Introduction to large deviation theory: Theory, applications, simulations

Hugo Touchette

School of Mathematical Sciences
Queen Mary, University of London

3rd International Summer School on Modern Computational Science:
Simulation of Rare and Extreme Events
Oldenburg, Germany
15-26 August 2011

Outline

Themes

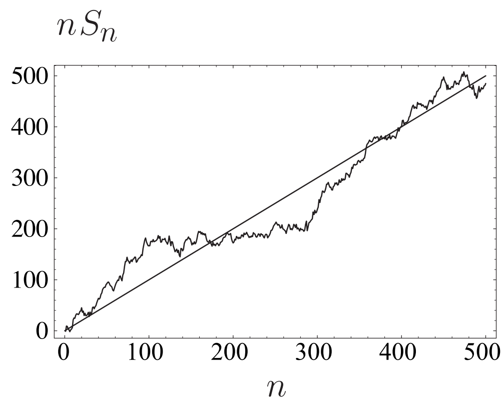
- Random variables / stochastic systems
- Most probable values / typical states
- Fluctuations around these states
- Small vs large deviations / fluctuations / rare events

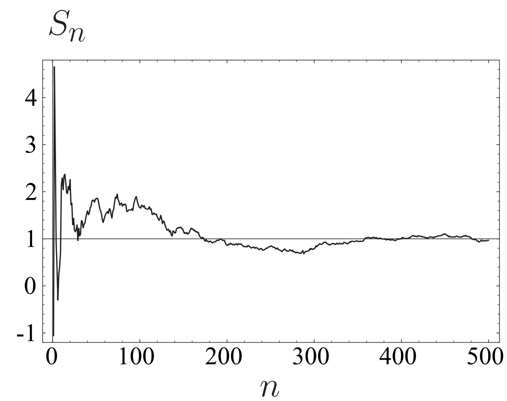
1. Large deviation theory
2. Applications
3. Simulations I
4. Simulations II

- Lecture notes: `arxiv:1106.4146`
- H. Touchette, The large deviation approach to statistical mechanics, Physics Reports **478**, 1-69, 2009. `arxiv:0804.0327`

Example: Sum of Gaussian random variables

$$S_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad p(X_i = x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)}$$





Basic observations

- $S_n \to \mu$ in probability
- Fluctuations $\sim 1/\sqrt{n} \to 0$

Sum of Gaussian random variables (cont'd)

- Probability density function (pdf) of S_n:

$$p(S_n = s) = \sqrt{\frac{n}{2\pi\sigma^2}} e^{-n(s-\mu)^2/(2\sigma^2)}$$

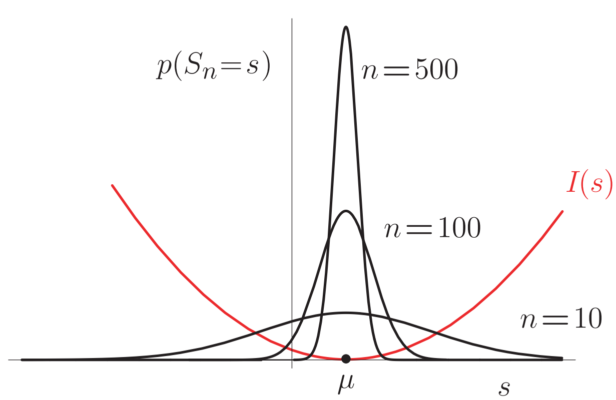
- Variance: $\mathrm{var}(S_n) = \frac{\sigma^2}{n} \to 0$

- Dominant part:

$$p(S_n = s) \approx e^{-nI(s)}$$

- Rate function:

$$I(s) = \frac{(s-\mu)^2}{2\sigma^2}$$



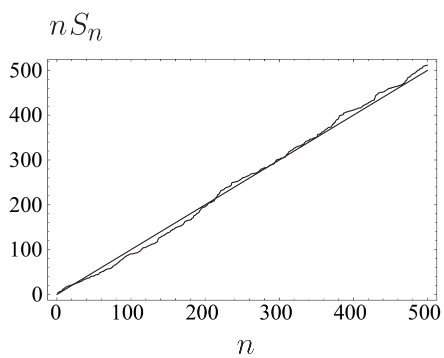
Exercise: Calculation of $p(S_n)$ via generating functions

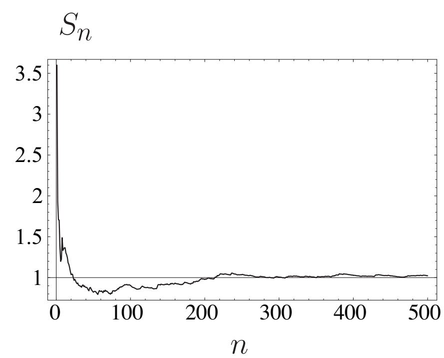
[Exercise 2.7.1]

Example: Sum of exponential random variables

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

$$p(x) = \frac{1}{\mu} e^{-x/\mu}$$



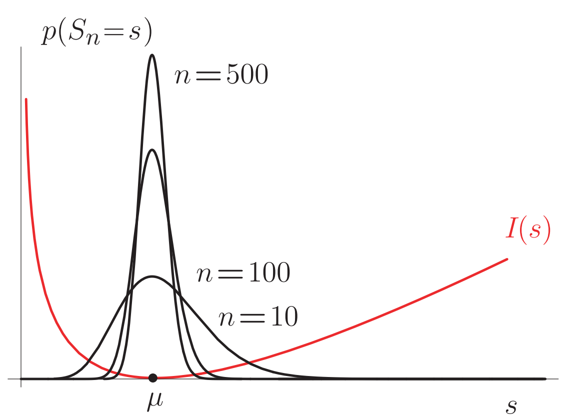


- Large deviation probability:

$$p(S_n = s) \approx e^{-nI(s)}$$

- Rate function:

$$I(s) = \frac{s}{\mu} - 1 - \ln\frac{s}{\mu}$$



Example: Random bits

- Sequence of bits:

$$\omega = \underbrace{010011100101}_{n \text{ bits}}, \qquad \begin{aligned} P(0) &= p_0 \\ P(1) &= p_1 = 1 - p_0 \end{aligned}$$

- Empirical vector:

$$\left.\begin{aligned} L_{n,0}(\omega) &= \frac{\#\ 0\text{'s in } \omega}{n} \\ L_{n,1}(\omega) &= \frac{\#\ 1\text{'s in } \omega}{n} \end{aligned}\right\} \quad \mathbf{L}_n = (L_{n,0}, L_{n,1})$$

- Example:

$$\omega = \underbrace{0001101001}_{n=10}, \qquad \begin{aligned} L_{10,0} &= \frac{6}{10} = \frac{3}{5} \\ L_{10,1} &= \frac{4}{10} = \frac{2}{5} \end{aligned}$$

Random bits (cont'd)

- Probability of a sequence:

$$P(\omega) = p_0^{nL_{n,0}} p_1^{nL_{n,1}}$$

- Probability of $\mathbf{L}_n$:

$$\begin{aligned} P(\mathbf{L}_n = \boldsymbol{\mu}) &= P(L_{n,0} = \mu_0, L_{n,1} = \mu_1) \\ &= \frac{n!}{(n\mu_0)!(n\mu_1)!} p_0^{n\mu_0} p_1^{n\mu_1} \end{aligned}$$

[Exercise 2.7.5]

Large deviation probability

$$P(\mathbf{L}_n = \boldsymbol{\mu}) \approx e^{-nD(\boldsymbol{\mu})}, \quad D(\boldsymbol{\mu}) = \mu_0 \ln \frac{\mu_0}{p_0} + \mu_1 \ln \frac{\mu_1}{p_1}$$

- D = relative entropy
- Zero of D: $\boldsymbol{\mu} = (p_0, p_1)$
- $\mathbf{L}_n \to (p_0, p_1)$ in probability

Example: Spin system

- Spin chain (configuration):

$$\omega = \underbrace{\omega_1, \omega_2, \dots, \omega_n}_{n \text{ spins}}, \quad \omega_i \in \{-1, +1\}$$

- Mean magnetization:

$$M_n = \frac{1}{n} \sum_{i=1}^{n} \omega_i$$

- Density of states:

$$\Omega(m) = \# \text{ configurations } \omega \text{ with } M_n = m$$

Large deviation form

$$\Omega(m) \approx e^{ns(m)}, \quad s(m) = -\frac{1-m}{2} \ln \frac{1-m}{2} - \frac{1+m}{2} \ln \frac{1+m}{2}$$

Large deviation theory

- Random variable: A_n
- Probability density: $p(A_n = a)$

Large deviation principle (LDP)

$$p(A_n = a) \approx e^{-nI(a)}$$

- Meaning of $\approx$:

$$\ln p(a) = -nI(a) + o(n)$$

$$\lim_{n \to \infty} -\frac{1}{n} \ln p(a) = I(a)$$

- Rate function: $I(a) \geq 0$

Goals of large deviation theory

1. Prove that a large deviation principle exists
2. Calculate the rate function

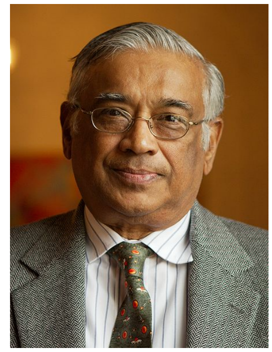
Varadhan's Theorem

- Exponential average:

$$\langle e^{nf(A_n)} \rangle = \int p(A_n = a)\, e^{nf(a)}\, da$$

- Assume LDP for A_n:

$$p(A_n = a) \approx e^{-nI(a)}$$



- Courant Institute
- Abel Prize 2007

Theorem: Varadhan (1966)

$$\lambda(f) = \lim_{n\to\infty} \frac{1}{n} \log\langle e^{nf(A_n)} \rangle = \max_a \{f(a) - I(a)\}$$

Special case: $f(a) = ka$

$$\lambda(k) = \max_a \{ka - I(a)\}$$

Heuristic derivation of Varadhan's result

Gärtner-Ellis Theorem

Scaled cumulant generating function (SCGF)

$$\lambda(k) = \lim_{n\to\infty} \frac{1}{n} \ln\langle e^{nkA_n}\rangle, \qquad k \in \mathbb{R}$$

Theorem: Gärtner (1977), Ellis (1984)

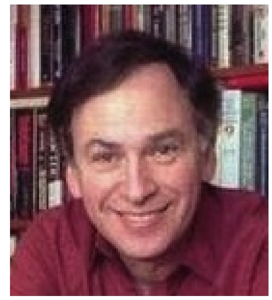
If $\lambda(k)$ is differentiable, then

1. Existence of LDP:

$$p(A_n = a) \approx e^{-nI(a)}$$

2. Rate function:

$$I(a) = \max_k \{ka - \lambda(k)\}$$



- Richard S. Ellis
- UMass, USA

- $I(a)$ = Legendre transform of $\lambda(k)$
- $I(a)$ convex in this case

Exercise: Legendre transforms

[Exercise 2.7.8]

Contraction principle

LDP $\qquad A_n \longrightarrow B_n \qquad$ LDP?

- LDP for A_n:

$$p(A_n = a) \approx e^{-nI_A(a)}$$

- Contraction: $B_n = f(A_n)$
- Probability for B_n:

$$p(B_n = b) = \int_{f^{-1}(b)} p(A_n = a)\, da$$

Contraction principle

- LDP for B_n:

$$p(B_n = b) \approx e^{-nI_B(b)}$$

- Rate function:

$$I_B(b) = \min_{a:f(a)=b} I_A(a) = \min_{f^{-1}(b)} I_A(a)$$

Sums of IID random variables

- Random variable:

$$S_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad X_i \sim p(x), \quad IID$$

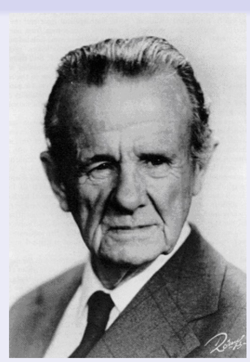
- SCGF:

$$\lambda(k) = \lim_{n\to\infty} \frac{1}{n} \ln\langle e^{nkS_n}\rangle = \lim_{n\to\infty} \frac{1}{n} \ln\left\langle \prod_{i=1}^{n} e^{kX_i}\right\rangle = \ln\langle e^{kX}\rangle$$

Gärtner-Ellis Theorem

$$I(s) = k(s)s - \lambda(k(s)), \qquad \lambda'(k(s)) = s$$

- $I(s)$ is convex
- Zero of $I(s)$ at $\langle X\rangle$
- Originally proved by Cramér (1938)



Example: Gaussian random variables

[Exercise 3.6.1]

$$S_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad p(X_i = x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)}$$

- Generating function:

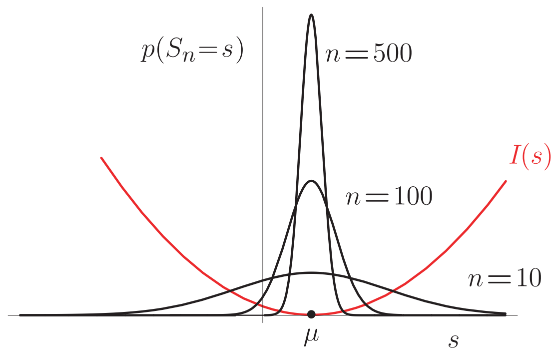
$$\langle e^{kX}\rangle = \ln \int_{-\infty}^{\infty} e^{kx}\,\frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)}\,dx = e^{k\mu+\sigma^2 k^2/2}$$

- Log-generating function:

$$\lambda(k) = \ln\langle e^{kX}\rangle = k\mu + \frac{\sigma^2}{2}k^2$$

- Rate function:

$$I(s) = k(s)s - \lambda(k(s)) = \frac{(s-\mu)^2}{2\sigma^2}$$



Example: Exponential random variables*

[Exercise 3.6.1]

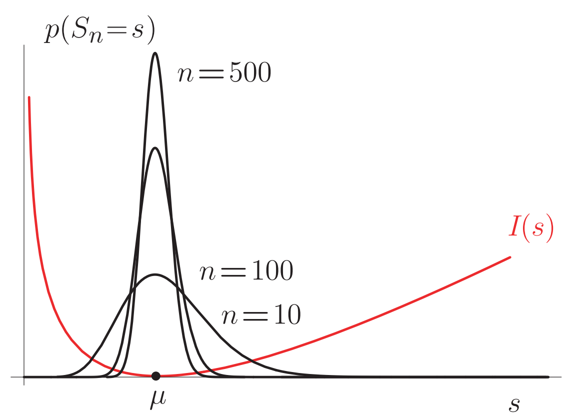
$$S_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad p(X_i = x) = \frac{1}{\mu} e^{-x/\mu}, \quad x > 0$$

- Log-generating function:

$$\lambda(k) = -\ln(1-\mu k), \qquad k < \frac{1}{\mu}$$

- Rate function:

$$I(s) = \frac{s}{\mu} - 1 - \ln\frac{s}{\mu}, \qquad s > 0$$



Example: Bernoulli sample mean

[Exercise 3.6.1]

$$S_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad \begin{array}{l} X_i \in \{0,1\} \\ P(X_i = 1) = \alpha \\ P(X_i = 0) = 1-\alpha \end{array}$$

- Discrete RV:

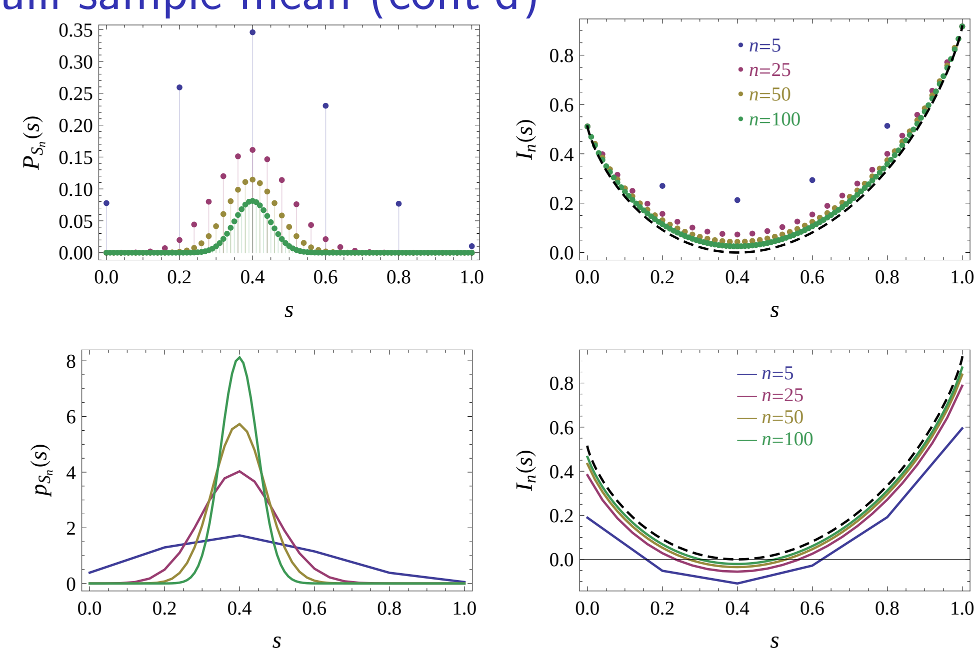
$$S_n \in \left\{0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}, 1\right\}$$

- Discrete probability distribution: $P(S_n = s)$
- Values of S_n "fill" unit interval $[0,1]$ as $n \to \infty$
- Continuous-limit probability density:

$$p(S_n = s) = \frac{P(S_n \in [s, s+\Delta s])}{\Delta s}$$

- Smoothed pdf (weak convergence)

Bernoulli sample mean (cont'd)



- LDP: $p(S_n = s) \approx e^{-nI(s)}$
- Rate function:

$$I(s) = s \ln \frac{s}{\alpha} + (1-s)\ln\frac{1-s}{1-\alpha}$$

Example: Cauchy random variables*

[Exercise 3.6.1]

$$S_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad p(X_i = x) = \frac{1}{\pi}\frac{1}{x^2+1}, \quad x \in \mathbb{R}$$

- Generating function:

$$\lambda(k) = \begin{cases} 0 & k = 0 \\ \infty & k \neq 0 \end{cases}$$

- No large deviations
- $I(s) = 0$
- $P(S_n = s)$ has power-law tails (not exponential)

Sanov's Theorem

[Exercise 3.6.8]

- n IID random variables: $\omega = \omega_1, \omega_2, \ldots, \omega_n$
- Empirical frequencies:

$$L_{n,j} = \frac{1}{n}\sum_{i=1}^{n} \delta_{\omega_i, j} = \frac{\#(\omega_i = j)}{n}$$

- Empirical vector: $\mathbf{L}_n = (L_{n,1}, L_{n,2}, \ldots, L_{n,q})$
- SCGF:

$$\lambda(\mathbf{k}) = \ln \sum_{j=1}^{q} p_j \, e^{k_j}$$

Gärtner-Ellis Theorem

- LDP: $P(\mathbf{L}_n = \boldsymbol{\mu}) \approx e^{-nD(\boldsymbol{\mu})}$
- Rate function: $D(\boldsymbol{\mu}) = \mathbf{k}(\boldsymbol{\mu}) \cdot \boldsymbol{\mu} - \lambda(\mathbf{k}(\boldsymbol{\mu})) = \sum_{j=1}^{q} \mu_j \ln \frac{\mu_j}{p_j}$

Markov processes

Donsker and Varadhan (1975)

- Markov chain:

$$\omega = \omega_1, \omega_2, \ldots, \omega_n, \quad p(\omega) = \rho(\omega_1)\pi(\omega_2|\omega_1)\cdots\pi(\omega_n|\omega_{n-1})$$

- Additive process:

$$S_n = \frac{1}{n}\sum_{i=1}^{n} f(\omega_i)$$

Gätner-Ellis Theorem

- Tilted transition matrix: $\pi_k(\omega_n|\omega_{n-1}) = \pi(\omega_n|\omega_{n-1})e^{kf(\omega_n)}$
- Dominant eigenvalue: $\zeta(\pi_k)$
- SCGF:

$$\lambda(k) = \ln \zeta(\pi_k)$$

- LDP:

$$p(S_n = s) \approx e^{-nI(s)}, \quad I(s) = \max_k\{ks - \lambda(k)\}$$

Exercise: SCGF for Markov chains

Example: Binary Markov chain*

[Exercise 3.6.10]

- Sample mean:

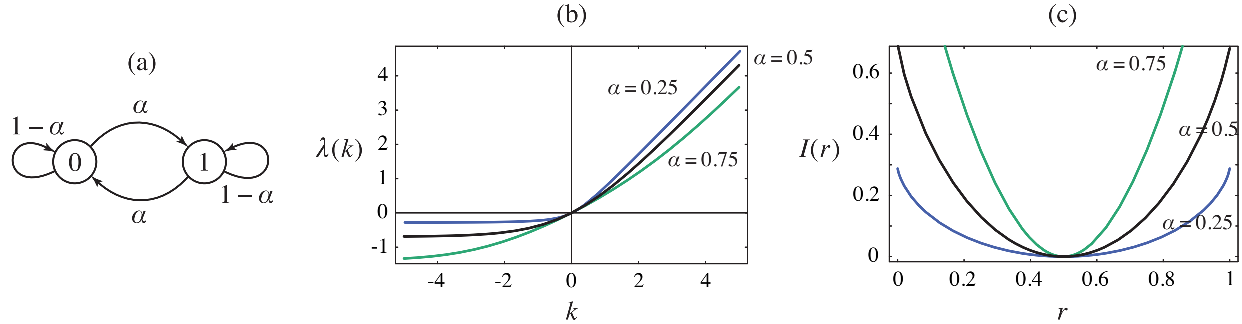
$$R_n = \frac{1}{n}\sum_{i=1}^{n}\omega_i, \qquad \omega_i \in \{0,1\}$$

- Transition matrix:

$$\Pi = \begin{pmatrix} \pi(0|0) & \pi(0|1) \\ \pi(1|0) & \pi(1|1) \end{pmatrix} = \begin{pmatrix} 1-\alpha & \alpha \\ \alpha & 1-\alpha \end{pmatrix}, \qquad \alpha \in (0,1)$$

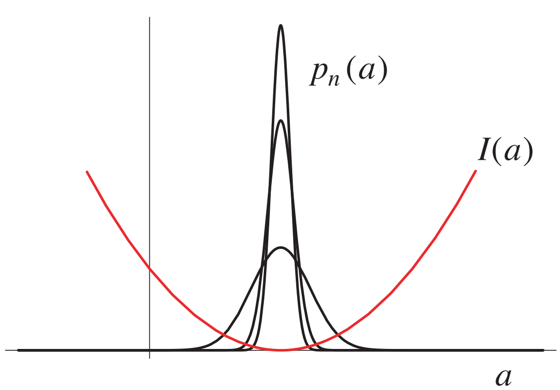
- Rate function:

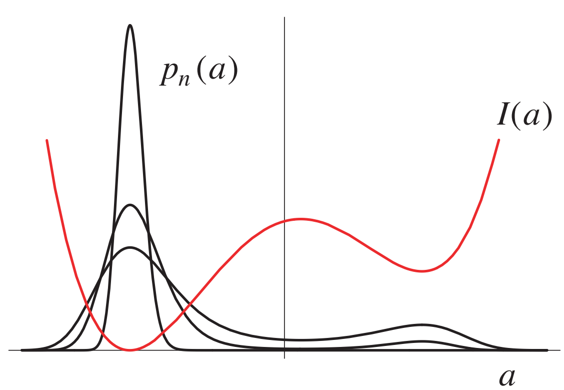
$$I(r) \approx I_0(r) + 2(1-2r)^2\varepsilon + (2 - 32r^2 + 64r^3 - 32r^4)\varepsilon^2, \qquad \varepsilon = 1/2 - \alpha$$



General properties

- Most probable value = min and zero of I
- Zero of I = Law of Large Numbers
- Parabolic minimum = Central Limit Theorem





- $I(a) \neq \max_k\{ka - \lambda(k)\}$ when I is nonconvex

  [Exercise 3.6.2]

Summary

Large deviation principle

$$p(A_n = a) \approx e^{-nI(a)}$$

- Valid for uncorrelated and correlated processes
- Exponential term is dominant
- Describes **small** and **large** fluctuations
- Most probable value: min of $I(a)$
- Law of large numbers: min (zero) of $I(a)$
- Central limit theorem: Parabolic minimum of $I(a)$
- Methods for obtaining $I(a)$:
  - Gärtner-Ellis Theorem
  - Contraction principle

Exercises

- 2.7.1 – 2.7.10
- 3.6.1 – 3.6.10

Continuous-time processes

- Stochastic process: $x(t)$
- Path pdf:

$$\underbrace{p[x]}_{\text{functional}} = p(\{x(t)\}_{t=0}^{T}) = \text{Probability of path } x(t)$$

- Observable: $A_T[x]$
- Observable distribution:

$$p(A_T = a) = \int_{x(t):A_T[x]=a} \mathcal{D}[x]\, p[x]$$

Problems

- Find $p(A_T = a)$
- Find most probable value of A_T
- Determine fluctuations around steady state
- Scaling limits:
  - Long-time: $T \to \infty$
  - Low noise

Low-noise limit of stochastic differential equations

- Dynamical system:

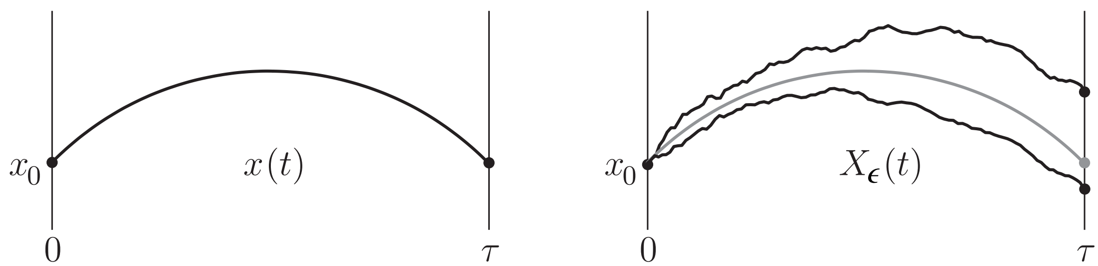
$$\dot{x}(t) = F(x(t))$$

- Perturbed dynamics:

$$\dot{X}(t) = F(X(t)) + \underbrace{\sqrt{\epsilon}\,\xi(t)}_{\text{perturbation}}$$

- Low-noise limit: $\epsilon \to 0$
- Gaussian white noise:

$$\langle \xi(t) \rangle = 0, \quad \langle \xi(t)\xi(t') \rangle = \delta(t - t')$$



Exercise: Simulating SDEs

LDP for the random paths

Functional LDP

$$p[x] \approx e^{-J[x]/\epsilon}, \qquad J[x] = \underbrace{\frac{1}{2}\int_0^T \underbrace{[\dot{x}(t) - F(x(t))]^2}_{\text{Lagrangian}} \, dt}_{\text{Action}}$$

- Low-noise limit: $\epsilon \to 0$
- Derived in maths by Freidlin and Wentzell (1984)
- Derived in physics by Onsager and Machlup (1953)

- Zero of $J[x]$ = most probable dynamics = unperturbed dynamics:

$$\dot{x}^* = F(x^*)$$

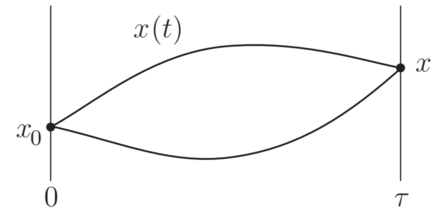
- Gaussian fluctuations around $x^*(t)$

Exercise: Derivation of the action

Other LDPs by contraction

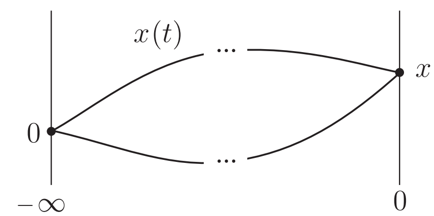
- Transition probability:

$$p(x,T|x_0) \approx e^{-V(x,T|x_0)/\epsilon}$$
$$V(x,T|x_0) = \min_{x(t):x(0)=x_0,x(T)=x} J[x]$$



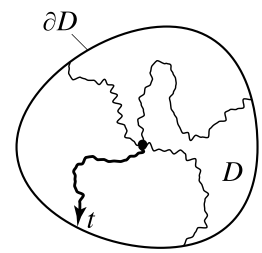
- Stationary distribution:

$$p(x) \approx e^{-V(x)/\epsilon}$$
$$V(x) = \min_{x(t):x(-\infty)=0,x(0)=x} J[x]$$



- Exit time:

$$\tau_\epsilon \approx e^{V^*/\epsilon} \quad \text{in probability}$$
$$V^* = \min_{x\in\partial D} \min_{t\geq 0} V(x,t|x_s)$$



Example: Ornstein-Uhlenbeck process

[Exercise 3.6.14]

- System:

$$\dot{x}(t) = -\gamma x(t) + \sqrt{\epsilon}\,\xi(t)$$

- Stationary distribution:

$$p(x) \approx e^{-V(x)/\epsilon}, \qquad V(x) = \min_{x(t):x(0)=x_0,x(\infty)=x} I[x]$$

- Euler-Lagrange equation:

$$\delta I[x^*] = 0 \quad \Longleftrightarrow \quad \frac{d}{dt}\frac{\partial L}{\partial \dot{x}} - \frac{\partial L}{\partial x} = 0, \quad L = \frac{1}{2}(\dot{x}+\gamma x)^2$$

- Solution: $V(x) = I[x^*] = \gamma x^2$

General result

$$\dot{x} = -U'(x) + \sqrt{\epsilon}\,\xi(t) \quad \Longrightarrow \quad V(x) = 2U(x)$$

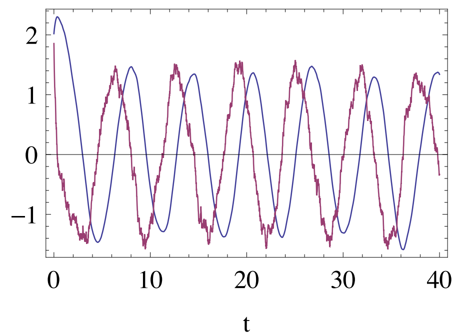
Example: Noisy Van der Pol oscillator*

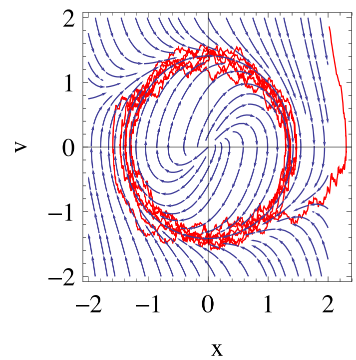
[Exercise 3.6.17]

- Equations of motion:

$$\begin{aligned}\dot{x} &= v \\ \dot{v} &= -x + v(\alpha - x^2 - v^2) + \sqrt{\epsilon}\,\xi(t)\end{aligned}$$

  - Coupled Langevin equations
  - Bifurcation: Stable fixed point ($\alpha < 0$) to stable limit cycle ($\alpha > 0$)



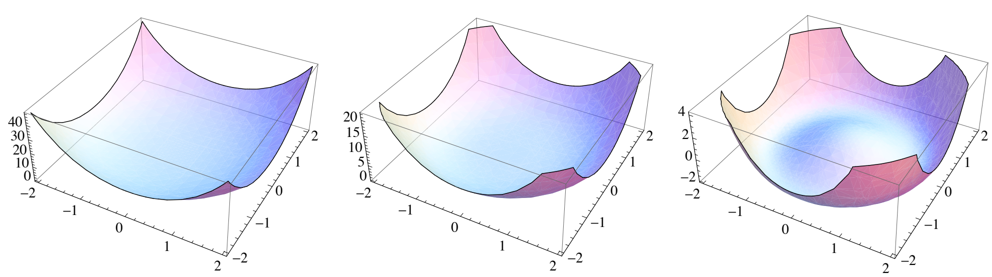


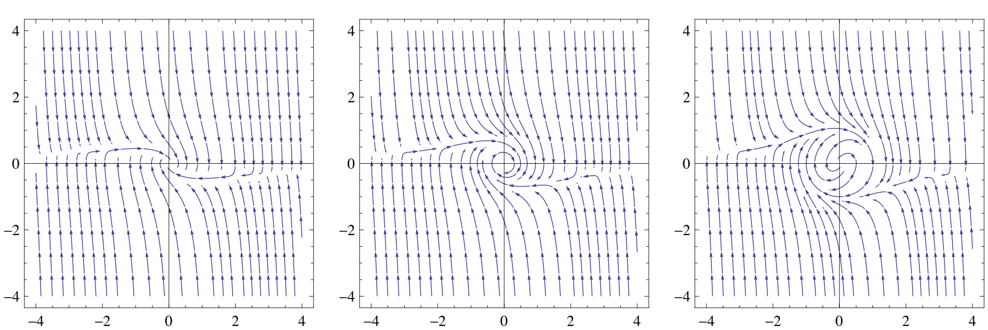
- Stationary distribution: $p(r, \theta) \approx e^{-W(r)/\epsilon}$

Noisy Van der Pol oscillator (cont'd)

- Solution:

$$W(r) = -\alpha r^2 + \frac{r^4}{2}$$





Long-time limit of additive observables

- Stochastic process: $x(t)$
- Observable:

$$A_T[x] = \frac{1}{T}\int_0^T f(x(t))\,dt$$

Gärtner-Ellis

- SCGF:

$$\lambda(k) = \lim_{T\to\infty}\frac{1}{T}\ln\langle e^{TkA_T}\rangle, \qquad \langle e^{TkA_T}\rangle = \int \mathcal{D}[x]\,p[x]\,e^{TkA_T[x]}$$

- Rate function: $I(a) = \max_k\{ka - \lambda(k)\}$

Donsker-Varadhan

- Generator: L
- Tilted generator: $L_k = L + kf$
- SCGF: $\lambda(k) = \zeta(L_k)$

Example: Pulled Brownian particle

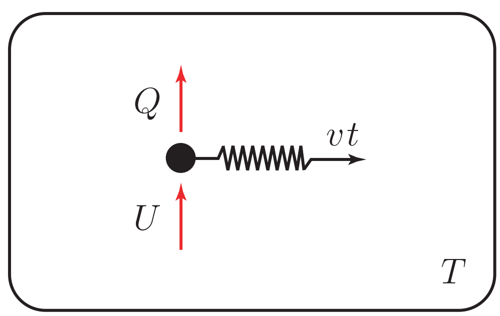
[Exercise 3.6.18]

- Langevin dynamics:

$$m\ddot{x}(t) = \underbrace{-\alpha\dot{x}}_{\text{drag}}\underbrace{-k[x(t)-vt]}_{\text{spring force}}+\underbrace{\xi(t)}_{\text{noise}}$$

- Work:

$$W_T = \frac{1}{T}\int_0^T F(t)\,v\;dt = \Delta U + Q_T$$



Large deviation principle

- SCGF:

$$\lambda(k) = \lim_{T\to\infty}\frac{1}{T}\ln\langle e^{TkW_T}\rangle = ck(1+k), \qquad c = v^2$$

- Rate function:

$$I(w) = \max_k\{kw - \lambda(k)\} = \frac{(w-c)^2}{4c}$$

Summary

Path LDPs

$$p[x] \approx e^{-I[x]/\epsilon} \qquad\qquad p(A_T[x] = a) \approx e^{-TI(a)}$$

- LDPs for time-evolving or steady-state processes
- Describes most probable state (or trajectory)
- Describes fluctuations around most probable states
- Most probable state = zero of I = min of I
- Most probable state given by variational principle

Connection with classical mechanics

$$\delta I[x^*] = 0 \quad \Longrightarrow \quad \begin{cases} \text{Euler-Lagrange equation} \\ \text{Hamiltonian equation} \\ \text{Hamiltonian-Jacobi equation} \end{cases}$$

Exercises

- 3.6.11 – 3.6.18

Other applications

- Equilibrium statistical mechanics
- Multifractals
- Chaotic systems (thermodynamic formalism)
- Disorded systems
  - Random walks in random environments
  - Spin glasses
  - Quenched/annealed large deviations
- Nonequilibrium systems
- Interacting particle models
  - Zero-range process
  - Exclusion process
  - Current, density profile
  - Fluctuation relations
  - Space/time large deviations

Equilibrium many-particle systems

- N particles
- Microstate: $\omega = \omega_1, \omega_2, \ldots, \omega_N$
- Space of one particle: $\omega_i \in \Lambda$
- Space of N particles: $\Lambda_N = \Lambda^N$
- Probability distribution on Λ_N: $P(\omega)$
- Macrostate: $M_N(\omega)$
- Probability distribution for M_N:

$$P(M_N = m) = \sum_{\omega : M_N(\omega) = m} P(\omega)$$

Problems

- Find $P(M_N = m)$
- Find most probable values of M_N (= equilibrium states)
- Study fluctuations around most probable value
- Consider thermodynamic limit $N \to \infty$

Thermodynamic LDPs

Ensembles

Microcanonical	Canonical
$P_u(M_N = m) \approx e^{-NI_u(m)}$	$P_\beta(M_N = m) \approx e^{-NI_\beta(m)}$

- Generalize and refine Einstein's theory of fluctuations
- Equilibrium and fluctuation properties given by rate function
- Equilibrium states = min of $I(m)$ = zero of $I(m)$
- Equilibrium states given by variational principles

Entropy

$$P(U_N = u) \approx e^{Ns(u)}, \qquad s(u) = \min_\beta\{\beta u - \varphi(\beta)\}$$

- Legendre transform of thermo = Legendre transform of LDT
- Legendre transform is valid only if $s(u)$ is concave

Problem

- Sequence of RVs: $\omega = X_1, X_2, \dots, X_n$
- Observable (RV): $S_n(\omega)$
- Probability density:

$$p(S_n = s) = \frac{P(S_n \in [s, s+\Delta s])}{\Delta s} = \frac{P(S_n \in \Delta_s)}{\Delta s}$$

Goals

1. Sample S_n
2. Estimate $p(S_n = s)$
3. Test existence of LDP
4. Estimate rate function $I(s)$

$n \to \infty$
$\Delta s \to 0$

Continuous-time problem: Time discretization

$$\{X(t)\}_{t=0}^{T} \longrightarrow \{X_i\}_{i=1}^{n}, \quad X_i = X(i\Delta t)$$

Direct estimators

- Configuration sample:

$$\{\omega^{(1)}, \omega^{(2)}, \dots, \omega^{(L)}\}$$

  - L copies or realizations of ω
  - Prior pdf: $p(\omega)$

- Observable sample:

$$\{s^{(1)}, s^{(2)}, \dots, s^{(L)}\}, \qquad s^{(j)} = S_n(\omega^{(j)})$$

- Estimator of $p(S_n = s)$:

$$\hat{p}(s) = \frac{\hat{P}(\Delta_s)}{\Delta s} = \frac{1}{L\Delta s}\sum_{j=1}^{L} \mathbf{1}_{\Delta_s}(s^{(j)}).$$

  - Empirical pdf
  - Unbiased estimator [Exercise 4.7.3]

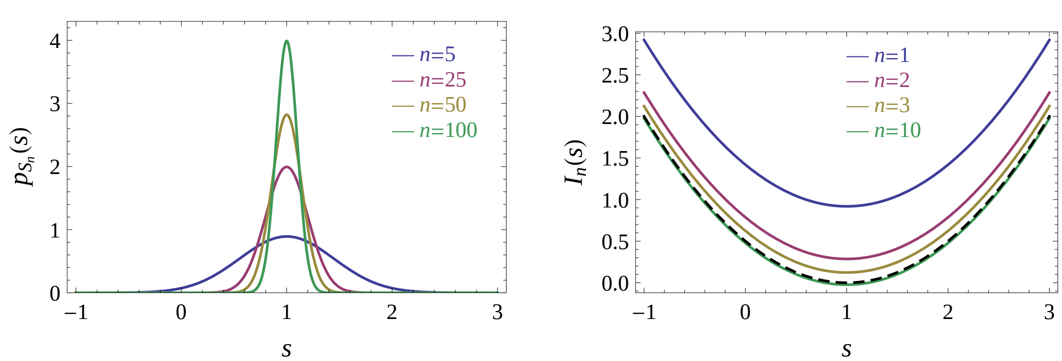
Exercise: What is an empirical pdf?

Direct estimators (cont'd)

- Estimator of rate function:

$$\hat{I}(s) = -\frac{1}{n} \ln \hat{p}(s)$$

1. Estimate $\hat{p}(s)$ for fixed L and n
2. Estimate $\hat{I}(s)$ for fixed L and n
3. Repeat for increasing values of n



Use large enough L to get good statistics

Problem with direct sampling

- $p(S_n = s) \approx e^{-nI(s)}$
- Event $S_n = s$ (or $S_n \in \Delta_s$) is exponentially rare
- Choose $L \sim e^{nI(s)}$ to see this event
- L exponential with n

Solution: Importance sampling

- Sample ω with another pdf $q(\omega)$
- Choose $q(\omega)$ to make $S_n = s$ probable
- Correct estimation of $p(S_n = s)$ according to $q(\omega)$ chosen

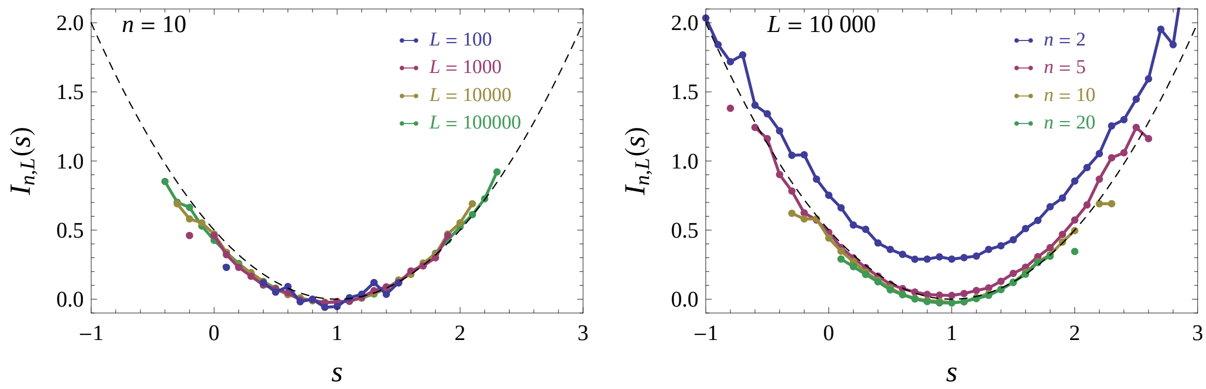
Example: Gaussian sample mean

[Exercise 4.7.2]

$$S_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad p(X_i = x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)}$$

1. Generate $x_1, x_2, \ldots, x_n \sim$ Gaussian variates
2. Compute S_n
3. Repeat L times to obtain sample $\{s^{(1)}, s^{(2)}, \ldots, s^{(L)}\}$
4. Compute $\hat{p}(s)$ of sample
5. Compute $\hat{I}(s)$
6. Repeat for different n and L

Gaussian sample mean (cont'd)



- Increase L to sample tails
- Increase L to smooth results (smaller error bars) [Exercise 4.7.4] [Exercise 4.7.5]
- Increase L for increasing n
- Choose $L \sim e^n$

Importance sampling

- Original sampling pdf: $p(\omega)$
- New sampling pdf: $q(\omega)$

New estimator

$$\hat{q}(s) = \frac{1}{L\Delta s} \sum_{j=1}^{L} \mathbf{1}_{\Delta_s}(S_n(\omega^{(j)}))\, R(\omega^{(j)})$$

- Importance sampling or likelihood ratio: $R(\omega) = \dfrac{p(\omega)}{q(\omega)}$

- $\hat{q}(s)$ is unbiased:

$$\langle \hat{q}(s) \rangle_q = \langle \hat{p}(s) \rangle_p = p(S_n = s)$$

- $\hat{q}(s)$ may have smaller variance than $\hat{p}(s)$

Choose q such that $\mathrm{var}_q(\hat{q}) < \mathrm{var}_p(\hat{p})$

Exercise: Derivation of new estimator

Exponential change of measure

- Original sampling pdf: $p(\omega)$
- Exponentially tilted pdf:

$$p_k(\omega) = \frac{e^{nkS_n(\omega)}}{W_n(k)} p(\omega), \qquad k \in \mathbb{R}$$

- Generating function:

$$W_n(k) = \langle e^{nkS_n} \rangle_p = \int e^{nkS_n(\omega)} \, p(\omega) \, d\omega$$

- Likelihood ratio:

$$R(\omega) = \frac{p(\omega)}{p_k(\omega)} = e^{-nkS_n(\omega)} \, W_n(k) \approx e^{-n[kS_n(\omega) - \lambda(k)]}$$

- $p_k(\omega)$ is efficient
- Zero-variance estimator as $n \to \infty$
- How to choose k?

Example: Gaussian sample mean

[Exercise 4.7.9]

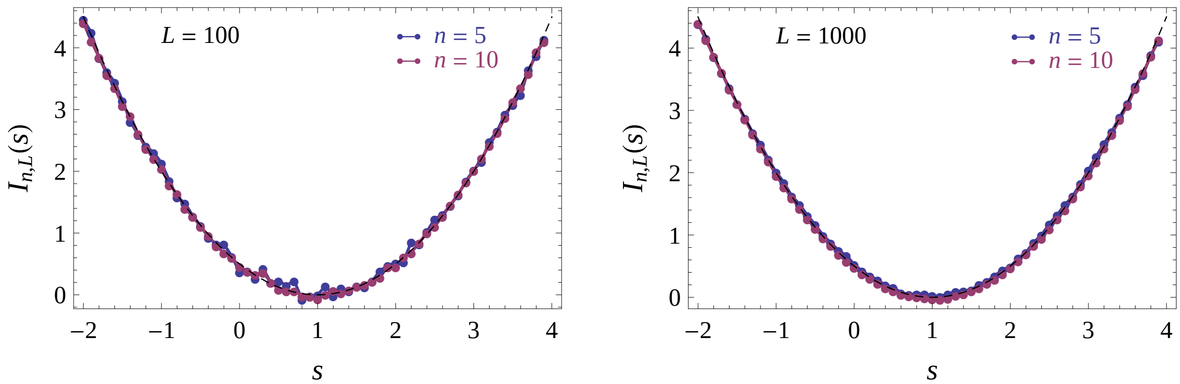
$$S_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad p(\omega) = \prod_{i=1}^{m} p(X_i), \quad p(X_i = x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)}$$

- Choose $k \in \mathbb{R}$
- Generate variate $x_1, x_2, \ldots, x_n$ according to tilted pdf:

$$p_k(x_i) = \frac{e^{kx_i} p(x_i)}{W(k)} = \frac{e^{-(x_i-\mu-\sigma^2 k)^2/(2\sigma^2)}}{\sqrt{2\pi\sigma^2}}$$

- Compute S_n
- Repeat L times to obtain sample $\{s^{(1)}, s^{(2)}, \ldots, s^{(L)}\}$
- Compute $\hat{q}(s)$
- Compute $\hat{I}(s)$
- Repeat for $k \in [k_{\min}, k_{\max}]$

Gaussian sample mean (cont'd)



- p_k is efficient to recover $I(s)$ at $s = \lambda'(k)$
- Scan $k \in [k_{\min}, k_{\max}]$ to obtain desired range $s \in [s_{\min}, s_{\max}]$
- Non-parametric: choose k such that $s = \lambda'(k)$ to obtain I at s

Gaussian sample mean (cont'd)

- SCGF:

$$\lambda(k) = \mu k + \frac{1}{2}\sigma^2 k^2$$

- Concentration point:

$$\lambda'(k) = \mu + \sigma^2 k = s \quad \Rightarrow \quad k(s) = \frac{s-\mu}{\sigma^2}$$

- Explicit form of tilted pdf:

$$p_{k(s)}(x_i) = \frac{e^{-(x_i-s)^2/(2\sigma^2)}}{\sqrt{2\pi\sigma^2}}$$

Importance sampling interpretation

$$S_n = s \text{ large deviation under } p(\omega)$$
$$\downarrow$$
$$S_n = s \text{ typical event under } p_{k(s)}(\omega)$$

Metropolis (Monte Carlo) sampling

Problem

- Tilted pdf p_k involves $W_n(k)$
- Tilted pdf p_k assumes knowledge of SCGF $\lambda(k)$
- Knowledge of SCGF $\Rightarrow I(s)$

Solutions

1. Use other types of estimators
2. Use Metropolis (Monte Carlo) sampling with p_k
   - Based on $p_k(\omega)/p_k(\omega')$
   - Free of $W_n(k)$ and $\lambda(k)$

[Exercise 4.7.11]

Exercise: Illustration of Metropolis sampling

Application to SDEs

Path pdf form

- Original path pdf: $p[x]$
- Tilted pdf:

$$p_k[x] = \frac{e^{TkS_T[x]}\, p[x]}{W_T(k)} \approx e^{-TI_k[x]}, \qquad W_T(k) = \langle e^{TkS_T} \rangle_p$$

Transition probability form

- Original transition matrix: $\Pi(\Delta t)$
- Tilted transition matrix:

$$\Pi_k(\Delta t) = \frac{e^{k\Delta t x'}\Pi(\Delta t)}{e^{\lambda(k)\Delta t}} = e^{[kx' + G - \lambda(k)]\Delta t}$$

Generator form

- Original generator: G
- Tilted generator: $G_k = G + kx' - \lambda(k)$

Example: Gaussian additive process

- SDE: $\dot{x}(t) = \xi(t)$
- Pure Brownian motion without drift
- Observable:

$$D_T[x] = \frac{1}{T}\int_0^T \dot{x}(t)\,dt = \frac{x(T)}{T}$$

Intuitive observations

- Typical state: $D_T = 0$
- Modified process with typical state $D_T = d$:

$$\dot{x}(t) = d + \xi(t)$$

- Effective dynamics for large deviation

- Explicit result:

$$I_{k(d)}[x] = I[x] - dD_T[x] + \frac{d^2}{2} = \frac{1}{2T}\int_0^T (\dot{x} - d)^2\,dt,$$

Exercise: Simulation of SDEs

Summary

- Direct sampling of S_n with $p(\omega)$ inefficient
- Requires exponential sample size L
- Change sampling distribution to $q(\omega)$
- Make rare event under p more probable under q
- Possible change of measure: Exponential measure
- Exponential sampling efficient
- Subexponential sample size
- Structure of exponential measure = structure of LDT
- Combine exponential sampling with Metropolis sampling
- Other methods?

Sample mean method

Sample $\lambda(k)$ instead of $p(S_n = s)$

- Sampling with $p(\omega)$:

$$S_n \to \lambda'(0) \quad \text{in probability}$$

- Sampling with $p_k(\omega)$:

$$S_n \to \lambda'(k) \quad \text{in probability}$$

- Estimator of S_n:

$$\hat{s}(k) = \frac{1}{L}\sum_{j=1}^{L} S_n(\omega^{(j)})$$

- Estimator of $\lambda(k)$:

$$\hat{\lambda}(k) = \int_0^k \hat{s}(k')\,dk'$$

- Compute $\hat{I}(s)$ by Legendre transform
- No need to estimate $\hat{p}$

Example: Gaussian sample mean

[Exercise 5.4.1]

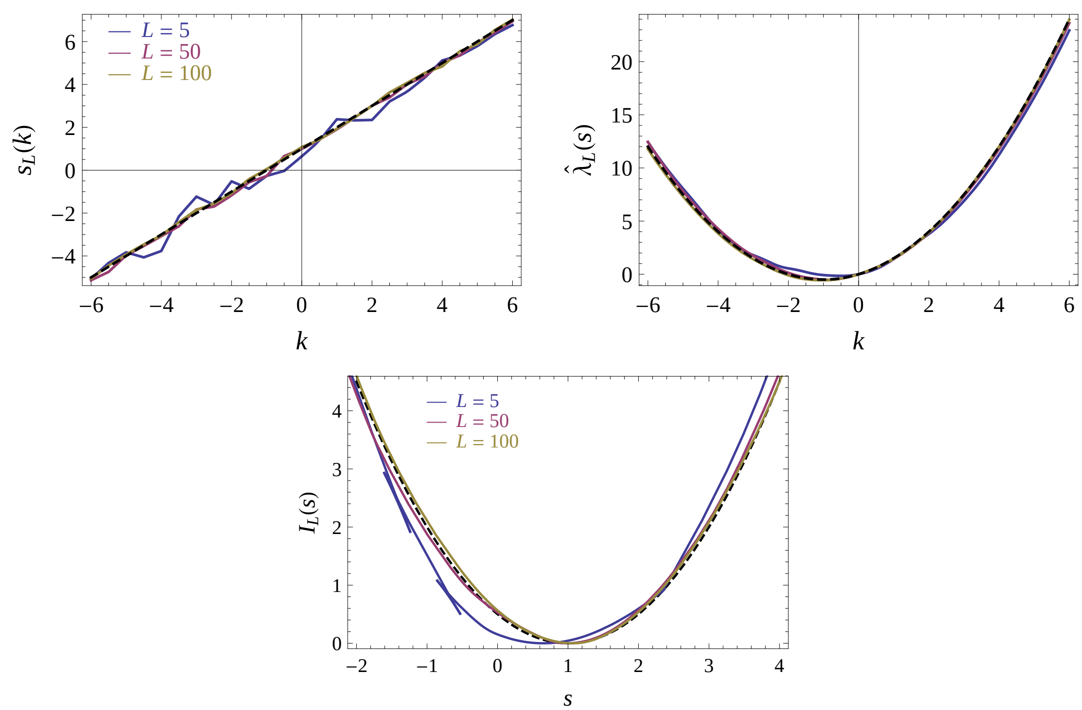
$$S_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad p(\omega) = \prod_{i=1}^{m} p(X_i), \quad p(X_i = x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)}$$

- Choose $k \in \mathbb{R}$
- Generate variates $x_1, x_2, \ldots, x_L$ according to tilted pdf:

$$p_k(x_i) = \frac{e^{kx_i} p(x_i)}{W(k)} = \frac{e^{-(x_i-\mu-\sigma^2 k)^2/(2\sigma^2)}}{\sqrt{2\pi\sigma^2}}$$

- Compute $\hat{s}$ for L large
- Repeat for different k
- Integrate results to obtain $\hat{\lambda}(k)$
- Obtain $\hat{I}$ by Legendre transform

Gaussian sample mean (cont'd)



- No empirical pdf calculated
- Non-convex artefacts for small L ($\hat{\lambda}(k)$ not convex)

Empirical generating function

- IID sample mean: $\lambda(k) = E[e^{kX}]$
- Estimator for $\lambda(k)$:

$$\hat{\lambda}(k) = \ln \frac{1}{L} \sum_{j=1}^{L} e^{kX^{(j)}}$$

- Alternative estimator:

$$\hat{s}(k) = \hat{\lambda}'(k) = \frac{\sum_{j=1}^{L} X^{(j)}\, e^{kX^{(j)}}}{\sum_{j=1}^{L} e^{kX^{(j)}}}$$

- Markov chain:

$$\underbrace{X_1 + \cdots + X_b}_{Y_1} + \underbrace{X_{b+1} + \cdots + X_{2b}}_{Y_2} + \cdots + \underbrace{X_{n-b+1} + \cdots + X_n}_{Y_m}$$

- Markov estimator:

$$\hat{\lambda}(k) = \frac{1}{b} \ln \frac{1}{L} \sum_{j=1}^{L} e^{kY^{(j)}}, \qquad m = \frac{n}{b}$$

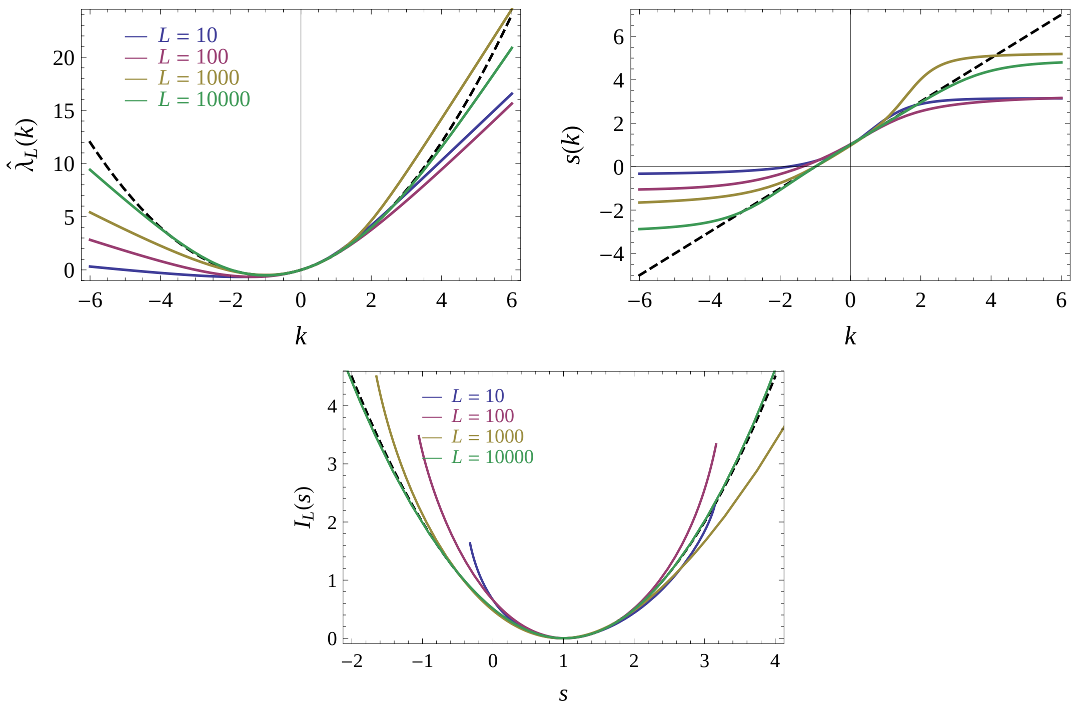
Example: Gaussian sample mean

[Exercise 5.4.2]

$$S_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad p(\omega) = \prod_{i=1}^{m} p(X_i), \quad p(X_i = x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)}$$

- Choose $k \in \mathbb{R}$
- Generate variates $x_1, x_2, \ldots, x_L$ according to original pdf $p(x_i)$
- Compute $\hat{\lambda}(k)$ for L large
- Repeat for different k
- Obtain $\hat{I}$ by Legendre transform
- Repeat for larger L

Gaussian sample mean (cont'd)



- Efficient for RVs with bounded support [Exercise 5.4.2]
- Less efficient for unbounded RVs

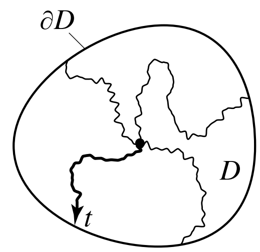
Application for SDEs*

[Exercise 5.4.3]

Other methods

- Optimal path methods

$$p(x, T|x_0) \approx e^{-V(x,T|x_0)/\epsilon}$$
$$V(x, T|x_0) = \min_{x(t):x(0)=x_0, x(T)=x} J[x]$$



  - [Exercise 5.4.4 – 5.4.6]
- Transition path method
  - Monte Carlo for paths
  - See Christoph Dellago
- Splitting / cloning methods
  - See notes for references
- Eigenvalue method

$$\lambda(k) = \text{dominant eigenvalue of tilted generator } G_k$$

  - See notes for references

Reading

H. Touchette
The large deviation approach to statistical mechanics
Physics Reports **478**, 1-69, 2009
arxiv:0804.0327

Y. Oono
Large deviation and statistical physics
Prog. Theoret. Phys. Suppl. **99**, 165-205, 1989

A. Dembo and O. Zeitouni
Large Deviations Techniques and Applications
Springer, New York, 1998

J. A. Bucklew
Introduction to Rare Event Simulation
Springer, New York, 2004