Exponentially Rare Large Fluctuations: Computing their Frequency

Robert S. Maier
Mathematics and Physics Departments
University of Arizona
rsm@math.arizona.edu

Dec. 1, 2000

THE UNIVERSITY OF
ARIZONA
TUCSON ARIZONA

Abstract

Random processes that model asset price evolution, such as geometric Brownian motion or more general diffusion processes with drift, may occasionally deviate radically from their expected values. The smaller the volatility, the less frequently a deviation of specified size will occur. Such occurrence rates fall off exponentially, and are hard to estimate via Monte Carlo simulation or numerical integration of diffusion equations.

Large deviation theory provides a way, not based on simulation, of estimating the rate at which such rare events take place. The key concept is that of an *optimal trajectory*: the asymptotically most likely trajectory along which a fluctuation of specified size will occur.

I explain a new, enhanced version, which may be implemented numerically, and applied to multidimensional diffusion processes with drift. It employs a *system of ordinary differential equations*, which must be integrated along any optimal trajectory.

The new technique permits the estimation not only of asymptotic exponential falloff rates, but also of the all-important pre-exponential factor: the constant that multiplies the exponential. This yields precise predictions, in the small-volatility limit at least, for the (exponentially small!) rate at which rare fluctuations occur.

Tail probabilities computed from diffusion models often do not agree well with empirical data. The new technique applies, suitably modified, to jump-diffusion processes as well. So it should facilitate choosing a model that fits empirical data, even on the tails.

Recent Work

- R. S. Maier and D. L. Stein, "Noise-activated escape from a sloshing potential well." Explains the theoretical and numerical computation of first passage time asymptotics for *periodically modulated* 1-D diffusion processes. Available at `http://www.math.arizona.edu/~rsm`.

- R. S. Maier and D. L. Stein, "Limiting exit location distributions in the stochastic exit problem." Treats a boundary crossing problem for 2-D drift-diffusion processes, and the computation of limiting crossing location distributions. *SIAM J. on Applied Mathematics* 57 (1997), 752–790.

- R. S. Maier, "Communications networks as stochastically perturbed nonlinear systems." Stresses that large deviation theory for discrete jump processes is not the same as for diffusion processes. *Proc. 30th Allerton Conference on Communication, Control, and Computing.*

Related Approaches

- Rare event probability estimation via simulation; importance sampling; optimal ‘exponential twisting’ defined variationally. [ECE community; queueing theory and stochastic networks context.]

- Quasi-analytic weak-volatility limits for drift-diffusion processes via singular perturbation theory and matched asymptotic approximations. [Applied mathematics community, B. Matkowsky, Z. Schuss, et al., before early 1990s.]

- Quasi-analytic weak-volatility limits via optimal trajectory formalism, Hamilton’s equations rather than Euler–Lagrange equations. [Theoretical physics community, M. I. Dykman et al. and RSM; since early 1990s.]

A Simple Large Deviation Example

- If $\{X_i\}_i$ are i.i.d., with $P(X_i = 1)$ and $P(X_i = -1)$ equal to $1/2$, let (with "large deviation scaling")

$$S_n := \sum_{i=1}^{n} X_i$$

$$x^{(n)}(t) := n^{-1} S_{\lfloor nt \rfloor}.$$

  The n-indexed family of processes $x^{(n)}(t)$, $t \geq 0$ converges to zero as $n \to \infty$, and naively resembles the family of scaled Brownian motions $n^{-1/2}B(t)$, $t \geq 0$. (Central limit theorem!)

- However, the central limit theorem applies as $n \to \infty$ only in the *small deviation regime*: when $x^{(n)}(t)$ and $n^{-1/2}B(t)$ are an $O(1)$ number of standard deviations from their mean (zero), i.e., they are $O(n^{-1/2})$, i.e., if $S_n = O(n^{1/2})$.

- For fixed $t > 0$, the event $S_{\lfloor nt \rfloor} \geq cn$ is an *exponentially unlikely event* as $n \to \infty$. In terms of the rescaled family of processes $x^{(n)}(t)$, this is the *above-a-level* event $x^{(n)}(t) \geq c$.

- Naively, this is like the event $n^{-1/2}B(t) \geq c$: an above-a-level event for a small-volatility Brownian motion, or a tail event $B(t) \geq cn^{1/2}$ for a standard Brownian motion.

- But at any fixed $t > 0$, $P(x^{(n)}(t) \geq c)$ and the (seemingly similar!) $P(n^{-1/2}B(t) \geq c)$ have *different exponential falloff rates* as $n \to \infty$.

- A trivial example of this: if $c > 1$, the first is identically zero, but the second is not. *In the large deviation regime*, the two families differ.

From Cramér's Theorem to Optimal Trajectories

- Central Limit Theorem: $n^{-1/2} S_n \sim Z$ as $n \to \infty$, where Z is standard normal: $f_Z(z) = (2\pi)^{-1/2} \exp(-z^2/2)$.

- Cramér's Theorem: $n^{-1} S_n$, as $n \to \infty$, has density asymptotics that are *not* those of $n^{-1/2} Z$, i.e., $(2\pi/n)^{-1/2} \exp(-nz^2/2)$. Rather, $f_{n^{-1}S_n}$ satisfies:

$$f_{n^{-1}S_n}(v) \sim \exp(-nI(v)), \quad n \to \infty,$$

  to *logarithmic accuracy*. (Various assumptions. . .)

- $I = I(v)$ is a *rate function*, quantifying the *exponential rareness* of large deviations from the mean. If the $\{X_i\}$ are i.i.d. standard normal, then $I(v) = v^2/2$. But if the $\{X_i\}$ are i.i.d. ± 1, then $I(v) = +\infty$ for $|v| > 1$.

- Rate function result: $I = I(v)$ is the *Legendre transform* of the *cumulant generating function* of any of the i.i.d. increments $\{X_i\}$.

$$\begin{aligned} M(p) &:= E[\exp(pX)] \\ \log M(p) &:= \log E[\exp(pX)] \\ I(v) &:= \sup_{p \in \mathbf{R}} [pv - \log M(p)]. \end{aligned}$$

- For example, if X is standard normal then $\log M(p) = p^2/2$ and $I(v) = v^2/2$. But if X takes ± 1 values with equal probability, then

$$\begin{aligned} M(p) &= \cosh p, \\ \log M(p) &= \log \cosh p, \\ I(v) &= \frac{1}{2}[(1+v)\log(1+v) + (1-v)\log(1-v)] \end{aligned}$$

(provided $|v| < 1$; otherwise $I(v) = +\infty$).

- Generalization to the random processes $x^{(n)}(t)$? (Wentzell–Freidlin and Donsker–Varadhan, 1970s.) A *large deviation principle* for the associated measures states that the "probability that $x^{(n)}(t)$ tracks any specified trajectory $t \mapsto x(t)$", uniformly over $t \in [0,T]$, has asymptotics

$$\sim \exp\left(-n\int_0^T I(\dot{x}(t))\,dt\right), \qquad n \to \infty,$$

to logarithmic accuracy. $x(0) = 0$ is required. This is an *asymptotic Feynman–Kac formula*.

- More generally, most reasonable events, such as $x^{(n)}(t_0) \geq c > 0$ at a fixed time $t_0 \in [0,T]$, will become *exponentially rare* as $n \to \infty$, with the exponential falloff rate equalling the *infimum* of the above integral, computed over all trajectories $t \mapsto x(t)$ comprised by the event. The integral is a *rate functional*.

Optimal Trajectories

- Mathematical definition: An optimal trajectory for the indexed family of processes $x^{(n)}(t)$ is the trajectory $t \mapsto x(t)$ which (1) satisfies $x(0) = 0$, (2) satisfies $x(t_0) = x'$ for some specified $t_0 > 0$ and x' ('final endpoint condition'), and (3) minimizes the rate functional.

- Intuitive definition: An optimal trajectory from $x = 0$ at time zero to $x = x'$ at time t_0 is the *most probable such trajectory* in the $n \to \infty$ limit. In that limit, all trajectories that do not maintain a constant zero value are exponentially suppressed. The optimal trajectory extending to a point x' at time t_0 is the *least suppressed* one.

- Easy to check: for processes like $x^{(n)}(t)$, whose increment distribution is time- and value-independent, the optimal trajectories are *straight lines* emanating from $x = 0$.
  (If $X = \pm 1$, slope $\in (-1, 1)$.)

Extension to General Drift-Diffusion Processes

A stochastic differential equation for an n-indexed family of processes $X^{(n)}(t)$, $t \geq 0$:

$$dX = b(X)\,dt + n^{-1/2}\sigma(X)\,dB$$

(b=value-dependent drift, σ=normalized value-dependent volatility; Itô interpretation).

The exponential falloff generalizes to:

$$\sim \exp\left(-n \int_0^T L(x(t), \dot{x}(t))\,dt\right), \quad n \to \infty.$$

For a scaled family of jump processes, $L(x, \bullet)$ would be the Legendre transform of the (possibly value-dependent!) cumulant generating function of the increments. For this drift-diffusion process,

$$L(x, \dot{x}) := \frac{1}{2\sigma^2(x)} \left|\dot{x} - b(x)\right|^2.$$

The Vector-Valued Generalization

The stochastic differential equation for an n-indexed family of $\mathbf{R}^d$-valued drift-diffusion processes $\mathbf{X}^{(n)}(t)$:

$$d\mathbf{X} = \mathbf{b}(\mathbf{X})\,dt + n^{-1/2}\sigma(\mathbf{X}) \cdot d\mathbf{B}$$

($\mathbf{b}$=state-dependent drift field, σ=normalized state-dependent volatility matrix; Itô interpretation).

The exponential falloff generalizes to:

$$\sim \exp\left(-n\int_0^T L(\mathbf{x}(t), \dot{\mathbf{x}}(t))\,dt\right), \quad n \to \infty,$$

where the *Lagrangian function* L is defined by

$$L(\mathbf{x}, \dot{\mathbf{x}}) := \frac{1}{2}\,[\dot{\mathbf{x}} - \mathbf{b}(\mathbf{x})] \cdot \mathbf{D}^{-1}(\mathbf{x}) \cdot [\dot{\mathbf{x}} - \mathbf{b}(\mathbf{x})].$$

Here $\mathbf{D} := \sigma\sigma^t$ is the (in general, value-dependent) *diffusivity matrix*.

General Optimal Trajectories

Optimal trajectories

1. are not straight lines in $\mathbf{R}^d$, in general. (Cf. the small-volatility limit of geometric Brownian motion.)

2. may be computed variationally by minimizing the rate functional. This leads to the Euler–Lagrange equations
$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\mathbf{x}}}\right) - \frac{\partial L}{\partial \mathbf{x}} = 0.$$

3. are only a starting point. The probability density for the random process $\mathbf{X}^{(n)}$ equalling $\mathbf{x}'$ at some $t = t_0 > 0$ will fall off exponentially as $n \to \infty$, if $\mathbf{x}'$ is *not* on the integral curve of the drift field extending from the initial value of $\mathbf{X}^{(n)}$.

   The exponential falloff rate must be computed by integrating L along the optimal trajectory terminating at this endpoint.

Application to Stationary Distributions

- Let an n-indexed family of $\mathbf{R}^d$-valued drift-diffusion processes $\mathbf{X}^{(n)}(t)$ be defined as above by a state-dependent drift field $\mathbf{b} = \mathbf{b}(\mathbf{x})$, and a normalized state-dependent volatility matrix, $\sigma = \sigma(\mathbf{x})$. (*Reminder*: true volatility $\propto n^{-1/2}$.)

- Suppose that these processes are *globally stable*, with unique stationary (i.e., time-invariant) probability distributions $\rho_0 = \rho_0^{(n)}(\mathbf{x})$, and that the drift $\mathbf{b}$ has a single *attractor*, near which $\rho_0^{(n)}$ becomes *exponentially concentrated* as $n \to \infty$.

  – *Example*: $\mathbf{b}(\mathbf{x}) := -\mathbf{x}$ and $\sigma := \sigma\mathbf{I}$ yields a d-dimensional Ornstein–Uhlenbeck process, with $\mathbf{x} = \mathbf{0}$ the only attractor: an isolated point. Here, $\rho_0^{(n)}(\mathbf{x}) = (2\pi\sigma^2/n)^{-1/2}\exp(-n\,|\mathbf{x}|^2/2\sigma^2)$.

- Then, the asymptotics of $\rho_0^{(n)}$ can be worked out to logarithmic accuracy: $\rho_0^{(n)}(\mathbf{x}) \sim \exp\left(-nW(\mathbf{x})\right)$.

The *Action Function* $W : \mathbf{R}^d \to \mathbf{R}^+$

1. is zero on the attractor, and strictly positive elsewhere.

2. quantifies the frequency with which the random process $\mathbf{X}^{(n)}$ visits the neighborhood of any point $\mathbf{x}' \in \mathbf{R}^d$, in the $n \to \infty$ (small-volatility) limit. It is a *state-dependent exponential falloff rate*.

3. can be computed from the rate functional, by taking the infimum not only over all trajectories $t \mapsto \mathbf{x}(t)$ extending from the attractor to $\mathbf{x}'$, but over all *transit times*. This yields the *dominant fluctuational trajectory* terminating at $\mathbf{x}'$.

   - *Example*: If $\mathbf{b}(\mathbf{x}) := -\mathbf{x}$ and $\sigma := \sigma\mathbf{I}$, then the dominant trajectory extending from $\mathbf{x} = \mathbf{0}$ to $\mathbf{x} = \mathbf{x}'$ is a straight line with nonuniform speed, namely $\mathbf{x}(t) = e^t\,\mathbf{x}'$. It has *infinite transit time*: it emerges from $\mathbf{x} = \mathbf{0}$ at $t = -\infty$ and approaches $\mathbf{x} = \mathbf{x}'$ at, say, $t = 0$.

Computing Dominant Trajectories and the Action Function

- A pattern of outgoing dominant trajectories, and the value of W at each of their endpoints, can be generated more easily from *Hamilton's equations* (1st-order) than from the Euler–Lagrange equations (2nd-order).

- For this, both the trajectory $t \mapsto \mathbf{x}(t)$ and an auxiliary 'momentum trajectory' $t \mapsto \mathbf{p}(t)$ would be numerically generated, by:

$$\dot{\mathbf{x}} = \frac{\partial H}{\partial \mathbf{p}}(\mathbf{x}, \mathbf{p}), \qquad \dot{\mathbf{p}} = -\frac{\partial H}{\partial \mathbf{x}}(\mathbf{x}, \mathbf{p}).$$

  The *Hamiltonian function* $H : \mathbf{R}^d \times \mathbf{R}^d \to \mathbf{R}$ is defined by:

$$H(\mathbf{x}, \mathbf{p}) := \frac{1}{2}\mathbf{p} \cdot \mathbf{D}(\mathbf{x}) \cdot \mathbf{p} + \mathbf{b}(\mathbf{x}) \cdot \mathbf{p}.$$

- In theoretical physics, H would be an *energy function*, and energy is conserved: $H \equiv \mathrm{const}$ along any optimal or dominant trajectory.

- Remarkably, for *dominant* trajectories (with the infimum taken over all transit times, normally yielding an infinite transit time), this energy is *zero*. The trajectory $t \mapsto (\mathbf{x}, \mathbf{p}) \in \mathbf{R}^d \times \mathbf{R}^d$ lies in a *zero-energy hypersurface*, which has codimension unity in $\mathbf{R}^d \times \mathbf{R}^d$.

- The numerical punchline:

$$W(\mathbf{x}') = \int \mathbf{p} \cdot \dot{\mathbf{x}} \, dt = \int \mathbf{p} \cdot d\mathbf{x},$$

  the line integral being taken along the dominant trajectory from the attractor to $\mathbf{x}'$.

The Boundary Crossing Application

- Suppose the attractor for the drift $\mathbf{b}$ on $\mathbf{R}^d$ is in the interior of a region Ω, which is attracted to it. (Boundary is denoted $\partial\Omega$, with coordinate s.)

- Goal: the *boundary crossing location distribution* for $\mathbf{X}^{(n)}$ on $\partial\Omega$, as $n \to \infty$ (small-volatility limit). Harmless assumption: $\mathbf{X}^{(n)}(0)$ is on the attractor.

- Result (Donsker–Varadhan and Wentzell–Freidlin, 1970s): If the action function W has a unique minimum on $\partial\Omega$ at some point $s_{\min}$, then the crossing location distribution $p^{(n)}(s)\,ds$ *concentrates* at $s = s_{\min}$ as $n \to \infty$.

- Formal extension: If $W = W(s)$ has a continuous 2nd derivative at $s = s_{\min}$, then the crossing location distribution is *asymptotically normal*:

$$\propto \exp(-nW''(s_{\min})\,|s - s_{\min}|^2/2)\,ds.$$

THE UNIVERSITY OF ARIZONA
TUCSON ARIZONA

Precise Small-Volatility Asymptotics

- How to go beyond the small-volatility exponential falloff rate provided by the action function W?
- How to compute second partial derivatives $\partial^2 W/\partial x_i \partial x_j$ efficiently?
- An analytic (non-probabilistic!) approach to the first: introduce a refined approximation

$$\rho_0^{(n)}(\mathbf{x}) \sim K(\mathbf{x}) \exp\left(-nW(\mathbf{x})\right), \qquad n \to \infty,$$

to the stationary solution $\rho_0^{(n)}$ of the forward diffusion equation (a parabolic PDE):

$$\dot{\rho} = \frac{1}{2n} \sum_{i,j=1}^{d} \frac{\partial^2}{\partial x_i \partial x_j} [D_{ij}(\mathbf{x})\rho] - \sum_{i=1}^{d} \frac{\partial}{\partial x_i} [b_i(\mathbf{x})\rho].$$

- Substitute, and set the coefficients of n^1 and n^0 to zero. Result: equations for both W and (new!) the *pre-exponential function* K.

- A *Hamilton–Jacobi equation* for W:

$$H(\mathbf{x}, \nabla W(\mathbf{x})) = 0.$$

  Solving this equation yields the familiar dominant trajectories (“zero-energy optimal trajectories”), and the familiar numerical integration scheme.

- An equation for the pre-exponential function K:

$$\dot{K} = -\left(\nabla \cdot \mathbf{b}(\mathbf{x}) + \frac{1}{2}\sum_{i,j=1}^{d} D_{ij}(\mathbf{x})\frac{\partial^2 W}{\partial x_i \partial x_j}(\mathbf{x})\right) K.$$

  To yield $K(\mathbf{x}')$, this must be *integrated along the dominant trajectory* from the attractor to $\mathbf{x}'$.

A Matrix Riccati Equation

- How to compute the *Hessian matrix* $(W_{,ij}) \equiv (\partial^2 W/\partial x_i \partial x_j)$, along any dominant trajectory?

- Answer: Manipulation of the H.–J. equation yields a *matrix Riccati equation*: $\dot{W}_{,ij}$ equals

  $$-\sum_{k,l=1}^{d} W_{,ik} D_{kl} W_{,lj} - \sum_{k=1}^{d} [W_{,ik} b_{k,j} + (i \leftrightarrow j)] - \sum_{l=1}^{d} b_{l,ij}\, p_l$$

  along any dominant trajectory.

- A *triangular numerical scheme* now follows:
  - Compute a dominant trajectory $t \mapsto (\mathbf{x}(t), \mathbf{p}(t))$.
  - Simultaneously, integrate the matrix Riccati equation along the dominant trajectory.
  - Also simultaneously, integrate the equation for K along the dominant trajectory.

Potential Numerical Problems

- The coupled ordinary differential equations for $t \mapsto (\mathbf{x}(t), \mathbf{p}(t))$, for $t \mapsto W(\mathbf{x}(t))$, for the matrix $t \mapsto (W_{,ij}(\mathbf{x}(t)))$, and $t \mapsto K(\mathbf{x}(t))$, are quite stiff.

  $\Rightarrow$ Solution: Use a good integration algorithm.

- Dominant trajectories $t \mapsto \mathbf{x}(t)$ extending from the attractor may *cross*, yielding *multivaluedness* of W and K. (More than one trajectory to $\mathbf{x}'$!)

  $\Rightarrow$ Interpretation: The *least-action* dominant trajectory is the relevant one (since action is an exponential falloff factor).

- The prefactor K, when integrated along outgoing dominant trajectories, may diverge.

  $\Rightarrow$ Fact: This only happens on 'irrelevant' dominant trajectories, which may bounce off *caustics*.

Exotic Extensions

- What if the drift field $\mathbf{b}$ on $\mathbf{R}^d$ has *two attractors*? Then as $n \to \infty$ (small-volatility limit), fluctuations from the vicinity of one to the other become exponentially rare.

  $\Rightarrow$ The frequency of *fluctuations between attractors* can be computed from the flux of probability, in a transient situation, over the separatrix between them. Result: exponential falloff as $n \to \infty$, and the pre-exponential factor, are both computable. The limiting *separatrix crossing location distribution* is computable too. (See SIAM J. Appl. Math. paper).

- What if the drift field is periodically modulated, rather than static?

  $\Rightarrow$ Work on an expanded state space: $\mathbf{R}^d \times [0, T)$, rather than on $\mathbf{R}^d$. On this *cylinder*, dominant trajectories spiral out from a loop-like attractor. (See new preprint.)

Summary and Conclusions

1. This approach is useful in numerically approximating exponentially small quantities, such as:
   - the probability that a Markov diffusion process with small volatility will wander far from where its drift would take it, and
   - the frequency that a *recurrent* Markov diffusion process will undergo a fluctuation of specified size, in a specified direction.

   Moreover, it yields approximations to boundary crossing location distributions.

2. It extends large deviation theory, by constructing (analytically!) small-volatility approximations to solutions of the diffusion equation.

3. This approach owes much to theoretical physics. (Hamilton's eqns., PDE approximations, etc.).

4. Extensions to jump processes are possible.