ in the punctured neighborhood.

Note: See Reiman and Weiss (1986) for a discussion of why (A1)-(A3) are not restrictive.

REFERENCES

[1] Bickel, P. J. and Doksum, K. A. "Mathematical Statistics: Basic Ideas with Selected Topics," Holden-Day, San Francisco, 1977.

[2] Reiman, M. I. and Weiss, A. "Sensitivity Analysis for Simulations via Likelihood Ratios," submitted for publication, 1986.

[3] Rubinstein, R. "Sensitivity Analysis and Performance Extrapolation for Computer Simulation Models," submitted for publication, 1986.

[4] Siegmund, D. "Importance Sampling in the Monte Carlo Study of Sequential Tests," Ann. Stat., Vol. 14, No. 4, pp. 673-684, 1976.

[5] Tolman, R. C. "The Principles of Statistical Mechanics," Dover, New York, 1979.