Analysis of an Interacting Particle Scheme for rare event estimation

Paul Dupuis
Division of Applied Mathematics
Brown University
(Yi Cai, Brown)

RESIM/Newton Institute

June, 2010
Outline

- Problem formulation
Outline

- Problem formulation
- Construction of IPS estimator
Outline

- Problem formulation
- Construction of IPS estimator
- Performance measure–large deviation asymptotics of second moment
Outline

- Problem formulation
- Construction of IPS estimator
- Performance measure–large deviation asymptotics of second moment
- Sources of difficulty
Outline

- Problem formulation
- Construction of IPS estimator
- Performance measure—large deviation asymptotics of second moment
- Sources of difficulty
- Proper perspective—occupation measure indexed by thresholds
Outline

- Problem formulation
- Construction of IPS estimator
- Performance measure—large deviation asymptotics of second moment
- Sources of difficulty
- Proper perspective—occupation measure indexed by thresholds
- One-dimensional problem

Main results:
- Large deviation limit for second moment with fixed number of particles
- Asymptotics of decay rate as number of particles grows

Remarks on higher dimensions
Problem formulation
Construction of IPS estimator
Performance measure–large deviation asymptotics of second moment
Sources of difficulty
Proper perspective–occupation measure indexed by thresholds
One-dimensional problem
Relative entropy representations
Outline

- Problem formulation
- Construction of IPS estimator
- Performance measure—large deviation asymptotics of second moment
- Sources of difficulty
- Proper perspective—occupation measure indexed by thresholds
- One-dimensional problem
- Relative entropy representations
- Main results:
Problem formulation
Construction of IPS estimator
Performance measure—large deviation asymptotics of second moment
Sources of difficulty
Proper perspective—occupation measure indexed by thresholds
One-dimensional problem
Relative entropy representations
Main results:
  - Large deviation limit for second moment with fixed number of particles
Outline

- Problem formulation
- Construction of IPS estimator
- Performance measure—large deviation asymptotics of second moment
- Sources of difficulty
- Proper perspective—occupation measure indexed by thresholds
- One-dimensional problem
- Relative entropy representations
- Main results:
  - Large deviation limit for second moment with fixed number of particles
  - Asymptotics of decay rate as number of particles grows
Outline

- Problem formulation
- Construction of IPS estimator
- Performance measure—large deviation asymptotics of second moment
- Sources of difficulty
- Proper perspective—occupation measure indexed by thresholds
- One-dimensional problem
- Relative entropy representations
- Main results:
  - Large deviation limit for second moment with fixed number of particles
  - Asymptotics of decay rate as number of particles grows
- Remarks on higher dimensions
Problem formulation

We consider a sequence of Markov processes
\( \{X^m_i, i = 1, 2, \ldots, m = 1, 2, \ldots\} \) for which a large deviation result holds.

Loosely speaking, for \( T < 1; \alpha > 0 \) small, continuous trajectory and large \( m \),
\[
\Pr(\sup_{0 \leq t \leq mT} |X^m_t| = x = (0)) \approx e^{-m \int_0^T L(x, \_ \alpha) \, dt}
\]

Typically \( m \) scales the system size, e.g., buffer size.
We consider a sequence of Markov processes \( \{X^m_i, i = 1, 2, \ldots, m = 1, 2, \ldots\} \) for which a large deviation result holds. Loosely speaking, for \( T < \infty, \delta > 0 \) small, continuous trajectory \( \phi \) and large \( m \),

\[
P \left\{ \sup_{0 \leq i \leq mT} |X^m_i - \phi(i/m)| \leq \delta \bigg| X^m_i = x = \phi(0) \right\} \approx e^{-mI_T(\phi)},
\]

for a rate function of the standard form

\[
I_T(\phi) = \int_0^T L(\phi, \dot{\phi}) dt.
\]
Problem formulation

We consider a sequence of Markov processes \( \{X^m_i, i = 1, 2, \ldots, m = 1, 2, \ldots\} \) for which a large deviation result holds. Loosely speaking, for \( T < \infty, \delta > 0 \) small, continuous trajectory \( \phi \) and large \( m \),

\[
P\left\{ \sup_{0 \leq i \leq mT} |X^m_i - \phi(i/m)| \leq \delta \left| X^m_i = x = \phi(0) \right. \right\} \approx e^{-mI_T(\phi)},
\]

for a rate function of the standard form

\[
I_T(\phi) = \int_0^T L(\phi, \dot{\phi}) dt.
\]

[Typically \( m \) scales the system size, e.g., buffer size.]
A is a stable set and entry into $B$ is rare.
Problem formulation (cont’d)

A is a stable set and entry into $B$ is rare.

Let $C = \{ \text{trajectories that hit } B \text{ prior to } A \}$. To estimate:

$$p_m(x) = P \{ X^m \in C \mid X_0^m = x \}.$$
Problem formulation (cont’d)

Under mild conditions:

\[-\frac{1}{m} \log p_m(x) \to \inf \{ I_T(\phi) : \phi \text{ enters } B \text{ prior to } A \text{ before } T, T < \infty \} = \gamma(x)\]
Due to Del Moral et. al. A number [proportional to $m$] of splitting thresholds $C_k^m$ are defined,

and $N$ particles are started at $x$. 
Construction of IPS estimator (cont’d)

Construction is recursive with respect to threshold level, and divides into transportation step and resampling step. Transportation to $k$th level. Using transition law of $X_i^m$:

Diagram:
- Particles needing transportation
- Particles already in $C_k^m$
Construction of IPS estimator (cont’d)

Resampling at $k$th level. Suppose $N_k^m$ particles reach $C_k^m$ after transportation, *including those already there*. Sample from uniform distribution on location of these $N_k^m$ particles to yield $N$ particles:
Construction of IPS estimator (cont’d)

Since each particle is distributed according to first entrance distribution into threshold $k$, $N^m_k/N$ is an unbiased estimate of

$$P \{ X^m \text{ reaches threshold } k \mid X^m \text{ reached threshold } k - 1 \}.$$
Since each particle is distributed according to first entrance distribution into threshold $k$, $N_k^m / N$ is an unbiased estimate of

$$P \{ X^m \text{ reaches threshold } k \mid X^m \text{ reached threshold } k - 1 \}.$$

If $x \in C_0^m$,\[\theta^m = \prod_{k=1}^{K^m} \frac{N_k^m}{N}\]
is an unbiased estimate of

$$p_m(x) = \prod_{k=1}^{K^m} P \{ X^m \text{ reaches threshold } k \mid X^m \text{ reached threshold } k - 1 \}.$$
Since each particle is distributed according to first entrance distribution into threshold $k$, $N^m_k / N$ is an unbiased estimate of

$$ P \{ X^m \text{ reaches threshold } k \mid X^m \text{ reached threshold } k - 1 \}.$$ 

If $x \in C^m_0$,

$$ \theta^m = \prod_{k=1}^{K^m} \frac{N^m_k}{N} $$

is an unbiased estimate of

$$ p_m(x) = \prod_{k=1}^{K^m} P \{ X^m \text{ reaches threshold } k \mid X^m \text{ reached threshold } k - 1 \} $$

We then average independent copies of $\theta^m$ for the final estimate.
Construction of IPS estimator (cont’d)

Performance measure–large deviation asymptotics of second moment

Since unbiased bounds on variance follow from bounds on second moment.

Well-known use of Jensen’s inequality gives

$$\limsup_{m \to 1} \frac{1}{m} \log \mathbb{E}(m)^2 \leq (x)^2,$$

i.e., best possible decay rate of $e^{2m(x)}$, keeping variance proportional to quantity estimated. Goal of design is high (hopefully $2(x)$) decay rate.
Performance measure–large deviation asymptotics of second moment

Since unbiased bounds on variance follow from bounds on second moment. Well-known use of Jensen’s inequality gives

$$\lim sup_{m \to \infty} -\frac{1}{m} \log E (\theta^m)^2 \leq 2\gamma(x)$$

i.e., best possible decay rate of $e^{-2m\gamma(x)}$, keeping variance proportional to quantity estimated. Goal of design is high (hopefully $2\gamma(x)$) decay rate.
Sources of difficulty

1. Natural time scale for process is not the natural time scale for the algorithm (threshold hitting times).

2. Analyses of other schemes based on branching exploit independence of trajectories after splitting. Here resampling recouples particles at every threshold.

3. Cannot directly use any large deviation properties of the original process under "small noise" scaling.

4. Estimates should be valid all the way from few particles (large deviation effects dominate) to many particles (law of large numbers effects dominate).
Sources of difficulty

1. Natural time scale for process is not the natural time scale for the algorithm (threshold hitting times).

2. Analyses of other schemes based on branching exploit independence of trajectories after splitting. Here resampling recouples particles at every threshold.
Sources of difficulty

1. Natural time scale for process is not the natural time scale for the algorithm (threshold hitting times).

2. Analyses of other schemes based on branching exploit independence of trajectories after splitting. Here resampling recouples particles at every threshold.

3. Cannot directly use any large deviation properties of the original process under “small noise” scaling.
Sources of difficulty

1. Natural time scale for process is not the natural time scale for the algorithm (threshold hitting times).

2. Analyses of other schemes based on branching exploit independence of trajectories after splitting. Here resampling recouples particles at every threshold.

3. Cannot directly use any large deviation properties of the original process under “small noise” scaling.

4. Estimates should be valid all the way from few particles (large deviation effects dominate) to many particles (law of large numbers effects dominate).
Proper perspective—occupation measure indexed by thresholds

Alternative expression for second moment

\[ E \left( \theta^m \right)^2 = E \prod_{k=1}^{K^m} \left( \frac{N_k^m}{N} \right)^2 = E e^{2 \sum_{k=1}^{K^m} \log \left( \frac{N_k^m}{N} \right)} = E e^{K^m \left[ \frac{1}{K^m} \sum_{k=1}^{K^m} 2 \log \left( \frac{N_k^m}{N} \right) \right]} \]
Proper perspective—occupation measure indexed by thresholds

Alternative expression for second moment

\[ E (\theta^m)^2 = E \prod_{k=1}^{K^m} \left( \frac{N_k^m}{N} \right)^2 = E e^{2 \sum_{k=1}^{K^m} \log \left( \frac{N_k^m}{N} \right)} = E e^{K^m \left[ \frac{1}{K^m} \sum_{k=1}^{K^m} 2 \log \left( \frac{N_k^m}{N} \right) \right]}. \]

Can be approximated if we know large deviation asymptotics for empirical measure

\[ L^m(dn) = \frac{1}{K^m} \sum_{k=1}^{K^m} \delta_{N_k^m}(dn). \]
One-dimensional problem

Most difficulties of analysis already present in one-dimensional iid random walk:

\[ X_{i+1}^m = X_i^m + \frac{1}{m} \theta_i, \quad \theta_i \text{ has distribution } \mu, \quad \int y \mu(dy) < 0, \]

and with

\[ A = (-\infty, 0], \quad B = [1, \infty), \quad X_0^m = x = \frac{1}{m}. \]
One-dimensional problem

Most difficulties of analysis already present in one-dimensional iid random walk:

\[ X_{i+1}^m = X_i^m + \frac{1}{m} \theta_i, \quad \theta_i \text{ has distribution } \mu, \quad \int y \mu(dy) < 0, \]

and with

\[ A = (-\infty, 0], \quad B = [1, \infty), \quad X_0^m = x = \frac{1}{m}. \]

Assume

\[ \text{supp } (\mu) \text{ is bounded, } \mu([0, \infty)) > 0, \quad C_k^m = \frac{k}{m}. \]
One-dimensional problem

Most difficulties of analysis already present in one-dimensional iid random walk:

\[ X_{i+1}^m = X_i^m + \frac{1}{m} \theta_i, \quad \theta_i \text{ has distribution } \mu, \quad \int y \mu(dy) < 0, \]

and with

\[ A = (-\infty, 0], \quad B = [1, \infty), \quad X_0^m = x = \frac{1}{m}. \]

Assume

\[ \text{supp}(\mu) \text{ is bounded, } \mu([0, \infty)) > 0, \quad C_k^m = \frac{k}{m}. \]

Note no particular relation between \( \mu \) and \( C_k^m \) is assumed.
Notation for the IPS. Let

\[ k \in \{1, 2, \ldots, m\} \text{ index thresholds} \]

\[ s \in \{1, \ldots, N\} \text{ index particles.} \]
One-dimensional problem (cont’d)

Notation for the IPS. Let

\[ k \in \{1, 2, \ldots, m\} \text{ index thresholds} \]

\[ s \in \{1, \ldots, N\} \text{ index particles.} \]

We use the recursive notation

\[ Z^m_k \xrightarrow{\text{transportation}} Y^m_{k+1} \xrightarrow{\text{resampling}} Z^m_{k+1}. \]
Notation for the IPS. Let

\[ k \in \{1, 2, \ldots, m\} \quad \text{index thresholds} \]

\[ s \in \{1, \ldots, N\} \quad \text{index particles.} \]

We use the recursive notation

\[ Z^m_{k,s} \xrightarrow{\text{transportation}} Y^m_{k+1,s} \xrightarrow{\text{resampling}} Z^m_{k+1,s}. \]

\[ Z^m_{k,s} \] is one of \( N \) particles located within threshold \( C^m_k \).
Notation for the IPS. Let

\[ k \in \{1, 2, \ldots, m\} \] index thresholds

\[ s \in \{1, \ldots, N\} \] index particles.

We use the recursive notation

\[ Z^m_k \xrightarrow{\text{transportation}} Y^m_{k+1} \xrightarrow{\text{resampling}} Z^m_{k+1}. \]

- \( Z^m_k \) is one of \( N \) particles located within threshold \( C^m_k \).
- \( Y^m_{k+1,s} \) is the location when stopped on first hitting \( (-\infty, 0] \cup C^m_{k+1} \).
Notation for the IPS. Let

\[ k \in \{1, 2, \ldots, m\} \text{ index thresholds} \]
\[ s \in \{1, \ldots, N\} \text{ index particles.} \]

We use the recursive notation

\[
\begin{align*}
Z_k^m & \xrightarrow{\text{transportation}} Y_{k+1}^m & \xrightarrow{\text{resampling}} Z_{k+1}^m.
\end{align*}
\]

- \( Z_{k,s}^m \) is one of \( N \) particles located within threshold \( C_k^m \).
- \( Y_{k+1,s}^m \) is the location when stopped on first hitting \((-\infty, 0] \cup C_{k+1}^m\).
- Let \( \Delta \) denote absorbing state/set \((-\infty, 0]\).
Let
\[
I_k^m = \sum_{s=1}^{N} C_k^m \setminus C_{k+1}^m (Z_k^m, s)
\]
\[
= \text{number particles needing transportation at stage } k
\]
\[
N_{k+1}^m = \sum_{s=1}^{N} 1 C_{k+1}^m (Y_{k+1, s}^m)
\]
\[
= \text{number particles make it to next threshold at stage } k
\]
One-dimensional problem (cont’d)

Let

\[ I_k^m = \sum_{s=1}^{N} \mathbb{1}_{C_k^m \setminus C_{k+1}^m} (Z_k^m, s) \]

\[ = \text{number particles needing transportation at stage } k \]

\[ N_{k+1}^m = \sum_{s=1}^{N} \mathbb{1}_{C_{k+1}^m} (Y_{k+1}^m, s) \]

\[ = \text{number particles make it to next threshold at stage } k \]

Object of study:

\[ E \prod_{k=1}^{m} \left( \frac{N_k^m}{N} \right)^2 = E e^m \left[ \frac{1}{m} \sum_{k=1}^{m} 2 \log \left( \frac{N_k^m}{N} \right) \right] = E e^m \left[ \sum_{n=0}^{N} 2 \log \left( \frac{n}{N} \right) L^m(n) \right], \]

with \( L^m \) the occupation measure of \( N_k^m \).
Relative entropy representations

Use famous duality between exponential integrals and relative entropy. Let $S$ have distribution $\gamma$ and let $f$ be bounded and measurable.
Use famous duality between \textit{exponential integrals} and \textit{relative entropy}. Let $S$ have distribution $\gamma$ and let $f$ be bounded and measurable. Then
\begin{equation*}
- \log E e^{-f(S)} = \inf E \left[ R(\eta \parallel \gamma) + f(\bar{S}) \right],
\end{equation*}
Use famous duality between exponential integrals and relative entropy. Let $S$ have distribution $\gamma$ and let $f$ be bounded and measurable. Then

$$- \log E e^{-f(S)} = \inf E \left[ R(\eta \parallel \gamma) + f(\bar{S}) \right],$$

where $\bar{S}$ has distribution $\eta$ and

$$R(\eta \parallel \gamma) = \begin{cases} \int \log \left( \frac{d\eta}{d\gamma}(s) \right) \eta(ds) & \eta \ll \gamma \\ \infty & \text{else} \end{cases}$$

is relative entropy.
Relative entropy representations (cont’d)

Suppose $(S_0, S_1)$ are two steps of a Markov chain with initial distribution $\gamma_0(ds_0)$ and transition $\gamma_1(ds_1|s_0)$. 

Using the duality and the chain rule to decompose the relative entropy

$$\log E e^{f(S_0; S_1)} = \inf E R(\gamma_0) + R(\gamma_1 | \gamma_0) + f(S_0; S_1);$$

where $S_0$ has distribution $\gamma_0(ds_0)$ and $S_1$ has conditional distribution $\gamma_1(ds_1|s_0)$. 

Paul Dupuis (Brown University) June, 2010
Relative entropy representations (cont’d)

Suppose \((S_0, S_1)\) are two steps of a Markov chain with initial distribution \(\gamma_0(ds_0)\) and transition \(\gamma_1(ds_1|s_0)\).

**Chain rule.** Writing

\[
\gamma(ds_0, ds_1) = \gamma_0(ds_0)\gamma_1(ds_1|s_0), \quad \eta(ds_0, ds_1) = \eta_0(ds_0)\eta_1(ds_1|s_0),
\]

\[
R(\eta || \gamma) = R(\eta_0 || \gamma_0) + \int R(\eta_1(\cdot|s_0) || \gamma_1(\cdot|s_0)) \eta_0(ds_0).
\]
Suppose \((S_0, S_1)\) are two steps of a Markov chain with initial distribution \(\gamma_0(ds_0)\) and transition \(\gamma_1(ds_1|s_0)\).

**Chain rule.** Writing
\[
\gamma(ds_0, ds_1) = \gamma_0(ds_0)\gamma_1(ds_1|s_0), \quad \eta(ds_0, ds_1) = \eta_0(ds_0)\eta_1(ds_1|s_0),
\]
\[
R(\eta \| \gamma) = R(\eta_0 \| \gamma_0) + \int R(\eta_1(\cdot |s_0) \| \gamma_1(\cdot |s_0)) \eta_0(ds_0).
\]

Using the duality and the chain rule to decompose the relative entropy
\[
- \log E e^{-f(S_0, S_1)} = \inf E \left[ R(\eta_0 \| \gamma_0) + R(\eta_1(\cdot |\bar{S}_0) \| \gamma_1(\cdot |\bar{S}_0)) + f(\bar{S}_0, \bar{S}_1) \right],
\]
where \(\bar{S}_0\) has distribution \(\eta_0(\cdot)\) and \(\bar{S}_1\) has conditional distribution \(\eta_1(\cdot |\bar{S}_0)\).
A representation for the IPS, which works with threshold hitting times (rather than ordinary time), can be proved.
Relative entropy representations (cont’d)

A representation for the IPS, which *works with threshold hitting times* (rather than ordinary time), can be proved. Let

\[ \alpha_k^m(x, dy) = P \left\{ X_{\sigma^m}^m \in \frac{k}{m} + dy | X_0^m = \frac{k - 1}{m} + x \right\}, \]

\[ \sigma^m = \inf \left\{ i : X_i^n \in C^m_k \cup \Delta \right\}. \]
A representation for the IPS, which works with threshold hitting times (rather than ordinary time), can be proved. Let

\[ \alpha_k^m(x, dy) = P \left\{ X_{\sigma^m}^m \in \frac{k}{m} + dy \mid X_0^m = \frac{k - 1}{m} + x \right\}, \]

\[ \sigma^m = \inf \{ i : X_i^n \in C_k^m \cup \Delta \}. \]

\( x \) is position relative to \( C_{k-1}^m \), \( y \) is position relative to \( C_k^m \):

\[ \begin{array}{c|c|c}
\hline
x & y \\
\hline
0 & \frac{k-1}{m} & \frac{k}{m} & \frac{k+1}{m} & 1 \\
\hline
\end{array} \]
Relative entropy representations (cont’d)

\[ \alpha_k^m(x, dy) \] becomes independent of \( m \) and \( k \) as \( m \to \infty \), and acts like stationary transition kernel \( \alpha(x, dy) \).
Relative entropy representations (cont’d)

\( \alpha^m_k (x, dy) \) becomes independent of \( m \) and \( k \) as \( m \to \infty \), and acts like stationary transition kernel \( \alpha(x, dy) \). We consider a controlled IPS system (denoted by bars)

\[
\begin{align*}
\tilde{Z}_k^m \xrightarrow{\text{transportation}} & \tilde{Y}_{k+1}^m \quad \xrightarrow{\text{resampling}} \tilde{Z}_{k+1}^m,
\end{align*}
\]

where threshold transition kernels \( \alpha^m_k (x, dy) \) are replaced by controlled transition kernels, and uniform resampling distributions replaced by controlled (biased) resampling distributions.
Relative entropy representations (cont’d)

\[- \frac{1}{m} \log E \prod_{k=1}^{m} \left( \frac{N_k^m}{N} \right)^2 \]

\[= \inf_{\text{controls}} E \left\{ \frac{1}{m} \sum_{k=1}^{m} \left[ \sum_{\text{parts } s \text{ at threshold } k-1 \text{ needing transp}} R \left( \text{new trans kern} \mid \alpha_{k-1}^m(\tilde{Z}_{k,s}, \cdot) \right) \right] \right\}

\[+ \sum_{\text{parts } s \text{ at threshold } k \text{ needing resampling}} R \left( \text{biased resamp} \mid \text{uniform} \right) \]

\[\left\{ - 2 \log \left( \frac{\tilde{N}_k^m}{N} \right) \right\} \]
Main results

This is a (somewhat nonstandard) average cost per unit time problem for controlled Markov chain.
Main results

This is a (somewhat nonstandard) average cost per unit time problem for controlled Markov chain. Can analyze limit $m \to \infty$ using weak convergence and prove

$$- \frac{1}{m} \log E \prod_{k=1}^{m} \left( \frac{N_k^m}{N} \right)^2 \to V^*.$$
Main results

This is a (somewhat nonstandard) average cost per unit time problem for controlled Markov chain. Can analyze limit \( m \to \infty \) using weak convergence and prove

\[- \frac{1}{m} \log E \prod_{k=1}^{m} \left( \frac{N_k^m}{N} \right)^2 \to V^*.\]

- Lower bound (upper bound on second moment) needs no additional assumptions.
Main results

This is a (somewhat nonstandard) average cost per unit time problem for controlled Markov chain. Can analyze limit $m \to \infty$ using weak convergence and prove

$$- \frac{1}{m} \log E \prod_{k=1}^{m} \left( \frac{N_k^m}{N} \right)^2 \to V^*.$$

- Lower bound (upper bound on second moment) needs no additional assumptions.
- Upper bound requires communication/ergodicity condition plus Feller property on $\alpha(x, dy)$. 

Paul Dupuis (Brown University) 
June, 2010
Main results

This is a (somewhat nonstandard) average cost per unit time problem for controlled Markov chain. Can analyze limit $m \to \infty$ using weak convergence and prove

$$-rac{1}{m} \log E \prod_{k=1}^{m} \left( \frac{N_k^m}{N} \right)^2 \to V^*.$$  

- Lower bound (upper bound on second moment) needs no additional assumptions.
- Upper bound requires communication/ergodicity condition plus Feller property on $\alpha(x, dy)$. 
Main results

This is a (somewhat nonstandard) average cost per unit time problem for controlled Markov chain. Can analyze limit $m \to \infty$ using weak convergence and prove

$$-\frac{1}{m} \log E \prod_{k=1}^{m} \left( \frac{N_k^m}{N} \right)^2 \to V^*.$$ 

- Lower bound (upper bound on second moment) needs no additional assumptions.
- Upper bound requires communication/ergodicity condition plus Feller property on $\alpha(x, dy)$. However, a priori upper bound of $2\gamma$ (twice LD rate) holds without assumptions.
Main results (cont’d)

**Expression for** $V^*$. A collection of linked quantities:

- $m(dz_1, \ldots, dz_N)$ a stationary distribution on $N$ particles before transportation.
Expression for $V^*$. A collection of linked quantities:

- $m(dz_1, \ldots, dz_N)$ a stationary distribution on $N$ particles before transportation
- $m_I(s)$ probability that $s$ of $N$ need transportation
Main results (cont’d)

Expression for \( V^* \). A collection of linked quantities:

- \( m(dz_1, \ldots, dz_N) \) a stationary distribution on \( N \) particles before transportation
- \( m_I(s) \) probability that \( s \) of \( N \) need transportation
- \( m(dz_1, \ldots, dz_N|s) \) conditional distribution given \( s \) (ordered as first \( s \)) need transportation
Main results (cont’d)

Expression for $V^*$. A collection of linked quantities:

- $m(dz_1, \ldots, dz_N)$ a stationary distribution on $N$ particles before transportation
- $m_I(s)$ probability that $s$ of $N$ need transportation
- $m(dz_1, \ldots, dz_N|s)$ conditional distribution given $s$ (ordered as first $s$) need transportation
- $\tau(dy_1, \ldots, dy_s|z_1, \ldots, z_N, s)$ transportation control, distribution with support on $([0, \infty) \cup \Delta)^s$
Main results (cont’d)

Expression for $V^*$. A collection of linked quantities:

- $m(dz_1, \ldots, dz_N)$ a stationary distribution on $N$ particles before transportation
- $m_I(s)$ probability that $s$ of $N$ need transportation
- $m(dz_1, \ldots, dz_N | s)$ conditional distribution given $s$ (ordered as first $s$) need transportation
- $\tau(dy_1, \ldots, dy_s | z_1, \ldots, z_N, s)$ transportation control, distribution with support on $([0, \infty) \cup \Delta)^s$
- $m_N(n)$ probability that $n$ make the next threshold
Main results (cont’d)

Expression for $V^*$. A collection of linked quantities:

- $m(dz_1, \ldots, dz_N)$ a stationary distribution on $N$ particles before transportation
- $m_I(s)$ probability that $s$ of $N$ need transportation
- $m(dz_1, \ldots, dz_N|s)$ conditional distribution given $s$ (ordered as first $s$) need transportation
- $\tau(dy_1, \ldots, dy_s|z_1, \ldots, z_N, s)$ transportation control, distribution with support on $([0, \infty) \cup \Delta)^s$
- $m_N(n)$ probability that $n$ make the next threshold
- $\tilde{m}(dy_1, \ldots, dy_n|n)$ distribution of those making next threshold
Main results (cont’d)

Expression for $V^*$. A collection of linked quantities:

- $m(dz_1, \ldots, dz_N)$ a stationary distribution on $N$ particles before transportation
- $m_I(s)$ probability that $s$ of $N$ need transportation
- $m(dz_1, \ldots, dz_N|s)$ conditional distribution given $s$ (ordered as first $s$) need transportation
- $\tau(dy_1, \ldots, dy_s|z_1, \ldots, z_N, s)$ transportation control, distribution with support on $([0, \infty) \cup \Delta)^s$
- $m_N(n)$ probability that $n$ make the next threshold
- $\bar{m}(dy_1, \ldots, dy_n|n)$ distribution of those making next threshold
- $\gamma(dy_{n+1}, \ldots, dy_N|n)$ resampling control determines location of new particles.
Consider $m$ that is **stationary** under application of $\tau$ and $\gamma$, with cost

$$C(m, \tau, \gamma) =$$

$$\sum_{s=0}^{N} \int R \left( \tau(\cdot, \ldots, \cdot | z_1, \ldots, z_N, s) \left\| \prod_{j=1}^{s} \alpha(z_j, \cdot) \right\| m(dz_1, \ldots, dz_N | s) m_I(s) \right)$$

$$+ \sum_{n=1}^{N} \int R \left( \gamma(\cdot, \ldots, \cdot | n) \left\| \prod_{j=n+1}^{N} \left[ \frac{1}{n} \sum_{\ell=1}^{n} \delta_{y_{\ell}} \right] (\cdot) \right\| \bar{m}(dy_1, \ldots, dy_n | n) m_N(n) \right)$$

$$- \sum_{n=1}^{N} 2 \log \left( \frac{n}{N} \right) m_N(n).$$
Consider \( m \) that is **stationary** under application of \( \tau \) and \( \gamma \), with cost

\[
C(m, \tau, \gamma) = \\
\sum_{s=0}^{N} \int R \left( \tau(\cdot, \ldots, \cdot | z_1, \ldots, z_N, s) \left| \prod_{j=1}^{s} \alpha(z_j, \cdot) \right| m(dz_1, \ldots, dz_N | s) m_I(s) \right) \\
+ \sum_{n=1}^{N} \int R \left( \gamma(\cdot, \ldots, \cdot | n) \left| \prod_{j=n+1}^{N} \left[ \frac{1}{n} \sum_{\ell=1}^{n} \delta_{y_{j\ell}} \right] (\cdot) \right| \bar{m}(dy_1, \ldots, dy_n | n) m_N(n) \right) \\
- \sum_{n=1}^{N} 2 \log \left( \frac{n}{N} \right) m_N(n). \\
\]

Then

\[
V^* = \inf_{m, \tau, \gamma} C(m, \tau, \gamma). 
\]
Main results (cont’d)

Now consider the dependence on the number of particles and write $V^{*,N}$. 

How does $V^{*,N}$ depend on $N$? Using LLN and CLT type approximations for the underlying distribution $(x, dy)$, one can show $V^{*,N} \leq C N$ for a constant $C$. 

Paul Dupuis (Brown University) June, 2010
Main results (cont’d)

Now consider the dependence on the number of particles and write $V^{*,N}$. With no additional assumptions

$$V^{*,N} \leq \liminf_{m \to \infty} -\frac{1}{m} \log E \prod_{k=1}^{m} \left( \frac{N_{k}^{m}}{N} \right)^{2}$$

$$\leq \limsup_{m \to \infty} -\frac{1}{m} \log E \prod_{k=1}^{m} \left( \frac{N_{k}^{m}}{N} \right)^{2} \leq 2\gamma^*.$$
Now consider the dependence on the number of particles and write $V^{*, N}$. With no additional assumptions

$$V^{*, N} \leq \liminf_{m \to \infty} - \frac{1}{m} \log E \prod_{k=1}^{m} \left( \frac{N_k^m}{N} \right)^2$$

$$\leq \limsup_{m \to \infty} - \frac{1}{m} \log E \prod_{k=1}^{m} \left( \frac{N_k^m}{N} \right)^2 \leq 2\gamma^*.$$

How does $V^{*, N}$ depend on $N$? Using LLN and CLT type approximations for the underlying distribution $\alpha(x, dy)$, one can show

$$V^{*, N} \geq 2\gamma^* - \frac{C}{N}$$

for a constant $C$. 
Remarks on higher dimensions

In prior work with T. Dean (following work with H. Wang on IS) have shown that good performance of standard splitting is tied to threshold being level curves of *subsolution* to related Hamilton-Jabobi-Bellman equation.
Remarks on higher dimensions

In prior work with T. Dean (following work with H. Wang on IS) have shown that good performance of standard splitting is tied to threshold being level curves of *subsolution* to related Hamilton-Jabobi-Bellman equation. We consider again the hitting problem and let $M$ be mean number of descendents at split.
With the rate function in the standard form

\[ I_T(\phi) = \int_0^T L(\phi, \dot{\phi}) \, dt, \]

let \( \mathbb{H}(y, \alpha) \) be the Legendre transform of the local rate function \( L(y, \beta) \):

\[ \mathbb{H}(y, \alpha) = \inf_{\beta} \left[ \langle \alpha, \beta \rangle + L(y, \beta) \right]. \]
With the rate function in the standard form

\[ I_T(\phi) = \int_0^T L(\phi, \dot{\phi}) dt, \]

let \( \mathcal{H}(y, \alpha) \) be the Legendre transform of the local rate function \( L(y, \beta) \):

\[ \mathcal{H}(y, \alpha) = \inf_{\beta} \left[ \langle \alpha, \beta \rangle + L(y, \beta) \right]. \]

Suppose the thresholds are level sets of some function \( W \) with spacing \( (\log M)/m \). Then expected number of descendants of a single particle started at \( y \) in the next threshold:

\[ e^{-\mathcal{H}(y, DW(y))}. \]
Remarks on higher dimensions (cont’d)

Notion of subsolution:

\[ \liminf_{m \to 1} \frac{1}{m \log E(m)} < 2 \]

\[ H(y, DW(y)) \geq 0 \]

\[ W(y) \leq 0 \]

\[ W(y) \leq \infty \]
Remarks on higher dimensions (cont’d)

Notion of subsolution:

Recall the notion of asymptotic efficiency:

\[
\liminf_{m \to \infty} \frac{1}{m} \log E (\theta^m)^2 \geq 2\gamma(x).
\]
Ordinary splitting. Consider a continuous function $W$ and suppose splitting thresholds are the level sets $\{W(y) \leq (K^m - k)(\log M)/m\}$, where $M$ is the number of particles per split.
Ordinary splitting. Consider a continuous function $W$ and suppose splitting thresholds are the level sets $\{W(y) \leq (K^m - k)(\log M)/m\}$, where $M$ is the number of particles per split.

- Then the number of particles needed to construct a single sample $\theta^n$ grows subexponentially if and only if $W$ is a viscosity subsolution.
Ordinary splitting. Consider a continuous function $W$ and suppose splitting thresholds are the level sets $\{W(y) \leq (K^m - k)(\log M)/m\}$, where $M$ is the number of particles per split.

- Then the number of particles needed to construct a single sample $\theta^n$ grows subexponentially if and only if $W$ is a viscosity subsolution.
- If $W$ is a viscosity subsolution

$$
\lim_{m \to \infty} - \frac{1}{m} \log E (\theta^m)^2 = W(x) + \gamma(x).
$$
Ordinary splitting. Consider a continuous function $W$ and suppose splitting thresholds are the level sets \( \{ W(y) \leq (K^m - k)(\log M)/m \} \), where $M$ is the number of particles per split.

- Then the number of particles needed to construct a single sample $\theta^n$ grows subexponentially if and only if $W$ is a viscosity subsolution.
- If $W$ is a viscosity subsolution

\[
\lim_{m \to \infty} - \frac{1}{m} \log E (\theta^m)^2 = W(x) + \gamma(x).
\]
Ordinary splitting. Consider a continuous function $W$ and suppose splitting thresholds are the level sets \( \{ W(y) \leq (K^m - k)(\log M)/m \} \), where $M$ is the number of particles per split.

- Then the number of particles needed to construct a single sample $\theta^n$ grows subexponentially if and only if $W$ is a viscosity subsolution.
- If $W$ is a viscosity subsolution

\[
\lim_{m \to \infty} -\frac{1}{m} \log E(\theta^m)^2 = W(x) + \gamma(x).
\]

Thus the design problem: find subsolutions with large value $W(x)$. 

Paul Dupuis (Brown University)  
June, 2010
Remarks on higher dimensions (cont’d)

Conjecture for IPS (need to complete proof of tightness for weak convergence):

Suppose that the splitting thresholds are of the form $C^m_k = f(x) W(x)$ and consider any monotone nondecreasing function $g: \mathbb{R} \rightarrow \mathbb{R}$ with $g(0) = 0$ that $g W$ is a subsolution for the associated HJB equation. Then

$$\liminf_{m \to 1} \frac{1}{m} \log E(m) g(W(x)) C^{N} = m g,$$

Then sup on $g$ for tightest bound.

Remarks:
In comparison to ordinary splitting, which requires that $W$ be a subsolution, here one extra degree of freedom. However, second moment rate for ordinary splitting $W(x) + g(x)$, so if $W$ is actually a subsolution with IPS we lose $C = N$ in rate of decay.
Conjecture for IPS (need to complete proof of tightness for weak convergence): Suppose that the splitting thresholds are of the form 
\[ C^m_k = \{ x : W(x) \leq (K^m - k)/m \} \]
and consider any monotone nondecreasing function \( g : \mathbb{R} \to \mathbb{R} \) with \( g(0) = 0 \) that \( g \circ W \) is a subsolution for the associated HJB equation. Then

\[
\liminf_{m \to \infty} \frac{1}{m} \log E(\theta^m)^2 \geq g(W(x)) - \frac{C}{N} + \gamma(x).
\]
Remarks on higher dimensions (cont’d)

**Conjecture for IPS** (need to complete proof of tightness for weak convergence): Suppose that the splitting thresholds are of the form $C^m_k = \{x : W(x) \leq (K^m - k)/m\}$, and consider any monotone nondecreasing function $g : \mathbb{R} \to \mathbb{R}$ with $g(0) = 0$ that $g \circ W$ is a subsolution for the associated HJB equation. Then

$$\liminf_{m \to \infty} \frac{1}{m} \log E (\theta^m)^2 \geq g(W(x)) - \frac{C}{N} + \gamma(x).$$

Then sup on $g$ for tightest bound.
Conjecture for IPS (need to complete proof of tightness for weak convergence): Suppose that the splitting thresholds are of the form $C_k^m = \{x : W(x) \leq (K^m - k)/m\}$, and consider any monotone nondecreasing function $g : \mathbb{R} \to \mathbb{R}$ with $g(0) = 0$ that $g \circ W$ is a subsolution for the associated HJB equation. Then

$$\liminf_{m \to \infty} -\frac{1}{m} \log E (\theta^m)^2 \geq g(W(x)) - \frac{C}{N} + \gamma(x).$$

Then sup on $g$ for tightest bound. Remarks:

- In comparison to ordinary splitting, which requires that $W$ be a subsolution, here one extra degree of freedom.

- However, second moment rate for ordinary splitting $W(x) + \gamma(x)$, so if $W$ is actually a subsolution with IPS we lose $C/N$ in rate of decay.
Remarks:

- Examples easily constructed to show the IPS scheme has poor asymptotic performance if there is no monotone mapping $g$ which makes $g \circ W$ into a subsolution.

- Unlike ordinary splitting, proof (assuming tightness) only for continuous statistics (so far).