SANDIA REPORT
SAND2020-9871
Printed September 2020

Sandia National Laboratories

Particle Sensitivity Analyses

S. D. Bond, B. C. Franke, R. B. Lehoucq, and S. A. McKinley

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185
Livermore, California 94550

Issued by Sandia National Laboratories, operated for the United States Department of Energy by National Technology & Engineering Solutions of Sandia, LLC.

Available to DOE and DOE contractors from

U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@osti.gov
Online ordering: http://www.osti.gov/scitech

Available to the public from

U.S. Department of Commerce
National Technical Information Service
5301 Shawnee Road
Alexandria, VA 22312

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.gov
Online order: https://classic.ntis.gov/help/order-methods

Particle Sensitivity Analyses

Stephen D. Bond, Brian C. Franke, R. B. Lehoucq
Sandia National Laboratories
P.O. Box 5800
Albuquerque, NM 87185

Scott A. McKinley
Tulane University
6823 St Charles Ave
New Orleans, LA 70118

SAND2020-9871

ABSTRACT

We propose to develop a computational sensitivity analysis capability for Monte Carlo sampling-based particle simulation relevant to Aleph, Cheetah-MC, Empire, Emphasis, ITS, SPARTA, and LAMMPS codes. These software tools model plasmas, radiation transport, low-density fluids, and molecular motion. Our report demonstrates how adjoint optimization methods can be combined with Monte Carlo sampling-based adjoint particle simulation. Our goal is to develop a sensitivity analysis to drive robust design-based optimization for Monte Carlo sampling-based particle simulation—a currently unavailable capability.

ACKNOWLEDGEMENTS

The authors acknowledge helpful conversations with Brian Adams and Don Bruss on the papers Adams et al. [2017, 2019], Pautz et al. [2017] and Drew Kouri over the relationship between adjoint optimization methods and partial differential equations.

Our report was supported by the Laboratory Directed Research and Development program at Sandia National Laboratories, a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Adminstration under contract DE-NA0003525.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

Our goal in this report is to develop the mathematical basis for a sensitivity analysis that ultimately drives a robust design-based optimization for Monte Carlo sampling-based particle simulation—a currently unavailable capability. Computational design-based optimization and uncertainty quantification techniques are well-used tools in science and engineering. Gradient-based methods crucially depend upon sensitivities, which are synonymous with the calculation of a derivative that measures the (instantaneous) change in a quantity with respect to an (instantaneous) change in another quantity. We also demonstrate how adjoint optimization methods can be combined with Monte Carlo sampling-based adjoint particle simulation.

An important class of optimization problems contain partial integral differential equations as constraints so that gradient-based methods require sensitivities for the various functionals involving the solution of these equations. Such problems include models for plasmas, radiation transport, low-density fluids where the partial integral differential equation is the Boltzmann equation or the Fokker-Planck equation when molecular motion is of interest. These equations embody a deterministic model for the aggregate behavior of particles where the solution represents the density of particles. In lieu of approximating the solution via numerical discretization of the partial integral differential equation, a large number of particle trajectories may be averaged via Monte Carlo sampling to approximate the functionals involving the density. Such a sampling represents a statistical approach and is implemented within the Aleph, Cheetah-MC, Empire, Emphasis, ITS, SPARTA, and LAMMPS codes.

The link between such a class of partial integral differential equations and particle trajectories is well-known and is a generalization of the relationship between the diffusion equation and Brownian motion. This link is also available for the adjoint partial integral differential equation and the corresponding particle trajectories. Both deterministic and Monte Carlo sampling-based approaches are useful and the situation at hand dictates whether a deterministic or statistical approach is selected.

Determining a sensitivity during Monte Carlo sampling for a partial integral differential equation is fraught with difficulty since a sample, a trajectory, is not differentiable. Our solution is to exploit tools from probability to develop the needed sensitivities; see for example the book [Glasserman, 2004, Chap.7] for applications to finance and [Asmussen and Glynn, 2007, Chap.VII] for a general treatment. These tools also include stochastic calculus based Malliavin estimators and what we introduce as Stein [1973] equation estimators. The former estimators originated in the mathematical probability literature to obtain regularity estimates of the probability measure in terms of the stochastic process [Malliavin, 1976]; these estimators have a well-understood application within mathematical finance; see e.g., [Chen and Glasserman, 2007]. The latter estimators originate in the mathematical statistics literature to obtain bounds on the distance between two probability distributions. Both Malliavin and Stein equation estimators are related to likelihood ratio based methods. The thread that binds these three classes of estimators is that they enable the sensitivity to be directly computed during the Monte Carlo sampling for the expectation via an integration by parts formula. This was also the pragmatic goal shared by the application-focused papers [Rief, 1994],[Plyasunov and Arkin, 2007],[Warren and Allen, 2012a].

Our report, in Section 2 demonstrates how adjoint optimization methods can be combined with Monte Carlo sampling-based adjoint particle simulation. The estimators are briefly introduced via the vehicle of integration by parts. Section 3 then introduces the Stein estimator and relationship with a Malliavin estimator. Sections 4–6 provide examples and a general theorem that justifies the integration by parts formula listed in section 2.

1.1. Mission-critical application

Our approach is general and so enables us to impact a broad spectrum of Monte Carlo particle simulation codes by a clean separation of the mathematical and the application-specific details. In contrast to an application-centric approach, we believe that our general approach aims to develop and demonstrate capability largely unexplored as a Sandia analysis tool, and, if successful, the results can be leveraged across several distinct mission applications.

A mission-critical application are robust satellite designs that minimize mass so as to lower launch costs. The papers Adams et al. [2017, 2019], Pautz et al. [2017] demonstrate the utility of sensitivities in deterministic particle transport, i.e. the Boltzmann equation is discretized, particle. These successes continue to drive the interest in satellite design, where Monte Carlo particle transport is preferred but yet available. Unfortunately, the current Monte Carlo approach is to fix a design followed by numerous Monte Carlo simulations, one per geometry constraint. Such a scheme provides a rudimentary sensitivity analysis by directly comparing two or more designs. A satellite analysis involves one thousand constraints, taking over three hours on one hundred compute nodes entailing four billion Monte Carlo samples. Our approach would sever the number of Monte Carlo simulations from the number of geometry constraints by instead performing one Monte Carlo simulation, recasting the geometry constraints as sensitivities, concurrently computed.

2. OPTIMIZATION PROBLEM

The optimization problem of interest is

$$\begin{cases} \min_{\theta} \rho(\theta;t,x) \qquad t>0\,, x\in\Omega\subset\mathbb{R}^d \\ \text{subject to } R=\int_{t,x} q(x)\,u(t,x)\,dx\,dt \leqslant R_{\max}\,, \end{cases} \tag{1}$$

where ρ is a scalar function, for example the mass of the object subject to radiation. We refer to R in the constraint inequality as the response functional, an inner product of scalar functions q and u. The response functional, as we shall see, is proportional to an expected value since u is the solution of the partial integral differential equation (PIDE)

$$\frac{\partial}{\partial t}u(t,x)-Lu(t,x)=0\,, \qquad t>0\,, x\in\Omega\subset\mathbb{R}^d\,, \tag{2a}$$

augmented with boundary and initial conditions

$$u(t,x) = g(x) \qquad t > 0\,, x \in \partial\Omega \tag{2b}$$
$$u(0,x) = f(x) \qquad x \in \Omega \subset \mathbb{R}^d \tag{2c}$$

where the backward Kolmogorov operator L is given by

$$\begin{aligned} Lu(t,x) = \frac{a(t,x)}{2}\Delta u(t,x) + b(t,x)\cdot\nabla u(t,x) + \sigma(t,x)\,u(t,x) \\ + \int \big(u(t,y) - u(t,x)\big)\,\eta(y,x)\,dy\,. \end{aligned} \tag{2d}$$

The PIDE is a model for diffusion, drift, absorption and jump-diffusion so that $u(t,x)$ is the density of particles.[1]

Gradient-based optimization approaches for problem (1) depend upon sensitivities of the response functional R in the inequality constraint. Such optimization approaches need response sensitivities, i.e.,

$$\frac{\mathrm{d}}{\mathrm{d}\theta}R = \int_{t,x} \frac{\mathrm{d}q(x)}{\mathrm{d}\theta}\,u(t,x)\,dx\,dt + \int_{t,x} q(x)\,\frac{\mathrm{d}u(t,x)}{\mathrm{d}\theta}\,dx\,dt\,. \tag{3}$$

The challenge is to approximate the second functional, in particular, $\mathrm{d}u/\mathrm{d}\theta$. One approach is to approximate the second functional via quadrature using a discrete approximation to $\mathrm{d}u/\mathrm{d}\theta$. A second approach is to exploit the probabilistic representation of u, which leads us to the concept of a particle sensitivity analysis.

Kolmogorov's representation formula for the initial value problem (2a) is the conditional expectation

$$u(t,x) = \mathbb{E}^x[f(X_t)] = \mathbb{E}[f(X_t)|X_0 = x]\,, \qquad t > 0\,, x \in \Omega = \mathbb{R}^d \tag{4}$$

where we now assume that $\Omega = \mathbb{R}^d$ for simplicity. The particle trajectory X_t is distributed according to the probability measure[2] for the expectation $\mathbb{E}$. The density $u(t,x)$ is then seen to be the average of $f(X_t)$ over all particles that start at x at time $t = 0$. Unfortunately, for many problems the probability measure is unknown. However the trajectory X_t can often be sampled and a Monte Carlo scheme can then be used to approximate $\mathbb{E}^x[f(X_t)]$.

Inserting the representation (4) into (3) leads to

$$\frac{\mathrm{d}}{\mathrm{d}\theta}R = \int_{t,x} \frac{\mathrm{d}q(x)}{\mathrm{d}\theta}\,\mathbb{E}^x[f(X_t)]\,dx\,dt + \int_{t,x} q(x)\,\frac{\mathrm{d}}{\mathrm{d}\theta}\mathbb{E}^x[f(X_t)]\,dx\,dt\,. \tag{5}$$

[1] The problem is steady-state when we assume that $\frac{\partial}{\partial t}u = 0$ so that $u(t,x) = u(x)$ is the equilibrium distribution so that the initial condition f is unnecessary. Instead the boundary condition g can be identified as a source for the steady-state problem.

[2] Also referred to as the transition density, heat kernel or fundamental solution.

We may now consider three classes of estimators suggested by the equalities

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\mathbb{E}^x[f(X_t)] = \mathbb{E}^x\left[f'(X_t)\,\frac{\mathrm{d}X_t}{\mathrm{d}\theta}\right] = \mathbb{E}^x[f(X_t)\,Y_t]\,. \tag{6}$$

The above equality explains that the three classes (see e.g., [Asmussen and Glynn, 2007, chap.VII]) are

1. a finite-difference approximation to the first expectation;
2. a derivative of f and the particle trajectories for the second expectation (also referred to as infinitesimal perturbation analysis or a pathwise derivative approach);
3. what amounts to a formal integration by parts for the third expectation (also referred to as likelihood ratio method or a Malliavin estimator or what we introduce as Stein equation estimators).

The third expectation depends upon the same measure as $\mathbb{E}^x[f(X_t)]$. Therefore the Monte Carlo scheme used to approximate the latter expectation can be reused. The practical impact of this mathematical statement is significant—the nontrivial understanding involved in Monte-Carlo approaches for approximating the expectation $\mathbb{E}^x[f(X_t)]$ can be applied to approximate the sensitivity. We caution the reader that in practise, the decision on which class of estimators to use depends upon the sensitivity of interest among several issues; see Sheppard et al. [2012] and also Section 4.

A standard approach in gradient-based optimization algorithms is to use an adjoint method. For example, if response sensitivities over a collection of initial conditions f are needed (where we continue to assume that $\Omega = \mathbb{R}^d$ for simplicity), then we can exploit the adjoint relation

$$R = \int_{t,x} q(x)\,u(t,x)\,dx\,dt = \int_{t,x} f(x)\,w(t,x)\,dx\,dt \tag{7}$$

where w is the solution to the adjoint PIDE

$$\begin{cases} \dfrac{\partial}{\partial t}w(t,x) + L^* w(t,x) = 0\,, & t > 0\,, x \in \mathbb{R}^d\,, \\ w(0,x) = q(x) & x \in \mathbb{R}^d\,. \end{cases} \tag{8}$$

The forward Kolmogorov operator L^* is the formal adjoint operator to L. The adjoint relation (7) is extremely useful within simulation-based optimization because adjoint (optimization) methods enable us to calculate numerous sensitivities given the solution of (8) in lieu of the solution of the PIDE (2a) for each initial f of interest.

This allows us to approximate the response function sensitivity R with a Monte Carlo scheme over the particle trajectories Z_t corresponding to the adjoint PIDE (8) so that

$$\frac{\mathrm{d}}{\mathrm{d}\theta}R = \int_{t,x} \frac{\mathrm{d}f(x)}{\mathrm{d}\theta}\,\mathbb{E}^x[q(Z_t)]\,dx\,dt + \int_{t,x} f(x)\,\frac{\mathrm{d}}{\mathrm{d}\theta}\mathbb{E}^x[q(Z_t)]\,dx\,dt \tag{9}$$

since

$$w(t,x) = \mathbb{E}^x[q(Z_t)] = \mathbb{E}[q(Z_t)|Z_0 = x]\,, \qquad t > 0\,, x \in \mathbb{R}^d\,. \tag{10}$$

The density $w(t,x)$ is then seen to be the average of $q(Z_t)$ over all particles that start at x at time $t = 0$. We now exploit the equalities (6) to approximate w within a Monte Carlo method and hence (9).

3. STEIN EQUATION ESTIMATORS FOR PARAMETER SENSITIVITY

In order to better understand the estimators that appear in (6), it is worth exploring some concrete examples. Through the next few sections, we will derive two estimators for the sensitivity of a scalar random variable X whose density $p(x;\theta)$ depends on some parameter $\theta \in \mathbb{R}$. It is natural to compute the mathematical expectation $\mathbb{E}_\theta[f(X)]$ by Monte Carlo methods when the integral $\int_{\mathbb{R}} f(x)p(x;\theta)\mathrm{d}x$ cannot be expressed explicitly. The purpose of this note is to explore multiple perspectives on how Monte Carlo methods can also be used to simultaneously estimate the sensitivity of mathematical expectation with respect to the parameter θ, i.e.

$$J(\theta\,; f) := \frac{\mathrm{d}}{\mathrm{d}\theta}\mathbb{E}_\theta[f(X)]\,.$$

In particular, the conceit of this exploration is that there is some random variable U such that (i) U does not depend on θ, (ii) U is "easy" to simulate, and (iii) X can be expressed in terms of U through some explicit monotone (for fixed θ) function $\phi_\theta : \mathbb{R} \to \mathbb{R}$, i.e. $X = \phi_\theta(U) = \phi(U\,;\theta)$. We will show how samples of U can be used to estimate the θ-sensitivity of X, comparing the efficiency of multiple methods.

In the development of this note, we were motivated by a particular inhomogeneous (two-rate) absorption problem. Suppose that X represents the location at which a particle is absorbed in a medium when the initial location is $x = 0$, the particle is moving to the right with a fixed velocity of 1. When the particle location is between 0 and some location ℓ, the absorption rate is $\lambda_1 > 0$ and when the particle has moved beyond ℓ the absorption rate is $\lambda_2 > 0$. Our goal will be to assess the sensitivity of expected absorption with respect to perturbations in the rate-change location ℓ. In particular, we wish to compute this sensitivity based on a Monte Carlo method that relies only on producing a sequence of unit rate exponential random variables $U_1, U_2, \ldots U_n$.

In Section 4 we work through the essential concepts by investigating the rate parameter sensitivity of a one-rate absorption problem ($X \sim \mathrm{Exp}(\lambda)$). We encounter four different types of estimators: $\widehat{J_1}$ a naïve difference estimator that establishes a baseline level of effectiveness; $\widehat{J_2}$ a coupled difference estimator that has been widely put into practice for continuous-time Markov chains; and $\widehat{J_3}$ and $\widehat{J_3}$ are two sides of an integration-by-parts formula that comes from two difference representations of the dependence on θ. In method for $\widehat{J_3}$ we encode the dependence on θ in the density of X, but in method for $\widehat{J_4}$ we encode the dependence in the mapping ϕ described above.

Once we see the general strategy for estimators $\widehat{J_3}$ and $\widehat{J_4}$, in Section 4 we will generalize the technique in Section 5 to a broader class of random variables where the dependence on θ can be encoded through a map ϕ_θ. Finally in Section 6 we address the two-rate absorption problem.

The core of this analysis is in the integration-by-parts formula that arises from encoding the dependence on θ in two different ways. This formula is an example of what is called a *Stein equation* in the probability and statistics literature. The existence of the integration-by-parts was first discovered by Stein while studying the mean of a multivariate normal distribution [Stein, 1973, 1981]. The idea was extended to the study of the Poisson distribution by Chen in collaboration with Stein a few years later [Peng, 1975] and the fact that this could be generalized to a wider class of random variables was observed soon thereafter by Hudson [1978].

While this work was developed from a statistics point of view, the probability community picked up on Stein's method and recognized it as a as a tool for measuring the distance between probability measures (see Meckes [2009] for a discussion of the Gaussian case). Since Stein equations can be generated in several different ways for a wide variety of random variables [Ley et al., 2017], they have been put to use in deriving upper and lower bounds on convergence rates of sequences of random variables that converge to final distributions that might be different than the standard Gaussian. One example of convergence to the exponential distribution was derived by Chatterjee et al. [2011] in a work on random graphs. The particular version of Stein equation that we derive and work with here has been called a *Parametric Stein Equation* by [Ley and Swan, 2016].

Though this literature on the Stein's method provides a deep context for the work we present here, the methods we employed borrow more heavily from an approach developed by Warren and Allen [2012b]. Their method, termed *trajectory reweighting* was in fact an instance of what are called *Malliavin estimators* in the stochastic processes community [Warren and Allen, 2014]. The estimator arises from Malliavin differentiation, which relies on a change-of-measure formula in path space. The connection between this work and what is presented here is that a change-of-measure formula is the vehicle through which our integration-by-parts is derived. This establishment of this connection between Stein's method and the Malliavin Calculus is attributed to Nourdin and Peccati [2009a] (see also [Nourdin and Peccati, 2009b], [Nourdin et al., 2010] and [Nourdin and Peccati, 2012]) and it has been an active field of research over the last decade. We expect that as we move to the numerical computation of sensitivities for jump processes, this connection between Stein's method and the Malliavin calculus will be essential.

3.1. Encoding dependence on parameters through a mapping

We take a moment to introduce some notation that will be useful in addressing method (4) for computing sensitivities, listed above. For one-dimensional random variables, it is, in principle, always possible to write a random variable in terms of a parameterless baseline distribution. In fact, this perspective is commonly used in simulation schemes for random variables that are defined in terms of arbitrary cumulative distribution functions (cdf). For a given value of the parameter θ, we denote the cdf by

$$F_\theta(x) = F(x;\theta) := \mathbb{P}_\theta[X \leqslant x].$$

Since a cdf is non-decreasing, it has a *left inverse* in the following sense:

$$F_\theta^{\leftarrow}(u) := \sup\{x \in \mathbb{R} : F_\theta(x) \leqslant u\}.$$

If we define $U \sim \text{Unif}([0,1])$, then the random variable $F_\theta^{\leftarrow}(U)$ has the cdf $F_\theta(x)$.

For the exponential distribution, for a given rate value $\lambda > 0$, let $X \sim \text{Exp}(\lambda)$ and $U \sim \text{Exp}(1)$. We can use the function $\phi_\lambda(u) = u/\lambda$ so that X is equal in distribution to the random variable U/λ. To see this, note that $\mathbb{P}[U > u] = e^{-u}$ and $\mathbb{P}[X > x] = e^{-\lambda x}$. It follows that $\mathbb{P}[U/\lambda > x] = \mathbb{P}[U > \lambda x] = e^{-\lambda x}$ as well.

We can define the inhomogeneous (two-rate) absorption problem similarly. In this case, we can again let $U \sim \text{Exp}(1)$ and then define

$$\phi_\Lambda(u) := \begin{cases} \dfrac{u}{\lambda_1}, & 0 \leqslant u \leqslant \lambda_1 \ell; \\ \dfrac{u - \lambda_1 \ell}{\lambda_2} + \ell, & u > \lambda_1 \ell. \end{cases} \tag{11}$$

where $\Lambda = (\lambda_1, \lambda_2, \ell)$. The inverse function will also play a large role. For each parameter set Λ, ϕ_Λ is invertible and has the form

$$\psi_\Lambda(x) = \begin{cases} \lambda_1 x, & 0 \leqslant x \leqslant \ell; \\ \lambda_2 x + (\lambda_1 - \lambda_2)\ell, & x > \ell. \end{cases} \tag{12}$$

Note that it immediately follows that the derivative with respect to x is

$$\psi'_\Lambda(x) = \begin{cases} \lambda_1, & 0 \leqslant x \leqslant \ell; \\ \lambda_2, & x > \ell. \end{cases} \tag{13}$$

To see that mapping yields the correct distribution, we first recall the derivation of the survival function $G_\Lambda(x) := \mathbb{P}_\Lambda[X > x]$. We can think of the state of the particle as a two-state Markov chain $\{S(x)\}_{x \in \mathbb{R}_+}$ that takes its values in $\{0,1\}$, which represent the not-absorbed and absorbed states respectively. Using the standard transition rate method for continuous-time Markov chains, if the initial condition is $S(0) = 0$, then the probability of no transition satisfies the ordinary differential equation

$$G'_\Lambda(x) = -\lambda(x) G_\Lambda(x), \qquad G_\Lambda(0) = 1,$$

where

$$\lambda(x) := \begin{cases} \lambda_1, & 0 \leqslant x \leqslant \ell; \\ \lambda_2, & x > \ell. \end{cases}$$

The solution is

$$G_\Lambda(x) = \exp\left(-\int_0^x \lambda(\xi) \mathrm{d}\xi\right) = \begin{cases} \exp\left(-\lambda_1 x\right), & 0 \leqslant x \leqslant \ell; \\ \exp\Big(-\big(\lambda_1 \ell + \lambda_2 (x - \ell)\big)\Big), & x > \ell. \end{cases}$$

Comparing to the inverse mapping function, we see the relation

$$\mathbb{P}_\Lambda[X > x] = G_\Lambda(x) = e^{-\psi_\Lambda(x)}.$$

This form allows us to immediately confirm that $\phi_\Lambda(U)$ has the right distribution. Since $\mathbb{P}[U > u] = e^{-u}$, we have

$$\mathbb{P}[\phi_\Lambda(U) > x] = \mathbb{P}[U > \psi_\Lambda(x)] = e^{-\psi_\Lambda(x)},$$

as expected.

In general, we will make the following assumptions about the functions ϕ and ψ.

Assumption 1. *Let the parameter space Θ be an open subset of $\mathbb{R}$. Let $\mathscr{O} \subset \mathbb{R}$ be the support of a probability density function $p : \mathscr{O} \to \mathbb{R}$ that does not depend on θ. Let $\phi : \mathscr{O} \times \Theta \to \mathbb{R}$ be an increasing, continuous function for each fixed $\theta \in \Theta$. Moreover we assume that the image $\Omega = \phi(\mathscr{O}; \theta)$ does not depend on the choice of $\theta \in \Theta$. Then we define the function $\psi : \Omega \times \Theta \to \mathbb{R}$ such that for each $\theta \in \Theta$, $\psi_\theta = \phi_\theta^{-1}$. For almost every $x \in \Omega$, we assume that $\psi(x; \cdot)$ and $\psi'(x; \cdot)$ are differentiable in θ.*

4. A MOTIVATING EXAMPLE: EXPONENTIAL RANDOM VARIABLES

Suppose that $X \sim \text{Exp}(\lambda)$ for some $\lambda > 0$. Then there are two eqivalent ways to express the dependence of the expected value of functions of X upon λ. The first, and most obvious, is to write the expected value in terms of the density function for X. Let

$$p(x; \lambda) = \lambda e^{-\lambda x}. \tag{14}$$

Note that $p(x; 1)$ is the density of a unit exponential random variable. When we wish to emphasize this point, we may write this density as simply $p(x)$.

$$\text{Dependence through density:} \quad \mathbb{E}_\lambda[f(X)] = \int_0^\infty f(x)p(x;\lambda)\mathrm{d}x. \tag{15}$$

As shown in the previous section, we can also encode the dependence on the rate parameter in writing X as a function of U,

$$X = \phi(U; \lambda) := U/\lambda. \tag{16}$$

Now, if we think of the expectation being taken with respect to U, we have

$$\begin{aligned} \text{Dependence through RV:} \quad \mathbb{E}_\lambda[f(X)] &= \mathbb{E}[f(\phi(U;\lambda))] \\ &= \int_0^\infty f(\phi(u;\lambda))p(u)\mathrm{d}u \end{aligned} \tag{17}$$

Assuming that f is sufficiently well-behaved so that we can differentiate under the integral, we have two formulas for the sensitivity that appear to be different. Differentiating (15) and then

rewriting the equation to that it is in the form of expectations, we get

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\lambda}\mathbb{E}_\lambda[f(X)] &= \frac{\mathrm{d}}{\mathrm{d}\lambda}\int_0^\infty f(x)\lambda e^{-\lambda x}\mathrm{d}x \\ &= \int_0^\infty f(x)(e^{-\lambda x} - \lambda x e^{-\lambda x})\mathrm{d}x \\ &= \int_0^\infty f(x)\left(\frac{1}{\lambda} - x\right)\lambda e^{-\lambda x}\mathrm{d}x \\ &= \mathbb{E}_\theta\left[f(X)\left(\frac{1}{\lambda} - X\right)\right]\end{aligned}$$

Note that this can be expressed in the following form

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\mathbb{E}_\lambda[f(X)] = -\mathbb{E}_\lambda[f(X)\,(X - \mathbb{E}_\lambda[X])]\,. \tag{18}$$

If we wish to express this in terms of a unit exponential U, we substitute $X = U/\lambda$ and attain

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\mathbb{E}_\lambda[f(X)] = -\mathbb{E}\left[f\left(\frac{U}{\lambda}\right)\left(\frac{U}{\lambda} - \mathbb{E}\left[\frac{U}{\lambda}\right]\right)\right]. \tag{19}$$

On the other hand, if we differentiate (17) we have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\lambda}\mathbb{E}_\lambda[f(X)] &= \frac{\mathrm{d}}{\mathrm{d}\lambda}\int_0^\infty f(\phi(u;\lambda))p(u)\mathrm{d}u \\ &= \int_0^\infty f'(\phi(u;\lambda))\frac{\partial}{\partial\lambda}\phi(u;\lambda))p(u)\mathrm{d}u \\ &= \int_0^\infty f'\left(\frac{u}{\lambda}\right)\frac{-u}{\lambda^2}p(u)\mathrm{d}u.\end{aligned}$$

So,

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\mathbb{E}_\lambda[f(X)] = -\mathbb{E}\left[f'\left(\frac{U}{\lambda}\right)\frac{U}{\lambda^2}\right] = -\mathbb{E}_\lambda\left[f'(X)\frac{X}{\lambda}\right]. \tag{20}$$

Equating (19) and (20) very much looks like an integration-by-parts formula:

$$\mathbb{E}\left[f'\left(\frac{U}{\lambda}\right)\frac{U}{\lambda^2}\right] = \mathbb{E}\left[f\left(\frac{U}{\lambda}\right)\left(\frac{U}{\lambda} - \mathbb{E}\left[\frac{U}{\lambda}\right]\right)\right]. \tag{21}$$

Note that we can also express this purely in terms of X:

$$\mathbb{E}_\lambda\left[f'(X)\frac{X}{\lambda}\right] = \mathbb{E}_\lambda\left[f(X)\big(X - \mathbb{E}_\lambda[X]\big)\right]. \tag{22}$$

This is the integration-by-parts formula that Ley and Swan refer to as a *parametric Stein equation* Ley and Swan [2016]. In Section 5, we will derive a generalization of this formula that can be used to generate estimators for the sensitivity to θ in a broad class of random variables.

4.1. Monte Carlo estimation of the exponential sensitivities

In this section we demonstrate the performance of four types of estimators for three different functions f. We choose

$$f^{(1)}(x) = x^2, \quad f^{(2)}(x) = \sin(x), \quad f^{(3)}(x) = 1_{[3,\infty)}(x). \tag{23}$$

The last of this is the indicator function on the set $[0.3, 0.5]$, recalling that the mathematical expectation of an indicator function is in fact a probability: $\mathbb{E}_\theta[1_A(X)] = \mathbb{P}_\theta[X \in A]$.

In each of these cases, the sensitivity can be computed explicitly. For reference, we have

$$\mathbb{E}_\lambda\left[X^2\right] = \frac{2}{\lambda^2}; \quad \mathbb{E}_\lambda[\sin(X)] = \frac{\lambda}{1+\lambda^2}; \text{ and } \quad \mathbb{E}_\lambda\left[1_{[a,b]}(X)\right] = e^{-\lambda a} - e^{-\lambda b}.$$

It follows that

$$J(\lambda\,; x^2) = -\frac{4}{\lambda^3}; \quad J(\lambda\,; \sin(x)) = \frac{1-\lambda^2}{(1+\lambda^2)^2}; \quad J(\lambda\,; 1_{[a,b]}(x)) = be^{-\lambda b} - ae^{-\lambda a}.$$

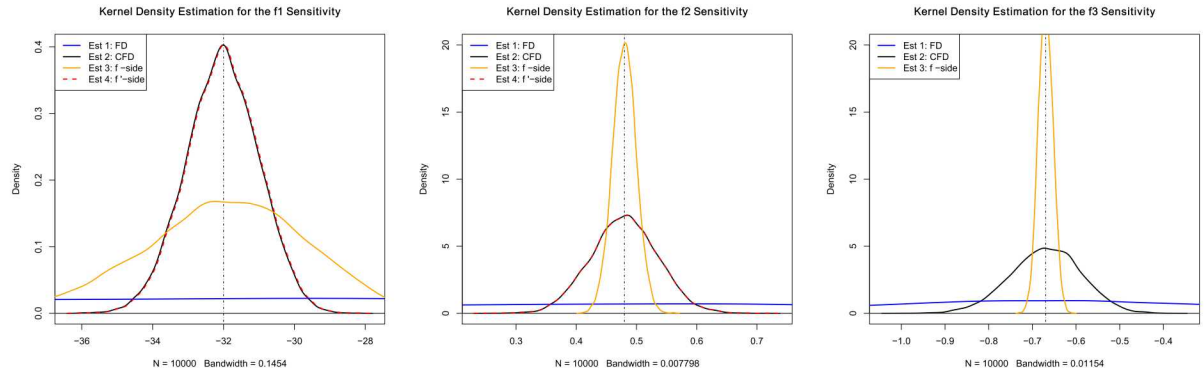

Figure 4-1 The results of a numerical computation for the sensitivity of an Exp(λ) random variable when the rate parameter is $\lambda = 1/2$. From left to right, we show a kernel denisty estimation for the distribution of $J_f(\lambda)$ for the three functions given in (23). Note that Estimators 2 and 4 are almost identical in the left and middle panels. This is addressed in Section 5. Note also that the best performing estimator depends on the function being used.

The estimators we use for these values are as follows. The first is the naïve finite difference estimator where we take N samples at two different values of λ and then estimate the derivative appropriately. To be precise, let $\varepsilon > 0$ be some small positive real number. Let $\{Y_i^+\}_{n=1}^N \sim \text{Exp}(\lambda+\varepsilon)$ and $\{Y_i^-\}_{n=1}^N \sim \text{Exp}(\lambda-\varepsilon)$ be two sets of independent random variables. Then the *Finite Difference (FD) esimator* is defined

$$\widehat{J_1}(\lambda\,; f) := \frac{1}{2\varepsilon}\left(\frac{1}{N}\sum_{n=1}^{N} f(Y_n^+) - f(Y_n^-)\right). \tag{24}$$

A much more efficient finite difference estimator is the *coupled finite difference* (CFD) estimator. The idea is to use "the same noise" to generate the $\lambda+\varepsilon$ and $\lambda-\varepsilon$ random variables. What that means for us, is that for each pair we generate a single unit-rate exponential and then use the mapping described in the previous section. Therefore, if $\{U_n\}_{n=1}^N \sim \text{Exp}(1)$, then the Coupled Finite Difference (CFD) estimator is defined to be

$$\widehat{J}_2(\lambda\,;f) := \frac{1}{2\varepsilon}\left(\frac{1}{N}\sum_{n=1}^{N} f\Big(\frac{U_n}{\lambda+\varepsilon}\Big) - f\Big(\frac{U_n}{\lambda-\varepsilon}\Big)\right). \tag{25}$$

We now define the two Parametric Stein Equation estimators are seen on both sides of the integration-by-parts formula (21). The right-hand side of (21) will serve as the basis of what we call the *f-side estimator*, $\widehat{J}_3$. Noting the negative sign in (19) we have

$$\widehat{J}_3(\lambda\,;f) := -\frac{1}{N}\sum_{n=1}^{N} f\left(\frac{U_n}{\lambda}\right)\left(\frac{U_n}{\lambda}-\frac{1}{\lambda}\right). \tag{26}$$

Meanwhile, the other side of (21) yields the *f'-estimator*:

$$\widehat{J}_4(\lambda\,;f) := -\frac{1}{N}\sum_{n=1}^{N} f\left(\frac{U_n}{\lambda}\right)\frac{U_n}{\lambda^2}. \tag{27}$$

In Figure 4-1, we display a kernel density estimation for the distribution of each of the four estimators. These simulations are the result of running 1000 independent trials, each producing an estimate for the true sensitivity based on 5000 independent samples. In all cases, the FD estimator (blue) has a much wider distribution compared to the others. This shows the efficacy of the CFD estimator (black), and why coupling (or correlated sampling) is so important for numerical computation. For the functions x^2 (left panel) and $\sin(x)$ (middle panel), the CFD and f'-estimators (dashed, red) are almost identical. We are able to explain this similarity in the next section. Note that for the indicator test function (right panel), which is equivalent to computing a probability of being in an interval, the f'-estimator is not well-defined. (This is natural because the indicator function is not differentiable at the points of discontuity.) The f-estimator (orange) *is* well-defined though, and considerably outperforms the other estimators for the indicator test function. Interestingly, it performs worse than the CFD and f' estimators for the other functions. This shows the power of the integration by parts formula. Depending on the test function, either the function itself or its derivative will be better behaved and yield a tighter estimation for the quanitity of interest. These results are similar to what we will see for the two-rate problem.

5. SENSITIVITY FOR RVS DEFINED THROUGH CONTINUOUSLY DIFFERENTIABLE MAPS

Generically, if X is a random variable with density $p(x;\theta)$ that depends on the parameter θ, then the sensitivity with respect to θ can be derived as follows:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\theta}\mathbb{E}_\theta[f(X)] &= \frac{\mathrm{d}}{\mathrm{d}\theta}\int_\Omega f(x)p(x;\theta)\mathrm{d}x \\ &= \int_\Omega f(x)\frac{\partial p}{\partial\theta}(x;\theta)\mathrm{d}x.\end{aligned}$$

The general technique for writing sensitivities in terms of mathematical expectation is to multiply and divide by the factor that is necessary to replace the density. In this case, this results in appearance of a logarithmic derivative of the density function:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\theta}\mathbb{E}_\theta[f(X)] &= \int_\Omega f(x)\frac{\frac{\partial p}{\partial\theta}(x;\theta)}{p(x;\theta)}p(x;\theta)\mathrm{d}x \\ &= \int_\Omega f(x)\frac{\partial}{\partial\theta}\ln(p(x;\theta))p(x;\theta)\mathrm{d}x \\ &= \mathbb{E}_\theta\left[f(X)\frac{\partial}{\partial\theta}\ln(p(X;\theta))\right].\end{aligned} \tag{28}$$

For the exponential random variable, we have $\frac{\partial}{\partial\lambda}\ln p(x;\lambda) = \frac{\partial}{\partial\lambda}(\ln\lambda - \lambda x) = \frac{1}{\lambda} - x$. Since $\mathbb{E}_\lambda[X] = 1/\lambda$, this matches right-hand side of (22).

The goal of this note, of course, is to develop a Monte Carlo method for estimating this sensitivity in terms of samples from some baseline random variable U, whose density $p(u)$ does not depend on θ. Just as for the exponential case, there are two natural (and equivalent) ways to represent the dependence of X on θ through the mapping ϕ_θ and its inverse ψ_θ. The connection between the two is captured by the standard change-of-measure formula

$$p_X(x;\theta) = p(\psi_\theta(x))\psi'_\theta(x). \tag{29}$$

In terms of mathematical expectation, this can be expressed:

$$\begin{aligned}\mathbb{E}_\theta[f(X)] &= \int_\Omega f(x)p(\psi_\theta(x))\psi'_\theta(x)\mathrm{d}x \\ &= \int_{\psi(\Omega)} f(\phi_\theta(u))p(u)\mathrm{d}u.\end{aligned} \tag{30}$$

When we differentiate these two forms of the expectation with respect to θ, we see that the latter will result in a derivative on f due to the chain rule, while the other does not. So, the change-of-measure formula for the density become an integration by parts formula for the sensitivities. This is the content of the following theorem, and as we saw in the previous section, the two sides of the integration-by-parts excel for different target functions f.

Theorem 5.1. *Let U be a base random variable with density $p(u)$ that does not depend upon the parameter θ and suppose that $X = \phi_\theta(U)$, where ϕ and ψ satisfy Assumption 1. Let*

$$\rho(u) := \frac{d}{du} \ln\big(p(u)\big). \tag{31}$$

Then

$$\begin{aligned} \frac{\partial}{\partial\theta}\mathbb{E}_\theta[f(X)] &= \mathbb{E}_\theta\left[f'(X)\,\frac{\partial\phi_\theta}{\partial\theta}\big(\psi_\theta(X)\big)\right] \\ &= \mathbb{E}_\theta\left[f(X)\left(\rho(\psi_\theta(X))\frac{\partial\psi_\theta}{\partial\theta}(X) + \frac{\partial}{\partial\theta}\ln(\psi'_\theta(X))\right)\right] \end{aligned} \tag{32}$$

Proof. For the first equality, we encode the dependence on θ through the random variable. To emphasize that ϕ can be seen as a function of both u and θ, we employ the suggestive notation $\phi(u;\theta) := \phi_\theta(u)$. Bringing the differentiation under the integral sign and applying the chain rule reveals

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}\theta}\mathbb{E}_\theta[f(X)] = \frac{\mathrm{d}}{\mathrm{d}\theta}\mathbb{E}[f(\phi(U;\theta))] &= \frac{\mathrm{d}}{\mathrm{d}\theta}\int_\Omega f(\phi(u;\theta))p(u)\mathrm{d}u \\ &= \int_\Omega f'(\phi(u;\theta))\frac{\partial\phi}{\partial\theta}\big(u;\theta\big)p(u)\mathrm{d}u \\ &= \mathbb{E}\left[f'\big(\phi_\theta(U)\big)\frac{\partial\phi_\theta}{\partial\theta}(U)\right]. \end{aligned} \tag{33}$$

Applying the inverse to express this in terms of X yields the middle expression in (32).

For the second equality, we write the density function of X in terms of ψ_θ and the density of U through the change-of-measure formula (29). We then pass the derivative under the integral, minding the product rule:

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}\theta}\mathbb{E}_\theta[f(X)] &= \frac{\mathrm{d}}{\mathrm{d}\theta}\int_\Omega f(x)p(\psi_\theta(x))\psi'_\theta(x)\mathrm{d}x \\ &= \int_\Omega f(x)p'(\psi_\theta(x))\frac{\partial\psi_\theta}{\partial\theta}(x)\psi'_\theta(x)\mathrm{d}x \\ &\qquad + \int_\Omega f(x)p(\psi_\theta(x))\frac{\partial\psi'_\theta}{\partial\theta}(x)\mathrm{d}x.. \end{aligned}$$

Similar the what was done in the general case in Equations 28, we multiply and divide each integrand by the factors that are necessary to recover the density for X:

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}\theta}\mathbb{E}_\theta[f(X)] &= \int_\Omega f(x)\frac{p'(\psi_\theta(x))}{p(\psi_\theta(x))}\frac{\partial\psi_\theta}{\partial\theta}(x)p_X(x)\mathrm{d}x \\ &\qquad + \int_\Omega f(x)p(\psi_\theta(x))\frac{\frac{\partial\psi'_\theta}{\partial\theta}(x)}{\psi'_\theta(x)}p_X(x)dx. \end{aligned}$$

Recognizing these ratios are logarithmic derivatives and writing in terms of mathematical expectation with respect to X, we have the third expression in (32). □

For the exponential random variable, recall that $\phi_\lambda(u) = u/\lambda$ and $\psi_\lambda(x) = \lambda x$. The quantities we need are

$$\frac{\mathrm{d}}{\mathrm{d}u}\ln(p(u)) = -1, \quad \frac{\partial \psi_\lambda}{\partial \lambda}(x) = x, \text{ and } \quad \frac{\partial}{\partial \lambda}\psi'_\lambda(x) = \frac{1}{\lambda}.$$

Substuting these quantities into the third expression in (32) yields (20) as expected. Meanwhile,

$$\frac{\partial \phi_\lambda}{\partial \lambda}(u) = -\frac{u}{\lambda^2}, \text{ so } \frac{\partial \phi_\lambda}{\partial \lambda}\big(\psi(x)\big) = -\frac{x}{\lambda}$$

which recovers the left-hand side of (22).

It is now easier to see why estimators 2 and 4 were so similar in the exponential random variable experiments displayed in Figure 4-1. The coupled finite difference estimator is, in fact, a discretization of the "f-prime" estimator. Indeed, using the expansion

$$\begin{aligned} f(\phi(u;\theta \pm \varepsilon)) &= f\left(\phi(u;\theta) \pm \frac{\partial}{\partial \theta}\phi(u;\theta)\varepsilon + O(\varepsilon^2)\right) \\ &= f(\phi(u;\theta)) \pm f'(\phi(u;\theta))\frac{\partial}{\partial \theta}\phi(u;\theta)\varepsilon + O(\varepsilon^2), \end{aligned}$$

we find that

$$\frac{1}{2\varepsilon}\big(\mathbb{E}[f(\phi(U;\theta+\varepsilon)) - f(\phi(U;\theta-\varepsilon))]\big) = \mathbb{E}\left[f'(\phi(U))\frac{\partial \phi_\theta}{\partial \theta}(U)\right] + O(\varepsilon). \tag{34}$$

The left-hand side is coupled finite difference estimator and the mathematical expectation on the right-hand side is the estimator derived in Theorem 5.1.

6. EXTENSION TO THE TWO-RATE ABSORPTION PROBLEM

We return to two-rate absorption problem. Here, the mapping function ϕ_Λ is defined piecewise with the split appear at a location $X = \ell$. Importantly, the two functions that comprise ϕ_Λ can depend on ℓ as well. In order to understand the result we will eventually derive, we first present a version of Theorem 5.1 and that addresses the sensitivity of piecewise defined random variables with respect to changes in the location of the change. The terms involving mathematical expectations will look the same, but there are boundary terms that emerge.

Theorem 6.1. *Suppose that $X = \phi_\Theta(U)$, where $\Theta = (\theta_1, \theta_2, \ell)$ and ϕ is a piecewise differentiable (expressed in terms of everywhere-differentiable functions $\phi_1(u;\theta_1,\ell)$ and $\phi_2(u;\theta_2,\ell)$), continuous and non-decreasing function. Let $\psi_1(\cdot;\theta_1,\ell)$ and $\psi_2(\cdot;\theta_2,\ell)$ be the inverse functions of $\phi_1(\cdot,\theta_1,\ell)$ and $\phi_2(\cdot;\theta_2,\ell)$ for each choice of parameter vector Θ. And suppose that ϕ is differentiable in θ_1, θ_2, and ℓ everywhere except for the set $\{x = \ell\}$. So,*

$$\phi(u;\Theta) := \begin{cases} \phi_1(u;\theta_1,\ell), & u \leqslant \psi(\ell) \\ \phi_2(u;\theta_2,\ell), & u > \psi(\ell) \end{cases} \tag{35}$$

If $p(u)$ is the density of U, and we recall the notation $\rho(u) := \frac{d}{du}\ln(p(u))$, we have

$$\begin{aligned}
\frac{d}{d\ell}\mathbb{E}_\Theta[f(X)] &= f(\ell)p(\psi(\ell))\left(\psi_1'(\ell;\theta_1,\ell)+\frac{\partial\psi_1}{\partial\ell}(\ell;\theta_1,\ell)\right.\\
&\qquad\left.-\psi_2'(\ell;\theta_2,\ell)-\frac{\partial\psi_2}{\partial\ell}(\ell;\theta_2,\ell)\right)\\
&\quad+\mathbb{E}_\Theta\left[f'(X)\frac{\partial\phi_1}{\partial\ell}\big(\psi_1(X)\big)1_{\{X\leqslant\ell\}}\right]\\
&\quad+\mathbb{E}_\Theta\left[f'(X)\frac{\partial\phi_2}{\partial\ell}\big(\psi_2(X)\big)1_{\{X>\ell\}}\right]\\
&= f(\ell)p(\psi(\ell))\big(\psi_1'(\ell;\theta_1,\ell)-\psi_2'(\ell;\theta_2,\ell)\big)\\
&\quad+\mathbb{E}_\Theta\left[f(X)\left(\rho\big(\psi_1(X)\big)\frac{\partial\psi_1}{\partial\ell}(X)+\frac{\partial}{\partial\ell}\ln\big(\psi_1'(X)\big)\right)1_{\{X\leqslant\ell\}}\right]\\
&\quad+\mathbb{E}_\Theta\left[f(X)\left(\rho\big(\psi_2(X)\big)\frac{\partial\psi_2}{\partial\ell}(X)+\frac{\partial}{\partial\ell}\ln\big(\psi_2'(X)\big)\right)1_{\{X>\ell\}}\right]
\end{aligned}\tag{36}$$

(Recall that, when there is no risk of ambiguity, the notation $'$ refers to differentiation with respect to x.)

Proof. The proof proceeds similarly to Theorem 5.1, beginning by encoding the dependence on ℓ in the random variable. We need to take special care though, when looking at the limits of integration. The breakpoint can be expressed as the inverse of the location $x=\ell$, i.e. $\psi(\ell)$, since ψ is continuous. But since it is not required to be (and indeed in our two-rate absorption problem, it will not be) differentiable, it is worth expressing in terms of the piecewise-defined functions ψ_1 and ψ_2:

$$\begin{aligned}
\mathbb{E}_\Theta[f(X)] &= \int_{-\infty}^{\psi_1(\ell;\theta_1,\ell)} f(\phi_1(u;\theta_1,\ell)p(u)\mathrm{d}u\\
&\qquad+\int_{\psi_1(\ell;\theta_1,\ell)}^{\infty} f(\phi_2(u;\theta_2,\ell))p(u)\mathrm{d}u.
\end{aligned}$$

When applying the Leibniz rule for differentiation of this integral, we will encounter total derivatives of the type

$$\begin{aligned}
\frac{\partial}{\partial\ell}\psi_i(\ell;\theta_i,\ell) &= \frac{\partial\psi_i}{\partial x}(x;\theta_i,\ell)\big|_{x=\ell}+\frac{\partial\psi_i}{\partial\ell}(\ell;\theta_i,\ell)\\
&= \psi_i'(\ell;\theta_i,\ell)+\frac{\partial\psi_i}{\partial\ell}(\ell;\theta_i,\ell)
\end{aligned}$$

where $i\in\{1,2\}$ and$\partial/\partial\ell$ refers to differentiation in the third coordinate. We note that, for the last equality, this is only a change of notation.

We now calculate the sensitivity, differentiate the mathematical expectation with respect to ℓ,

minding the chain rule and total derivative that appears in each term:

$$\begin{aligned}\frac{\partial}{\partial \ell}\mathbb{E}_\Theta[f(X)] = f(\ell)p(\psi(\ell))\Bigg(&\psi_1'(\ell;\theta_1,\ell) + \frac{\partial \psi_1}{\partial \ell}(\ell;\theta_1,\ell) \\ &\qquad -\psi_2'(\ell;\theta_2,\ell) - \frac{\partial \psi_2}{\partial \ell}(\ell;\theta_2,\ell))\Bigg) \\ &+ \int_{-\infty}^{\psi(\ell)} f'(\phi_1(u;\theta_1,\ell))\frac{\partial \phi_1}{\partial \ell}(u;\theta_1,\ell)p(u)\mathrm{d}u \\ &- \int_{\psi(\ell)}^{\infty} f'(\phi_2(u;\theta_2,\ell))\frac{\partial \phi_2}{\partial \ell}(u;\theta_2,\ell)p(u)\mathrm{d}u\end{aligned}$$

Changing variables to $x = \phi(u)$ yields the f-prime estimator.

For the other side of the integration by parts, we express the integral in terms of x instead of in terms of u. We have

$$\begin{aligned}\mathbb{E}_\Theta[f(X)] = &\int_{-\infty}^{\ell} f(x)p\big(\psi_1(x;\theta_1,\ell)\big)\psi_1'(x;\theta_1,\ell)\mathrm{d}x \\ &+ \int_{\ell}^{\infty} f(x)p\big(\psi_2(x;\theta_2,\ell)\big)\psi_2'(x;\theta_2,\ell)\mathrm{d}x\end{aligned}$$

Applying the Leibniz rule is more straightforward this time:

$$\begin{aligned}\frac{\partial}{\partial \ell}\mathbb{E}_\Theta[f(X)] = &f(\ell)p(\psi(\ell))\left(\psi_1'(\ell;\theta_1,\ell) - \psi_2'(\ell;\theta_2,\ell)\right) \\ &+ \int_{-\infty}^{\ell} f(x)\frac{\partial}{\partial \ell}\Big(p\big(\psi_1(x;\theta_1,\ell)\big)\psi'(x;\theta_1,\ell)\Big)\mathrm{d}x \\ &+ \int_{\ell}^{\infty} f(x)\frac{\partial}{\partial \ell}\Big(p\big(\psi_2(x;\theta_2,\ell)\big)\psi'(x;\theta_2,\ell)\Big)\mathrm{d}x\end{aligned}$$

The remainder of the argument for both of these integrands proceeds exactly as it did in the proof of Theorem 5.1. We multiply and divide each term appropriate to recover the densities for X when it is less than or greater than ℓ, respectively.

□

Corollary 6.2. *Let $U \sim \mathrm{Exp}(1)$ and $X = \phi_\Lambda(U)$ where the parameter set $\Lambda = (\lambda_1, \lambda_2, \ell)$ and functions ϕ_Λ and ψ_Λ are as defined in Section 3.1. Then*

$$\begin{aligned}\frac{\partial}{\partial \ell}\mathbb{E}_\Lambda[f(X)] &= -e^{-\lambda_1 \ell}\left(1 - \frac{\lambda_1}{\lambda_2}\right)\mathbb{E}\left[f'\left(\frac{U}{\lambda_2} + \ell\right)\right] \\ &= \quad e^{-\lambda_1 \ell}(\lambda_1 - \lambda_2)\left(f(\ell) - \mathbb{E}\left[f\left(\frac{U}{\lambda_2} + \ell\right)\right]\right)\end{aligned} \tag{37}$$

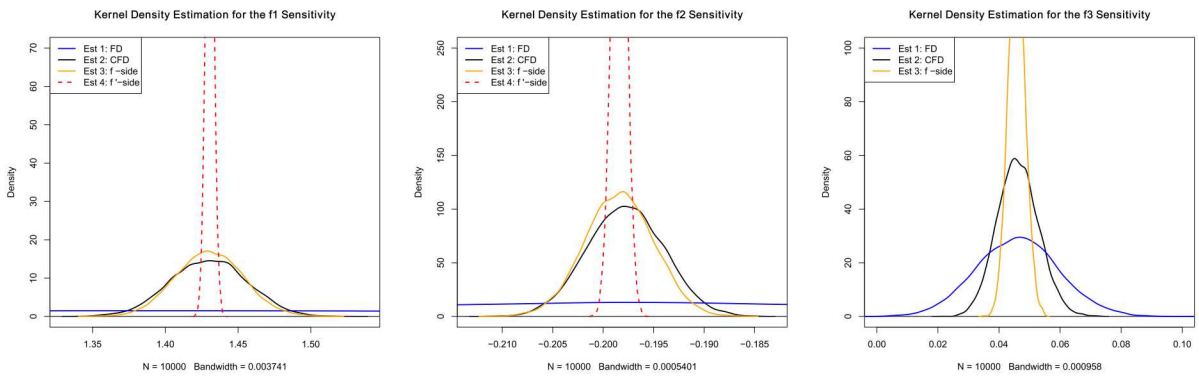

Figure 6-1 Sensitivity of a two-rate absorption problem when the rate parameters are $\lambda_1 = 1/3$, $\lambda_2 = 2$ and $\ell = 2$. From left to right, we show a kernel denisty estimation for the distribution of $J_f(2)$ for the three functions given in (23). Note that the CFD and f' estimators (2 and 4, respectively) are no longer almost identical like they were in the exponential case. This is due to the use of importance sampling, which improves the efficiency of the f-prime estimator. Note also that, again, the best performing estimator depends on the function being used.

Proof. For the two-rate absorption problem, note that we have the following partial derivatives:

$$\begin{aligned}
&\frac{\partial \phi_1}{\partial \ell}(u;\lambda_1,\ell) = 0, \quad \frac{\partial \phi_2}{\partial \ell}(u;\lambda_2,\ell) = 1 - \frac{\lambda_1}{\lambda_2};\\
&\frac{\partial \psi_1}{\partial \ell}(x;\lambda_1,\ell) = 0, \quad \frac{\partial \psi_2}{\partial \ell}(x;\lambda_2,\ell) = (\lambda_1 - \lambda_2);\\
&\frac{\partial \psi_1}{\partial x}(x;\lambda_1,\ell) = \lambda_1, \quad \frac{\partial \psi_2}{\partial x}(x;\lambda_2,\ell) = \lambda_2.
\end{aligned}$$

We look first at the f-prime estimator, and find that, remarkably, the boundary terms cancel. Moreover, the property $\frac{\partial \phi_1}{\partial \ell} = 0$ eliminates the mathematical expectation for X values less than ℓ. As a result, the sensitivity with respect to the location of the rate change is

$$\begin{aligned}
\frac{\partial}{\partial \ell}\mathbb{E}_\Lambda[f(X)] &= -\left(1 - \frac{\lambda_1}{\lambda_2}\right)\mathbb{E}_\Lambda\left[f'(X)1_{\{X \geqslant \ell\}}\right]\\
&= -\left(1 - \frac{\lambda_1}{\lambda_2}\right)\mathbb{E}\left[f'\left(\frac{U - \lambda_1\ell}{\lambda_2} + \ell\right)1_{\{U > \lambda_1\ell\}}\right]\Big).
\end{aligned}$$

The simulation effort associated with the right-most term is wasteful though, in the sense that all samples of U that are less than $\lambda_1\ell$ are simply discarded. With a simple substitution we can use *importance sampling* in order to focus only on values that matter. Indeed, letting $v = u - \lambda_1\ell$, we have

$$\begin{aligned}
\mathbb{E}\left[f'\left(\frac{U - \lambda_1\ell}{\lambda_2} + \ell\right)1_{\{U > \lambda_1\ell\}}\right] &= \int_{\lambda_1\ell}^{\infty} f'\left(\frac{u - \lambda_1\ell}{\lambda_2} + \ell\right)e^{-u}\mathrm{d}u\\
&= \int_0^{\infty} f'\left(\frac{v}{\lambda_2} + \ell\right)e^{-(v+\lambda_1\ell)}\mathrm{d}v.
\end{aligned}$$

Relabelling and factoring out $e^{-\lambda\ell}$, we have a final form

$$\frac{\mathrm{d}}{\mathrm{d}\ell}\mathbb{E}_\Lambda[f(X)] = -e^{-\lambda_1\ell}\Big(1-\frac{\lambda_1}{\lambda_2}\Big)\mathbb{E}\Big[f'\Big(\frac{U}{\lambda_2}+\ell\Big)\Big]. \tag{38}$$

Meanwhile, we can compute the other side of the integration-by-parts as follows. In terms of U, we have $\rho(u) = \frac{\mathrm{d}}{\mathrm{d}u}\ln p(u) = \frac{\mathrm{d}}{\mathrm{d}u}(-u) = -1$, and so

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\ell}\mathbb{E}_\Lambda[f(X)] &= f(\ell)e^{-\lambda_1\ell}(\lambda_1-\lambda_2) - \mathbb{E}_\Lambda\big[f(X)(\lambda_1-\lambda_2)1_{\{X>\ell\}}\big] \\ &= e^{-\lambda_1\ell}(\lambda_1-\lambda_2)\Big(f(\ell) - \mathbb{E}\Big[f\Big(\frac{U}{\lambda_2}+\ell\Big)\Big]\Big)\end{aligned}$$

□

The quantities in Corollary 6.2 yield natural estimators, and their performances are displayed in Figure 6-1.

7. LANGEVIN AND BROWNIAN DYNAMICS

In this section, we consider the task of computing the sensitivity of expectations (or averages) with respect to model parameters for Langevin and Brownian dynamics. We consider both static (or equilibrium) and dynamic (or non-equilibrium) averages. The results generally follow prior work by Rief [1994] on static averages and work by Warren and Allen [2012a], **?** on dynamic averages.

The Langevin equation for a system of N particles is a stochastic differential equation on a $6N$ dimensional phase space (position and velocity in 3 dimensional space). In the physics literature, the equation is compactly written as

$$m_i\dot{\vec{v}}_i = -\gamma_i\vec{v}_i + \vec{F}_{c,i}(\vec{x},t;\theta) + \sqrt{2\gamma_i k_B T}\vec{R}_i(t), \qquad \dot{\vec{x}}_i = \vec{v}_i, \tag{39}$$

where $\vec{x}_i, \vec{v}_i \in \mathbb{R}^3$ are the positions and velocities of the ith particle, $m_i \in \mathbb{R}$ is the mass, and $\vec{F}_{c,i}$ is the deterministic force. The friction parameter, γ_i, is assumed to be positive, and $k_B T$ is product of Boltzmann's constant and the temperature. The random force on the ith particle, $\vec{R}_i$, is a white noise process with zero mean, unit variance and uncorrelated in time,

$$\mathbb{E}\Big[\vec{R}_i(t)\Big] = 0, \qquad \mathrm{Cov}\Big[\vec{R}_i(t), \vec{R}_i^T(t')\Big] = \delta(t-t')\mathbb{I}. \tag{40}$$

Note that here we have assumed that the deterministic force is not a function of the velocity, which is a typical, but not a necessary simplification.

Given a function on phase space, $A : \mathbb{R}^{3N}\times\mathbb{R}^{3N} \to \mathbb{R}$, we can write the expectation of A at time t as

$$\mathbb{E}[A(Z(t;\theta))] = \int A(Z)p(Z,t;\theta)\mathrm{d}Z,$$

where $Z = [\vec{x}, \vec{v}]$ and p is the probability density of finding the system with configuration Z at time t for parameter θ. The function p can be estimated by generating an ensemble of trajectories, or by solving the associated Fokker–Planck equation. The sensitivity (or derivative) of the expectation with respect to θ is,

$$\begin{aligned}
\frac{d}{d\theta}\mathbb{E}\left[A(Z(t;\theta))\right] &= \int A(Z)\frac{d}{d\theta}p(Z,t;\theta)\mathrm{d}Z, \\
&= \int A(Z)\left(\frac{d}{d\theta}\ln p(Z,t;\theta)\right)p(Z,t;\theta)\mathrm{d}Z, \\
&= \mathbb{E}\left[A(Z(t;\theta))\frac{d}{d\theta}\ln p(Z,t;\theta)\right], \\
&= \mathbb{E}\left[A(Z(t;\theta))\omega(Z,t;\theta)\right].
\end{aligned}$$

where ω is a scalar weight. Hence, the sensitivity of the expectation can be viewed as an expectation with a non-constant weight.

As mentioned above, the function p can be estimated by generating an ensemble of trajectories. In the remainder of this report, we provide practical algorithms for estimating the sensitivity computing the weight, ω.

If we assume that the inertial component is small (which is a valid assumption over large time scales, for relatively small masses, or relatively large friction coefficients) the $m_i\dot{\vec{v}}_i$ term can be neglected, and we obtain the Brownian dynamics equations,

$$0 = -\gamma_i\vec{v}_i + \vec{F}_{c,i}(\vec{x},t;\theta) + \sqrt{2\gamma_i k_B T}\vec{R}_i(t), \qquad \dot{\vec{x}}_i = \vec{v}_i,$$

or

$$\gamma_i\dot{\vec{x}}_i = \vec{F}_{c,i}(\vec{x},t;\theta) + \sqrt{2\gamma_i k_B T}\vec{R}_i(t). \tag{41}$$

If the deterministic force is conservative, then $\vec{F}_{c,i} = -\frac{\partial}{\partial x_i}U$ for a conservative potential, U, one can show that the dynamics of Brownian and Langevin dynamics are ergodic, and that the stationary distribution is the Boltzmann distribution,

$$\rho(\vec{x};\theta) = C(\theta)\exp\left[-\frac{1}{k_B T}U(\vec{x};\theta)\right],$$

where $C(\theta)$ is a normalization constant.

7.1. Analytic results: 1D harmonic well

We start by considering the problem of a single walker in a one-dimensional harmonic well. In this case, the deterministic force is $F_c(x;\theta) = -U'(x;\theta) = -\theta(x-\mu)$ where $U(x;\theta) = \frac{1}{2}\theta(x-\mu)^2$. The harmonic well is centered at position μ, and the parameter θ is the "spring constant" which controls the expected range of motion of the walker.

For this problem, Brownian and Langevin dynamics are ergodic, and the stationary distribution is

$$\rho\left(\vec{x};\theta\right) = C(\theta)\exp\left[-\frac{1}{2}\frac{\theta\left(x-\mu\right)^2}{k_B T}\right], \tag{42}$$

with $C(\theta) = \sqrt{\theta/(2\pi k_B T)}$.

7.1.1. Brownian Dynamics

The Brownian dynamics equation for this example is

$$\gamma\dot{x} = -\theta\left(x-\mu\right) + \sqrt{2\gamma k_B T}R\left(t\right), \tag{43}$$

If we let $\alpha = \theta/\gamma$ and $\sigma = \sqrt{2k_B T/\gamma}$, then we can rewrite this as

$$dX_t = -\alpha\left(X_t-\mu\right)dt + \sigma dW_t, \tag{44}$$

which is the well–known Ornstein–Uhlenbeck equation. In this notation, W_t is a Wiener process, α is a positive dissipation parameter, and σ is a positive fluctuation parameter. This stochastic differential equation has an exact solution

$$X(t) = e^{-\alpha t}\left(X(0)-\mu\right) + \mu + \sigma\xi\sqrt{\frac{1-e^{-2\alpha t}}{2\alpha}}, \tag{45}$$

where ξ is drawn from a unit normal distribution. Note that in the literature, this exact solution is also known as an "autoregressive model of order 1" or an "AR(1) model". Since we know the distribution for ξ we use simple algebraic manipulation to derive the distribution for x for an ensemble of trajectories starting from the same initial condition, $x(0) = x_0$,

$$p(x,t;\theta) = \sqrt{\frac{\alpha}{\pi\sigma^2\left(1-e^{-2\alpha t}\right)}}\exp\left[-\alpha\frac{\left(x-e^{-\alpha t}(x_0-\mu)-\mu\right)^2}{\sigma^2\left(1-e^{-2\alpha t}\right)}\right]. \tag{46}$$

Note that this is solution of the associated Fokker–Planck equation

$$\frac{\partial p}{\partial t} = \alpha\frac{\partial}{\partial x}\left(\left(x-\mu\right)p\right) + \frac{\sigma^2}{2}\frac{\partial^2 p}{\partial x^2}, \tag{47}$$

with the appropriate initial and boundary conditions. Taking the limit as t goes to infinity results in the stationary limiting distribution in (42).

Now, suppose we want to compute the sensitivity of the expectation of A at time t for parameter θ. From the first section, we know that

$$\frac{d}{d\theta}\mathbb{E}\left[A(x(t;\theta))\right] = \mathbb{E}\left[A(x(t;\theta))\omega(x,t;\theta)\right]$$

where

$$\omega(x,t;\theta) = \frac{d}{d\theta}\ln p(x,t;\theta).$$

Inserting the distribution function p for this problem, we obtain

$$\begin{aligned}\omega(x,t;\theta) &= \frac{1}{2\theta} - t\frac{e^{-2\alpha t}}{\gamma(1-e^{-2\alpha t})} + 2t\alpha\frac{(x-e^{-\alpha t}(x_0-\mu)-\mu)^2 e^{-\alpha t}}{\gamma\sigma^2(1-e^{-2\alpha t})^2} \\ &\quad - \frac{(x-e^{-\alpha t}(x_0-\mu)-\mu)^2 + 2\alpha t\,(x-e^{-\alpha t}(x_0-\mu)-\mu)\,e^{-\alpha t}(x_0-\mu)}{\gamma\sigma^2(1-e^{-2\alpha t})}\end{aligned}$$

7.1.2. Langevin Dynamics

Langevin equation is

$$m\dot{v} = -\gamma v - \theta(x-\mu) + \sqrt{2\gamma k_B T}R(t), \qquad \dot{x} = v, \tag{48}$$

where the parameter θ is a "spring constant" that controls the width of the harmonic well.

If we let $Z = [x-\mu, v]$ then we can rewrite this as a linear first-order SDE,

$$\dot{Z} = \frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} v \\ x-\mu \end{bmatrix} = -\begin{bmatrix} \gamma/m & \theta/m \\ -1 & 0 \end{bmatrix}\begin{bmatrix} v \\ x-\mu \end{bmatrix} + \begin{bmatrix} \sqrt{2\theta k_B T} & 0 \\ 0 & 0 \end{bmatrix}\vec{R}(t), \tag{49}$$

or more compactly,

$$dZ_t = -AZ_t dt + S dW_t, \tag{50}$$

with

$$A := \begin{bmatrix} \gamma/m & \theta/m \\ -1 & 0 \end{bmatrix}, \qquad S := \begin{bmatrix} \sqrt{2\theta k_B T} & 0 \\ 0 & 0 \end{bmatrix}.$$

The solution for this system is

$$Z(t) = e^{-At}Z(0) + \int_0^t e^{-A(t-\tau)}S dW_\tau,$$

using the standard definition of the matrix exponential.

7.2. Numerical methods

The Euler–Maruyama approximation to the solution of the general Brownian dynamics equations in (41) is

$$\vec{x}_i^{\,n+1} - \vec{x}_i^{\,n} = \frac{\Delta t}{\gamma_i}\vec{F}_{c,i}(\vec{x}^{\,n}, t; \theta) + \sqrt{\frac{2k_B T\Delta t}{\gamma_i}}\xi_i^n,$$

where $\vec{x}_i^{\,n}$ is the position of the ith particle at time $t_n = n\Delta t$. Each ξ_i^n is a random number drawn from a standard normal distribution.

Given an ensemble of trajectories, we can estimate the expectation of A at time t_n with

$$\mathbb{E}\left[A(x(t_n;\theta))\right] \approx \frac{1}{N_{\mathrm{sim}}}\sum_{k=1}^{N_{\mathrm{sim}}} A(x_k^n(\theta)),$$

where x_k^n is the numerical solution at time t_n for the kth simulation. To estimate the sensitivity of this expectation, we note that the probability of a point on the trajectory is the product of conditional (or transition) probabilities,

$$Pr\left(x_k^n(\theta)\right) = Pr\left(x_k^n(\theta)|x_k^{n-1}(\theta)\right) Pr\left(x_k^{n-1}(\theta)|x_k^{n-2}(\theta)\right) \cdots Pr\left(x_k^0(\theta)\right).$$

Since the ξ_i^n are drawn from a standard normal distribution, we can compute the transition probability with simple algebraic manipulation,

$$Pr\left(X_{n+1}|X_n\right) = \frac{1}{\sqrt{(4\pi\gamma^{-1}\Delta t)^{3N}}} \exp\left[-\frac{\|X_{n+1} - X_n - \gamma^{-1}F_c(X_n;\theta)\Delta t\|^2}{4\gamma^{-1}k_BT\Delta t}\right],$$

where we have assumed that all the γ_i's are the same. Now, the sensitivity weight, ω, for parameter θ at time t_{n+1} is just

$$\omega\left(t_{n+1};\theta\right) = \frac{d}{d\theta}\ln Pr\left(X_{n+1}\right) = \frac{d}{d\theta}\ln Pr\left(X_{n+1}|X_n\right) + \omega\left(t_n;\theta\right).$$

Inserting the definition for the transition probability and applying the derivative results in

$$\omega_{n+1} = \omega_n + \sqrt{\frac{\Delta t}{2k_BT\gamma}}\xi_n \cdot \frac{d}{d\theta}F_c(X_n;\theta),$$

where F_c is the $3N$ dimensional force vector, and ξ_n is the $3N$ dimensional vector of random numbers used to advance the solution from X_n to X_{n+1}.

7.3. Numerical experiments

7.3.1. Harmonic well

Recall that the harmonic well potential and force are

$$U(x;\theta) = \frac{1}{2}\theta\left(x-\mu\right)^2, \qquad F_c(x;\theta) = \theta\left(x-\mu\right).$$

For the first numerical experiment, we compute the θ-sensitivity of the time-dependent mean-square displacement of x from μ. The mean-square displacement is defined as

$$\mathrm{MSD}(t) := \mathbb{E}\left[(x(t)-\mu)^2\right]$$

Applying results from the previous section, we know that for Brownian dynamics

$$\mathrm{MSD}(t) = e^{-2\theta t/\gamma}(x(0)-\mu)^2 + \frac{1-e^{-2\theta t/\gamma}}{\theta}k_BT$$

The exact θ-sensitivity for this problem is

$$\frac{d}{d\theta}\mathrm{MSD}(t) = \frac{-2t}{\gamma}e^{-2\theta t/\gamma}(x(0)-\mu)^2 + \frac{(-2t/\gamma)\theta - (1-e^{-2\theta t/\gamma})}{\theta^2}k_BT$$

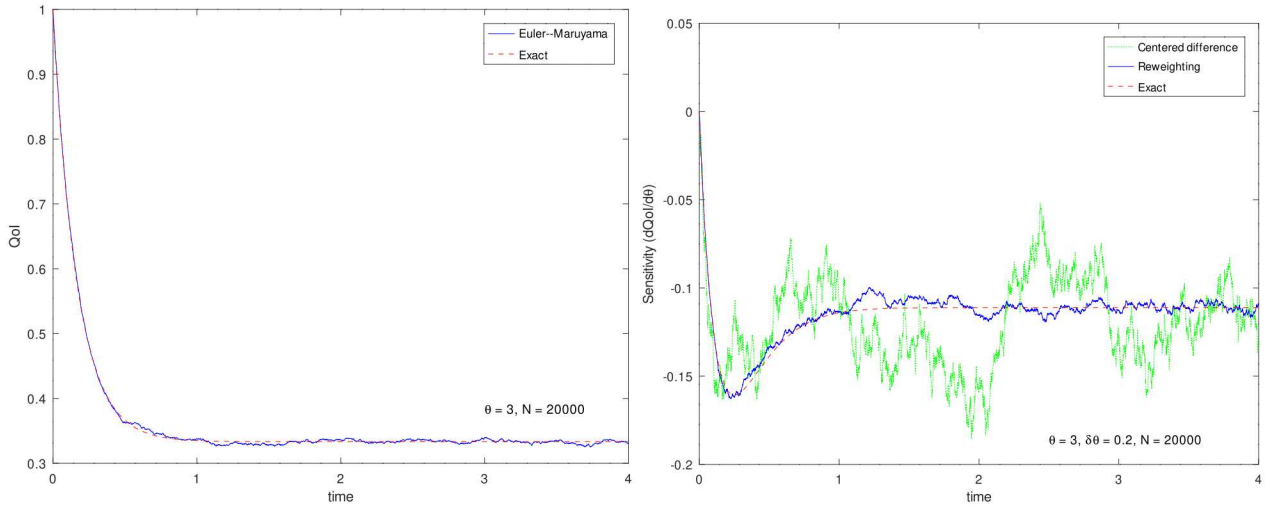

Figure 7-1 Using reweighting to estimate the derivative of the mean-square-displacement quantity of interest for the harmonic well. The blue curves are computed using 20,000 independent walkers, with a spring constant of $\theta = 3$, and using the trajectory reweighting scheme. The red curves are the analytic (exact) solution. On the left, the mean-square-displacement as a function of time. On the right, the sensitivity of the mean-square-displacement to changes in the spring constant, θ. The green curve on the right shows the result from using simple centered finite differencing with $\theta = 2.9, 3.1$.

In Figure 7-1, we show the results from an Euler–Maruyama simulation of the Brownian dynamics of 20,000 independent walkers in a potential well. For this simulation, we used the following parameter values: $\theta = 3$, $k_B T = 1$, $\mu = 1$, $\gamma = 1$, and $\Delta t = 0.001$. On the left, we show the evolution of the quantity of interest (the mean-square-displacement) as a function of time from $t = 0$ to $t = 4$. Due to the large number of walkers, the Euler–Maruyama result (in blue) closely matches the exact solution (in red). On the right, we show the sensitivity of the mean-square-displacement to changes in the spring constant, θ, using the trajectory reweighting scheme (in blue) and the analytic solution (in red). The numerical result matches the analytic result very closely for short times, with increased variance at longer times. To compare the reweighting estimator to a "black box" approach, we use simple (uncorrelated) centered finite differencing with $\theta_1 = 2.9$ and $\theta_2 = 3.1$. The result is the high-variance green curve on the right in Figure 7-1. In this example, the centered difference estimator has higher variance and is computationally more expensive, requiring two additional simulations. In addition, the centered difference estimator requires selecting a parameter step size to optimally balance the truncation error with the statistical error (something we did not do in this example). In contrast, the reweighting scheme has no "step size parameter", is lower variance, requires less computational effort, and has a higher floating point operations to memory access ratio.

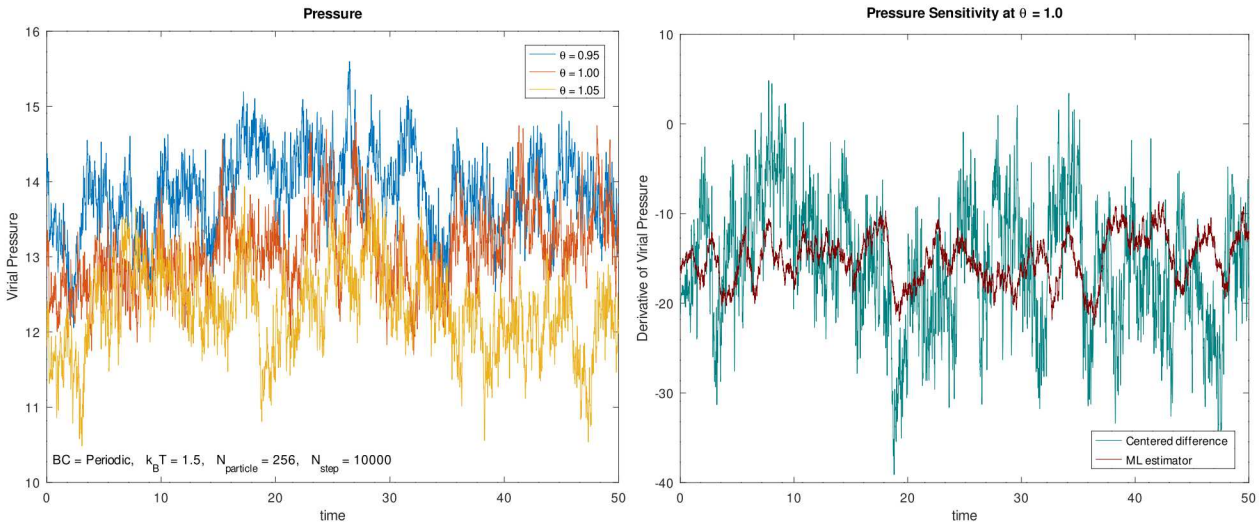

Figure 7-2 Using reweighting to estimate the derivative of the virial pressure with respect to the Lennard–Jones attractive constant, θ. On the left, the time-dependent virial pressure for $\theta = 0.95, 1.00,$ and 1.05. On the right, the sensitivity of the virial pressure to θ estimated using centered finite differences and trajectory reweighting.

7.3.2. Lennard–Jones gas

The Lennard–Jones potential is well-known molecular-dynamics pair potential. The total potential can be expressed as a sum over all pairs of atoms in the system,

$$U(\vec{x}) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i}^{N} U_{LJ}(r_{ij})$$

$$U_{LJ}(r) = 4\varepsilon\left(\left(\frac{r_0}{r}\right)^{12} - \theta\left(\frac{r_0}{r}\right)^{6}\right)$$

where r_{ij} is the minimum distance between a pair of atoms. The core potential is parameterized by ε, r_0, and θ. The parameter θ controls the strength of the attractive component of the force. The reduce boundary effects, we use periodic boundary conditions, and truncate the potential at half the box length, so an atom does not see its periodic image.

A standard observable computed in molecular dynamics simulations is the pressure,

$$P = \frac{1}{V}\left(Nk_BT - \frac{1}{6}\sum_{i=1}^{N}\sum_{j\neq i}^{N} r_{ij}U'_{LJ}(r_{ij})\right)$$

where V is the volume of the periodic cell and U'_{LJ} is the derivative of the Lennard–Jones pair potential.

We would like to compute the sensitivity of the pressure to changes in the attractive force constant, θ. We start by performing a series of simulations with 256 particles in a periodic box with number density of 0.95, $k_BT = 1.5$, and $\theta = 0.95, 1.00, 1.05$. On the left in Figure 7-2, we

show the pressure as a function of time over 10000 time steps with a step size of 0.05. As expected the pressure decreases as θ increases. On the right in Figure 7-2, we show two different estimates for the sensitivity of the pressure to changes in θ. The first is the simple uncorrelated centered difference estimator shown in cyan. The second is the reweighting estimator (also known as the maximum likelihood estimator) shown in red. For this problem we find that the reweighting estimator has lower variance when compared to the centered difference estimator, and requires less computational overhead, using one simulation instead of the two additional simulations required for centered differencing. It is also worth mentioning that the additional overhead for the reweighting scheme is one dot product between the conservative force and the random force (two quantities already required for a standard simulation). This additional dot product is of negligible cost compared to the cost of a full N-body force calculation, and can be computed in tandem with the forces.

8. PARTICLE TRANSPORT WITH TRANSITIONS BETWEEN PURE ABSORBERS

In this section, we connect the work of this report with the application area of radiation transport. We build upon the work accomplished in Bond et al. [2019] that related the differential operator approach Rief [1994] and the Malliavin derivative. The former had the goal of determining an approximation to a sensitivity by reusing samples used within the existing Monte Carlo method.

For a purely absorbing problem with no time-dependence and no internal source, the Boltzmann transport equation simplifies to the ordinary differential equation

$$\begin{cases} \mu \dfrac{\partial \phi(x,\mu)}{\partial x} + \sigma_a(x)\phi(x,\mu) = 0 \qquad x > 0, -1 < \mu < 1 \\ \phi(0,\mu) = \phi_0(\mu) \qquad 0 \leqslant \phi_0(\mu) \end{cases} \tag{51}$$

where ϕ is the angular particle flux, $\mu = \cos(\theta)$ is the particle direction, and σ_a is the absorption cross section. The flux is defined as the density of particles passing through position x traveling in direction μ prior to absorption given an initial density.

We consider a problem consisting of two regions with different interaction cross sections, such that

$$\sigma_a(x) = \begin{cases} \sigma_1, & 0 < x < \ell \\ \sigma_2, & \ell < x \end{cases}. \tag{52}$$

We note that Eqs. (51) and (52) are similar to problems analyzed in section 4, with the introduction of our physically motivated parameters of x to denote an axial position and an angle variable μ that relates axial distance traveled to total particle distance traveled, the basis on which interaction probability σ_a is expressed.

We are interested in obtaining sensitivities of the particle flux ϕ with respect to the boundary location ℓ. More generally, we are interested in reaction rates that are an inner product with the

particle flux, and we are interested in boundary locations that characterize the geometry dimensions of a system. In the following we demonstrate that we can calculate these sensitivities directly using Monte Carlo simulations while simultaneously solving Eq. (51). The value in this approach is the capability to leverage gradient-based design and uncertainty quantification tools, as discussed in the introduction.

The solution for this transport problem can be written directly as

$$\phi(x,\mu) = \begin{cases} \phi_0(\mu)\exp\left(-\sigma_1\dfrac{x}{\mu}\right), & 0 < x < \ell \\ \phi_0(\mu)\exp\left(-(\sigma_1-\sigma_2)\dfrac{\ell}{\mu}\right)\exp\left(-\sigma_2\dfrac{x}{\mu}\right), & \ell < x \end{cases} \tag{53}$$

and so provides the survival probability when we also suppose $\phi_0 = 1$. If X denotes the random variable distributed with respect to the flux ϕ, then the survival probability can be written as $\phi(x,\mu) = \mathbb{P}(X > x/\mu)$. We remark that the identity

$$\mathbb{P}\left(X > \frac{x}{\mu}\right) = \mathbb{E}(\mathbb{1}_{X>x/\mu}) \tag{54}$$

explains that the survival probability is equivalent to the expected fraction of particles absorbed beyond x/μ. The Monte-Carlo simulation approximates the expectation by sampling X and tallying the fraction of times it exceeds x/μ.

The sensitivity of the flux to the position of the geometry boundary at ℓ is

$$\frac{d}{d\ell}\phi(x,\mu) = \begin{cases} 0, & 0 < x < \ell \\ -\dfrac{\sigma_1-\sigma_2}{\mu}\phi(x,\mu), & \ell < x \end{cases} \tag{55}$$

This can then be cast as a relative sensitivity:

$$\frac{1}{\phi(x,\mu)}\frac{d}{d\ell}\phi(x,\mu) = \begin{cases} 0, & 0 < x < \ell \\ -\dfrac{\sigma_1-\sigma_2}{\mu}, & \ell < x \end{cases} \tag{56}$$

This factor for relative sensitivity due to change in a boundary location appears as the total cross section in the expression [Burke and Kiedrowski, 2018, Eq. (68)] derived using the differential operator. That is, in the absence of a scattering cross section, the total cross section is that same as the absorption cross section and the two expressions become identical.

Eq. (56) makes physical sense. The sensitivity of the particles transmitted to some depth x beyond the material boundary at ℓ is proportional to the change in the cross section at the material boundary. If the cross section of the second material is larger than the cross section of the first material and the material boundary is shifted so that there is less of the second material (that is, ℓ is increased), the transmission is increased. The sensitivity is inversely proportional to the cosine of the angle with respect to the material boundary surface. This follows simply from the effect on the particle pathlength in each material (the basis on which survival probability is defined) relative to the shift of the boundary in the axial direction x (the basis on which position ℓ is defined).

It should be noted that as a particle path approaches a direction tangential to the surface (i.e., $\mu \to 0$), the expected value for the sensitivity is unbounded. A similar problem arises with surface flux estimators in Monte Carlo simulations. An ad hoc solution is to modify the estimator by using an approximate contribution for particles that are nearly tangential, as discussed in Dupree and Fraley [2002].

Here we have considered an exponentially distributed random variable. We are interested in more complicated problems, in which Monte Carlo simulations provide the random process through a set of sampling procedures designed to simulate physical processes.

8.1. Numerical Results

We constructed a test problem with $\sigma_1 = 1.0$, $\sigma_2 = 2.0$, and $\ell = 1.5$. All particles were simulated in direction $\mu = 1$. We used a simple Monte Carlo code representative of particle transport codes. The algorithm samples distance to absorption from an exponential probability distribution function parameterized by the absorption cross section, moves the particle to the nearer of either the next boundary or the absorption location, and repeats until the particle is absorbed or passes the last geometry boundary. The particle flux was tallied at each boundary crossing. The tally is an indicator function, with a tally weight of 1 for each boundary crossing versus an implicit tally of 0 if the particle is absorbed before the boundary crossing. Results are normalized by the total number of particles simulated. The statistical average and standard deviation of each boundary flux were estimated. Simulations used 10^7 particles.

Sensitivities were calculated using three methods. The analytical sensitivites are given by (55). Sensitivities were sampled using an in-line implementation of (56), such that when a particle crossed the boundary a relative sensitivity weight was incremented by the difference in the cross sections. At each boundary crossing the flux sensitivity was tallied as the product of the flux tally and the sensitivity weight. Finally, sensitivities were also estimated using a finite difference approach in which an independent simulation was performed with the transition between cross sections ℓ shifted from a depth of 1.5 to 1.49. The sensitivity was estimated as the difference in results divided by the difference in boundary location.

Results for the boundary fluxes are provided in Table 8-1, and results for the sensitivities of the boundary fluxes to the location of the cross section transition are provided in Table 8-2. All results for the in-line sensitivity method are within reasonable statistical agreement of the analytical results, and since the cross section difference is 1.0, estimated flux sensitivities are in exact agreement with the estimated flux and have the same statistical uncertainty.

REFERENCES

Brian M. Adams, Shawn D. Pautz, Laura Painton Swiler, Brian C. Franke, and Ethan Blansett. Adjoint-enabled uncertainty quantification for satellite shield design. In *International Conference on Mathematics & Computational Methods Applied to Nuclear Science & Engineering*, April 2017. SAND2017-3731C.

Table 8-1 Flux Estimates

Depth (cm)	Analytical Flux	Estimated Flux
0.0	1.000000	1.000000 ± 0.000000
1.0	0.367879	0.367916 ± 0.000145
1.49	0.225373	0.225398 ± 0.000115
1.5	0.223130	0.223177 ± 0.000115
2.0	0.082085	0.082074 ± 0.000077
3.0	0.011109	0.011057 ± 0.000032

Table 8-2 Sensitivity Estimates

Depth (cm)	Analytical Sensitivity	Sampling Sensitivity	Finite Difference Sensitivity
0.0	0.000000	0.000000 ± 0.000000	0.000000
1.0	0.000000	0.000000 ± 0.000000	0.003370
1.49	0.000000	0.000000 ± 0.000000	0.003890
1.5	0.223130	0.223177 ± 0.000115	0.229910
2.0	0.082085	0.082074 ± 0.000077	0.083940
3.0	0.011109	0.011057 ± 0.000032	0.012460

Brian M. Adams, Shawn D. Pautz, and Don E. Bruss. Geometric uncertainity quantification and robust design for 2D satellite shielding. In *International Conference on Mathematics and Computational Methods Applied to Nuclear Science and Engineering*, 2019. SAND2019-4852C.

S. Asmussen and P.W. Glynn. *Stochastic Simulation: Algorithms and Analysis*. Stochastic Modelling and Applied Probability. Springer New York, 2007. ISBN 9780387690339. doi: 978-0-387-69033-9.

Stephen Bond, Brian Franke, R.B. Lehoucq, and J. Darby Smith. Developing and evaluating Malliavin estimators for intrusive sensitivity analysis of Monte Carlo radiation transport. Technical Report SAND2019-11524R, Sandia National Labs, 2019.

T.P. Burke and B.C. Kiedrowski. Monte Carlo perturbation theory estimates of sensitivities to system dimensions. *Nuclear Science and Engineering*, 189:199–223, 2018.

Sourav Chatterjee, Jason Fulman, and Adrian Röllin. Exponential approximation by Stein's method and spectral graph theory. *ALEA Lat. Am. J. Probab. Math. Stat*, 8:197–223, 2011.

Nan Chen and Paul Glasserman. Malliavin greeks without Malliavin calculus. *Stochastic Processes and their Applications*, 117(11):1689–1723, 2007. ISSN 0304-4149. doi: 10.1016/j.spa.2007.03.012. Recent Developments in Mathematical Finance.

S.A. Dupree and S.K. Fraley. *A Monte Carlo Primer: A Practical Approach to Radiation Transport*. Number v. 1. Springer US, 2002. ISBN 9780306467486.

P. Glasserman. *Monte Carlo Methods in Financial Engineering*. Applications of mathematics : stochastic modelling and applied probability. Springer, 2004. ISBN 9780387004518. doi: 10.1007/978-0-387-21617-1.

H Malcolm Hudson. A natural identity for exponential families with applications in multiparameter estimation. *The Annals of Statistics*, 6(3):473–484, 1978.

Christophe Ley and Yvik Swan. Parametric Stein operators and variance bounds. *Brazilian Journal of Probability and Statistics*, 30(2):171–195, 2016.

Christophe Ley, Gesine Reinert, and Yvik Swan. Stein's method for comparison of univariate distributions. *Probability Surveys*, 14:1–52, 2017.

Paul Malliavin. Stochastic calculus of variations and hypoelliptic operators. In *Proc. Int. Symp. on Stochastic Differential Equations (Kyoto, 1976)*, pages 195–263, New York, 1976. Wiley.

Elizabeth Meckes. On Stein's method for multivariate normal approximation. In *High dimensional probability V: the Luminy volume*, pages 153–178. Institute of Mathematical Statistics, 2009.

Ivan Nourdin and Giovanni Peccati. Stein's method and exact Berry–Esseen asymptotics for functionals of gaussian fields. *The Annals of Probability*, 37(6):2231–2261, 2009a.

Ivan Nourdin and Giovanni Peccati. Stein's method on Wiener chaos. *Probability Theory and Related Fields*, 145(1-2):75–118, 2009b.

Ivan Nourdin and Giovanni Peccati. *Normal approximations with Malliavin calculus: from Stein's method to universality*, volume 192. Cambridge University Press, 2012.

Ivan Nourdin, Giovanni Peccati, and Anthony Réveillac. Multivariate normal approximation using Stein's method and Malliavin calculus. In *Annales de l'IHP Probabilités et statistiques*, volume 46, pages 45–58, 2010.

Shawn D. Pautz, Brian M. Adams, Laura Painton Swiler, Brian C. Franke, and Ethan Blansett. Adjoint-based sensitivities for optimization of satellite electron/proton shields. In *International Conference on Mathematics & Computational Methods Applied to Nuclear Science & Engineering*, April 2017. SAND2017-1297C.

James CM Peng. Simultaneous estimation of parameters of independent poisson distribution. *Stanford Univ. Technical Report No. 78*, 1975.

Sergey Plyasunov and Adam P. Arkin. Efficient stochastic sensitivity analysis of discrete event systems. *Journal of Computational Physics*, 221(2):724–738, 2007. ISSN 0021-9991. doi: 10.1016/j.jcp.2006.06.047.

Herbert Rief. A synopsis of Monte Carlo perturbation algorithms. *Journal of Computational Physics*, 111(1):33–48, 1994. ISSN 0021-9991. doi: 10.1006/jcph.1994.1041.

Patrick W. Sheppard, Muruhan Rathinam, and Mustafa Khammash. A pathwise derivative approach to the computation of parameter sensitivities in discrete stochastic chemical systems. *The Journal of Chemical Physics*, 136(3):034115, 2012. doi: 10.1063/1.3677230.

Charles M Stein. Estimation of the mean of a multivariate normal distribution. *Stanford Univ. Technical Report No. 48*, 1973.

Charles M Stein. Estimation of the mean of a multivariate normal distribution. *The Annals of Statistics*, pages 1135–1151, 1981.

Patrick B. Warren and Rosalind J. Allen. Malliavin weight sampling for computing sensitivity coefficients in Brownian dynamics simulations. *Physical Review Letters*, 109(25):250601, 2012a. ISSN 0031-9007. doi: 10.1103/physrevlett.109.250601.

Patrick B. Warren and Rosalind J. Allen. Steady-state parameter sensitivity in stochastic modeling via trajectory reweighting. *The Journal of Chemical Physics*, 136(10):104106, 2012b. ISSN 0021-9606. doi: 10.1063/1.3690092.

Patrick B Warren and Rosalind J Allen. Malliavin weight sampling: A practical guide. *Entropy*, 16(1):221–232, 2014.

DISTRIBUTION

Hardcopy—External

Number of Copies	Name(s)	Company Name and Company Mailing Address

Hardcopy—Internal

Number of Copies	Name	Org.	Mailstop
1	D. Chavez, LDRD Office	1911	0359

Email—Internal [redacted]

Name	Org.	Sandia Email Address
Technical Library	01177	libref@sandia.gov

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.