Efficient Simulation for Rare Events

Jose Blanchet

Columbia Department of IEOR

Lunteren Conference 2010

Agenda

- **Introduction**
- A Simple Random Walk Example
- Systematic Approach
- Testing Efficiency
- Counter-examples, Heavy-tails and Beyond

Rare Event Analysis: Areas of Applicability

Rare events are consequential in many areas

- Insurance / Finance

Rare Event Analysis: Areas of Applicability

Rare events are consequential in many areas

- Insurance / Finance
- Congestion models (queues)

Rare Event Analysis: Areas of Applicability

Rare events are consequential in many areas

- Insurance / Finance
- Congestion models (queues)
- Environmental applications

Rare Event Analysis: Areas of Applicability

Rare events are consequential in many areas

- Insurance / Finance
- Congestion models (queues)
- Environmental applications
- Search problems

Rare Event Analysis: Areas of Applicability

Rare events are consequential in many areas

- Insurance / Finance
- Congestion models (queues)
- Environmental applications
- Search problems
- Reliabiliaty models

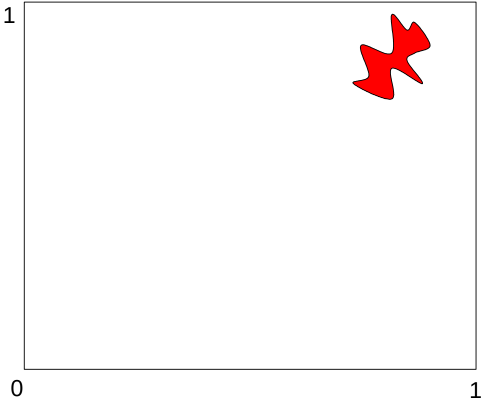
Rare Event Analysis: Areas of Applicability

Rare events are consequential in many areas

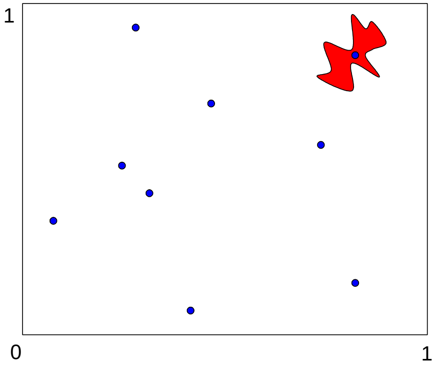
- Insurance / Finance
- Congestion models (queues)
- Environmental applications
- Search problems
- Reliabiliaty models
- Statistics

Why are Rare Events Difficult to Assess?

- Typically no closed forms (complex systems)
- But crudely implemented simulation might not be good



Why are Rare Events Difficult to Assess?



Why are Rare Events Difficult to Assess?

- **Relative mean squared error (RMSE)** PER TRIAL = stdev / mean

$$\frac{\sqrt{P(\text{RED})(1 - P(\text{RED}))}}{P(\text{RED})} \approx \frac{1}{\sqrt{P(\text{RED})}}$$

Performance Analysis

- *General Focus*: Estimate $P(A)$ assuming $P(A) \approx 0$.

Performance Analysis

- *General Focus*: Estimate $P(A)$ assuming $P(A) \approx 0$.
- *Goal*: Design estimator so that **relative mean squared error (RMSE)** is controlled

Performance Analysis

- *General Focus*: Estimate $P(A)$ assuming $P(A) \approx 0$.
- *Goal*: Design estimator so that **relative mean squared error (RMSE)** is controlled
- **STRONG EFFICIENCY:**

$$\textbf{RMSE} \ \text{ for } P(A) = O(1)$$

Performance Analysis

- *General Focus*: Estimate $P(A)$ assuming $P(A) \approx 0$.
- *Goal*: Design estimator so that **relative mean squared error (RMSE)** is controlled
- **STRONG EFFICIENCY:**

$$\textbf{RMSE} \;\; \text{for } P(A) = O(1)$$

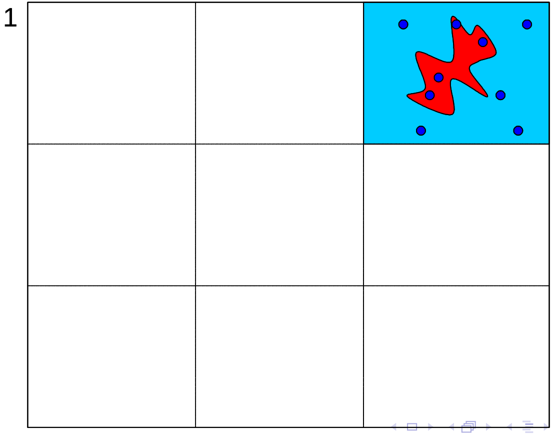
- **WEAK EFFICIENCY:** For each $\varepsilon > 0$

$$\textbf{RMSE} \;\; \text{for } P(A) = O\left(1/P(A)^{\varepsilon}\right)$$

Graphical Interpretation of Importance Sampling

- **Importance sampling (I.S.)**: sample from the important region and correct via likelihood ratio

$$\text{RED AREA} \approx \text{PROPORTION DARTS IN RED AREA} \times \frac{1}{9}$$



Importance Sampling

- **Goal:** Estimate $P(A) > 0$

Importance Sampling

- **Goal:** Estimate $P(A) > 0$
- Choose $\widetilde{P}$ and simulate ω from it

Importance Sampling

- **Goal:** Estimate $P(A) > 0$
- Choose $\widetilde{P}$ and simulate ω from it
- **Importance Sampling (I.S.)** estimator *per trial* is

$$I.S.Estimator = L(\omega)\, I(\omega \in A),$$

where $L(\omega)$ is the likelihood ratio (i.e. $L(\omega) = P(\omega)/\widetilde{P}(\omega)$).

Importance Sampling

- **Goal:** Estimate $P(A) > 0$
- Choose $\widetilde{P}$ and simulate ω from it
- **Importance Sampling (I.S.)** estimator *per trial* is

$$I.S.Estimator = L(\omega)\, I(\omega \in A),$$

where $L(\omega)$ is the likelihood ratio (i.e. $L(\omega) = P(\omega)/\widetilde{P}(\omega)$).

- **NOTE:** $\widetilde{P}$ is called a change-of-measure

- Suppose we choose $\widetilde{P}(\cdot) = P(\cdot|\, A)$

$$L(\omega) = \frac{P(\omega)\, I(\omega \in A)}{P(\omega)\, I(\omega \in A)\, / P(A)} = P(A)$$

Importance Sampling

- Suppose we choose $\widetilde{P}(\cdot) = P(\cdot | A)$

$$L(\omega) = \frac{P(\omega) I(\omega \in A)}{P(\omega) I(\omega \in A) / P(A)} = P(A)$$

- Estimator has *zero variance*, but requires knoweledge of $P(A)$

Importance Sampling

- Suppose we choose $\widetilde{P}(\cdot) = P(\cdot | A)$

$$L(\omega) = \frac{P(\omega) I(\omega \in A)}{P(\omega) I(\omega \in A) / P(A)} = P(A)$$

- Estimator has *zero variance*, but requires knoweledge of $P(A)$
- **Lesson:** *Try choosing* $\widetilde{P}(\cdot)$ *close to* $P(\cdot | A)$*!*

Agenda

- Introduction
- **A Simple Random Walk Example**
- Systematic Approach
- Testing Efficiency
- Counter-examples, Heavy-tails and Beyond

Insurance

- V_i's i.i.d. (independent and identially distributed) claim sizes

Insurance

- V_i's i.i.d. (independent and identially distributed) claim sizes
- τ_i's i.i.d. inter-arrival times

Insurance

- V_i's i.i.d. (independent and identially distributed) claim sizes
- τ_i's i.i.d. inter-arrival times
- A_n = time of the n-th arrival

Insurance

- V_i's i.i.d. (independent and identially distributed) claim sizes
- τ_i's i.i.d. inter-arrival times
- A_n = time of the n-th arrival
- Constant premium p

Insurance

- V_i's i.i.d. (independent and identially distributed) claim sizes
- τ_i's i.i.d. inter-arrival times
- A_n = time of the n-th arrival
- Constant premium p
- Reserve process

$$R(t) = b + pt - \sum_{j=1}^{N(t)} V_j$$

Insurance

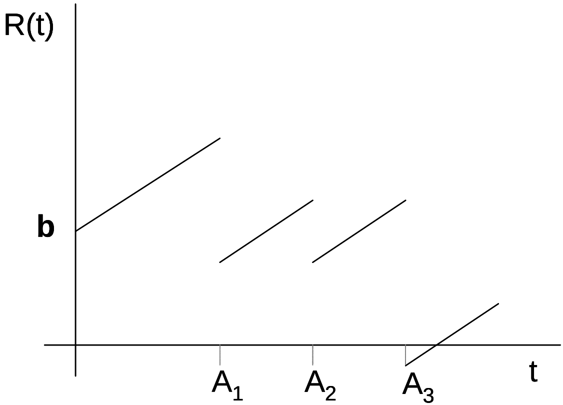
- V_i's i.i.d. (independent and identially distributed) claim sizes
- τ_i's i.i.d. inter-arrival times
- A_n = time of the n-th arrival
- Constant premium p
- Reserve process

$$R(t) = b + pt - \sum_{j=1}^{N(t)} V_j$$

- $N(t) = \#$ arrivals up to time t

Insurance

Plot of risk reserve



Insurance

- Evaluating reserve at arrival times we get random walk

Insurance

- Evaluating reserve at arrival times we get random walk
- Suppose $Y_1, Y_2, \ldots$ are i.i.d.

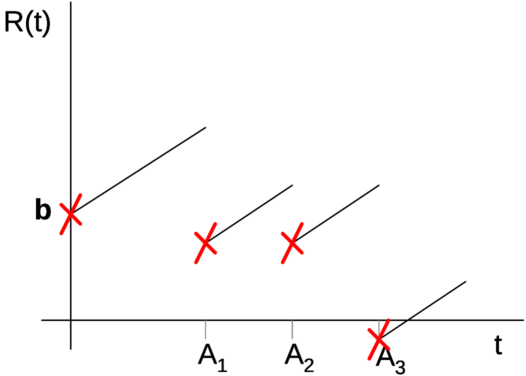
$$S(n) = b + Y_1 + \ldots + Y_n$$

Insurance

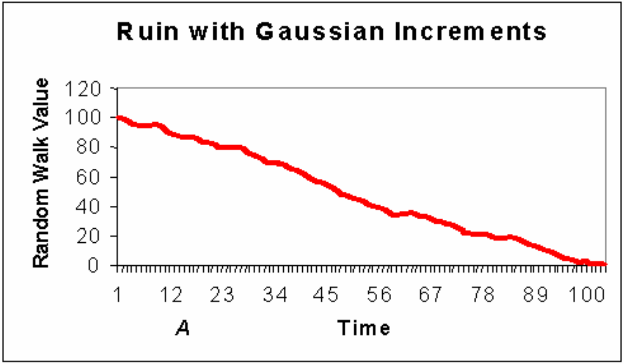
- Evaluating reserve at arrival times we get random walk
- Suppose $Y_1, Y_2, \ldots$ are i.i.d.

$$S(n) = b + Y_1 + \ldots + Y_n$$

- $R(A_n) = S(n)$ reserve at arrival times with $Y_n = p\tau_n - V_n$.

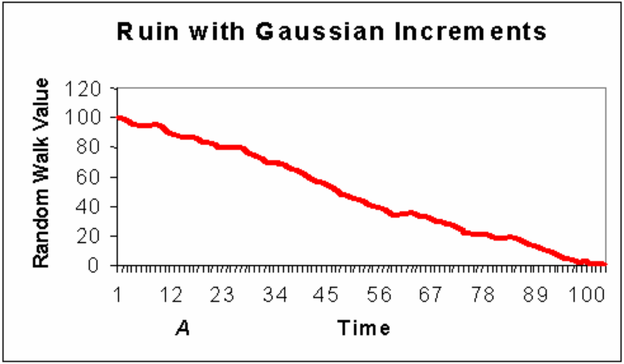


Insurance Process Conditioned on Ruin



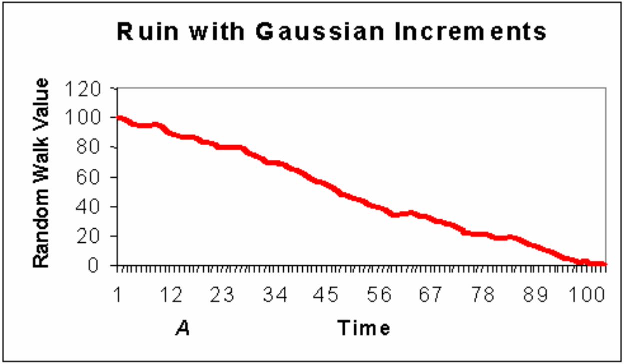
- Y_i's are $N(1,1)$, $EY_i = 1$

Insurance Process Conditioned on Ruin



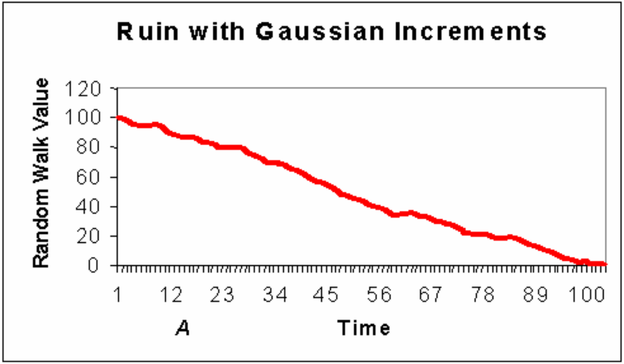
- Y_i's are $N(1,1)$, $EY_i = 1$
- Random walk conditioned on ruin

Insurance Process Conditioned on Ruin



- Y_i's are $N(1,1)$, $EY_i = 1$
- Random walk conditioned on ruin
- Light tails: Exponential, Gamma, Gaussian, mixtures of these, etc.

Insurance Process Conditioned on Ruin



- Y_i's are $N(1, 1)$, $EY_i = 1$
- Random walk conditioned on ruin
- Light tails: Exponential, Gamma, Gaussian, mixtures of these, etc.
- Picture generated with Siegmund's 76 algorithm

Light Tails Setting: Asymptotic Conditional Distributions

- In light-tailed cases there is large deviations theory (ref. Dembo and Zeitouni '99).

Light Tails Setting: Asymptotic Conditional Distributions

- In light-tailed cases there is large deviations theory (ref. Dembo and Zeitouni '99).
- Large deviations allows to obtain as $b \nearrow \infty$

$$P\left(Y_1 \leq x, ..., Y_k \leq x \right| \text{ruin starting at } b) \approx \widetilde{P}\left(Y_1 \leq x\right)...\widetilde{P}\left(Y_k \leq x\right),$$

Light Tails Setting: Asymptotic Conditional Distributions

- In light-tailed cases there is large deviations theory (ref. Dembo and Zeitouni '99).
- Large deviations allows to obtain as $b \nearrow \infty$

$$P\left(Y_1 \leq x, ..., Y_k \leq x \middle| \text{ruin starting at } b\right) \approx \widetilde{P}\left(Y_1 \leq x\right) ... \widetilde{P}\left(Y_k \leq x\right),$$

- Suggested change-of-measure: Sample Y_k's i.i.d. using $\widetilde{P}\left(\cdot\right)$

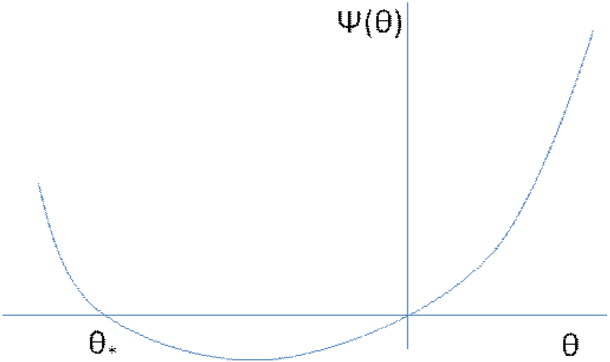
$$L = \frac{p\left(Y_1\right)}{\widetilde{p}\left(Y_1\right)} \cdot \frac{p\left(Y_2\right)}{\widetilde{p}\left(Y_2\right)} \cdot ... \cdot \frac{p\left(Y_{\text{ruin time}}\right)}{\widetilde{p}\left(Y_{\text{ruin time}}\right)}.$$

Light Tails Setting: Exponential Tilting

- More precisely if $p(\cdot)$ is the density of Y_i

$$\widetilde{p}(y) = p(y)\exp(\theta_* y)$$

where $\theta_* < 0$ solves $\psi(\theta_*) = \log E \exp(\theta_* Y_i) = 0$.



Siegmund's Algorithm for Random Walks

Theorem (Siegmund '76)

Assume $\psi(\theta_* - \delta) < \infty$ *for some* $\delta > 0$. *Then, the estimator*

$$L = \frac{p(Y_1)}{\widetilde{p}(Y_1)} \cdot \frac{p(Y_2)}{\widetilde{p}(Y_2)} \cdot \ldots \cdot \frac{p(Y_{ruin\ time})}{\widetilde{p}(Y_{ruin\ time})}$$
$$= \exp\left(-\theta_*[Y_1 + \ldots + Y_{ruin\ time}]\right)$$

is STRONGLY EFFICIENT. Moreover, $\widetilde{p}(\cdot)$ *is the ONLY STATE-INDEPENDENT change-of-measure that achieves efficiency.*

Moral of the Story for Light Tails...

1. Asymptotic conditional distribution well described using Large Deviations theory

Moral of the Story for Light Tails...

1. Asymptotic conditional distribution well described using Large Deviations theory
2. Description in terms of exponential tilting —> fundamental family of changes-of-measure

Moral of the Story for Light Tails...

1. Asymptotic conditional distribution well described using Large Deviations theory
2. Description in terms of exponential tilting —> fundamental family of changes-of-measure
3. Items 1) and 2) provide systematic tools for rare-event simulation

Agenda

- Introduction
- A Simple Random Walk Example
- **Systematic Approach**
- Testing Efficiency
- Counter-examples, Heavy-tails and Beyond

Description of Systematic Approach

- State-dependent random walk: $t \in \{0, \Delta, 2\Delta, \ldots\}$

$$Y_{t+\Delta} = Y_t + \Delta X_{t+\Delta}(Y_t)$$

Description of Systematic Approach

- State-dependent random walk: $t \in \{0, \Delta, 2\Delta, ...\}$

$$Y_{t+\Delta} = Y_t + \Delta X_{t+\Delta}(Y_t)$$

- Given $Y_t = y$, $X_{t+\Delta}(y)$ is random variable with finite moment generating function

$$\psi(\theta, y) = \log E[\exp(\theta X_{t+\Delta}(y))]$$

Description of Systematic Approach

- State-dependent random walk: $t \in \{0, \Delta, 2\Delta, ...\}$

$$Y_{t+\Delta} = Y_t + \Delta X_{t+\Delta}(Y_t)$$

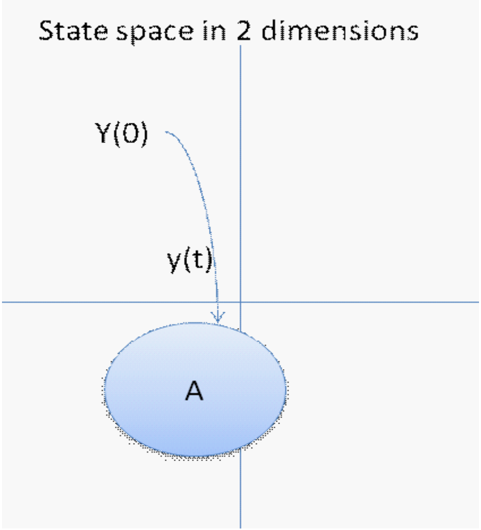
- Given $Y_t = y$, $X_{t+\Delta}(y)$ is random variable with finite moment generating function

$$\psi(\theta, y) = \log E[\exp(\theta X_{t+\Delta}(y))]$$

- As $\Delta \longrightarrow 0$ under mild assumptions $Y_{\cdot} \longrightarrow y(\cdot)$ so that

$$\dot{y}(t) = E[X(y(t))].$$

Fluid Limit of State-dependent Random Walk



Description of Systematic Approach

- Given $z(t)$ described by an ODE so that $z(t) \neq y(t)$ one often has

$$P(Y_t \approx z(t)) \approx \exp(-J(z)/\Delta)$$

Description of Systematic Approach

- Given $z(t)$ described by an ODE so that $z(t) \neq y(t)$ one often has

$$P(Y_t \approx z(t)) \approx \exp(-J(z)/\Delta)$$

- Associated Legendre transform

$$I(z, y) = \sup_{\theta}[\theta z - \psi(\theta, y)]$$

- Given $z(t)$ described by an ODE so that $z(t) \neq y(t)$ one often has

$$P(Y_t \approx z(t)) \approx \exp(-J(z)/\Delta)$$

- Associated Legendre transform

$$I(z, y) = \sup_{\theta}[\theta z - \psi(\theta, y)]$$

- Associated action integral

$$J(z) = \int_0^{\infty} I(\dot{z}(t), z(t))\, dt.$$

Description of Systematic Approach

- A generic rare-event estimation problem:

$$
\begin{aligned}
& \Delta \log P\,(\text{Hit } B \text{ prior to } A) \\
\approx\; & -\inf\{J\,(z) : z\,(\cdot) \text{ is path that hits } B \text{ prior to } A\}
\end{aligned}
$$

Description of Systematic Approach

- A generic rare-event estimation problem:

$$\begin{aligned} & \Delta \log P\left(\text{Hit } B \text{ prior to } A\right) \\ \approx\ & -\inf\{J\left(z\right) : z\left(\cdot\right) \text{ is path that hits } B \text{ prior to } A\} \end{aligned}$$

- Solution $z^{*}\left(\cdot\right)$ is called "optimal path"

Description of Systematic Approach

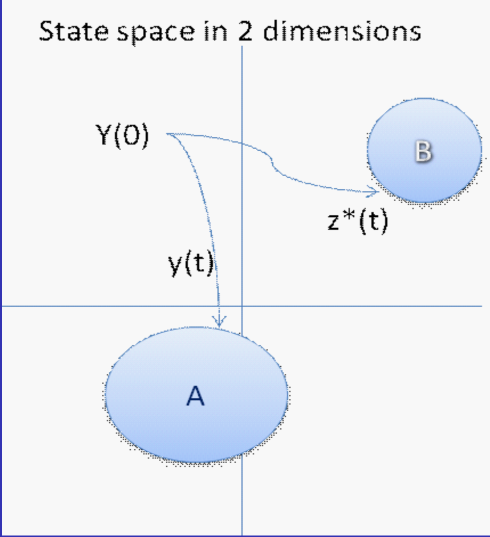
- A generic rare-event estimation problem:

$$
\begin{aligned}
& \Delta \log P\left(\text{Hit } B \text{ prior to } A\right) \\
\approx \; & -\inf\{J\left(z\right) : z\left(\cdot\right) \text{ is path that hits } B \text{ prior to } A\}
\end{aligned}
$$

- Solution $z^*\left(\cdot\right)$ is called "optimal path"
- **Tracking optimal path: Optimal Exponential Tilting $\theta_*\left(t\right)$ solves**

$$
\frac{\partial}{\partial\theta}\psi\left(\theta_*(t), z^*\left(t\right)\right) = \dot{z}^*\left(t\right).
$$

Description of Systematic Approach



Description of Systematic Approach

- **Tracking optimal path: Optimal Exponential Tilting** $\theta_*(t)$ **solves**

$$\frac{\partial}{\partial\theta}\psi\left(\theta_*(t), z^*\left(t\right)\right) = \dot{z}^*\left(t\right).$$

- **Tracking optimal path: Optimal Exponential Tilting** $\theta_*(t)$ **solves**

$$\frac{\partial}{\partial\theta}\psi\left(\theta_*(t), z^*(t)\right) = \dot{z}^*(t).$$

- **Importance Sampling Strategy:**

Description of Systematic Approach

- **Tracking optimal path: Optimal Exponential Tilting $\theta_*(t)$ solves**

$$\frac{\partial}{\partial\theta}\psi\left(\theta_*(t), z^*(t)\right) = \dot{z}^*(t).$$

- **Importance Sampling Strategy:**
  1. At time t apply exponential tilting $\theta_*(t)$ —> Corresponds to a so-called open-loop control (no feedback)...

Description of Systematic Approach

- **Tracking optimal path: Optimal Exponential Tilting $\theta_*(t)$ solves**

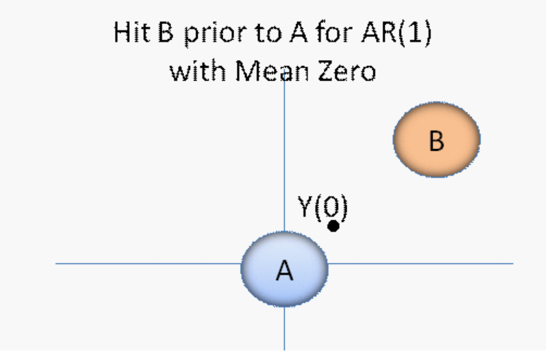
$$\frac{\partial}{\partial\theta}\psi\left(\theta_*(t), z^*(t)\right) = \dot{z}^*(t).$$

- **Importance Sampling Strategy:**

1. At time t apply exponential tilting $\theta_*(t)$ —> Corresponds to a so-called open-loop control (no feedback)...
2. Follow path to approximate $z^*(t)$...

Example: Two dimensional AR(1) Process

- **Model:** $Y_{t+\Delta} - Y_t = -\Delta Y_t + \Delta X_{t+\Delta}$, X_t's i.i.d. $N(0, I)$ and $y_0 = (1, 1)$.
- **Estimate:** $P_{y_0}(T_B < T_A)$, with $B = \{x : \left\| x - (e + 1/\sqrt{2}, e + 1/\sqrt{2}) \right\|_2 \leq 1\}$ & $A = \{x : ||x||_2 \leq 1\}$.



Example: Computing the Optimal Path

- **Calculus of Variations Problem:**

$$\min_{z \in C} \frac{1}{2} \int_0^T ||\dot{z}(t) + z(t)||_2^2 \, dt$$

where

$$\begin{aligned} C &= \{ z : z(0) = (1,1), \left\| z(T) - (e + 2^{-1/2})(1,1) \right\|_2 \leq 1, \\ &\quad T < \infty, \|z(t)\|_2 > 1, \ t < T \} \end{aligned}$$

Example: Computing the Optimal Path

- **Calculus of Variations Problem:**

$$\min_{z \in C} \frac{1}{2} \int_0^T ||\dot{z}(t) + z(t)||_2^2 \, dt$$

where

$$\begin{aligned} C &= \{z : z(0) = (1,1), \left\| z(T) - (e + 2^{-1/2})(1,1) \right\|_2 \leq 1, \\ &\quad T < \infty, \|z(t)\|_2 > 1, \; t < T\} \end{aligned}$$

- **Solution:**

$$\begin{aligned} z(t) &= (\exp(t), \exp(t)) \\ \dot{z}(t) &= z(t). \end{aligned}$$

Interpreting the Change-of-Measure

- **Importance Sampling Q:** W_t is i.i.d. N(0,1) under Q

$$\begin{aligned} Y_{t+\Delta} - Y_t &= \dot{z}(t)\,\Delta + \Delta W_{t+\Delta} \implies \text{Fluid } dy_t \approx \dot{z}(t)\,dt \\ Y_0 &= (1,1)\,, \\ X_{t+\Delta} &= W_{t+\Delta} + (\dot{z}(t) + Y_t). \end{aligned}$$

Interpreting the Change-of-Measure

- **Importance Sampling** Q**:** W_t is i.i.d. N(0,1) under Q

$$
\begin{aligned}
Y_{t+\Delta} - Y_t &= \dot{z}(t)\,\Delta + \Delta W_{t+\Delta} \Longrightarrow \text{Fluid } dy_t \approx \dot{z}(t)\,dt \\
Y_0 &= (1,1), \\
X_{t+\Delta} &= W_{t+\Delta} + (\dot{z}(t) + Y_t).
\end{aligned}
$$

- **Likelihood ratio** dP/dQ**:**

$$
L_0 = \exp\left(-\sum_{j=0}^{T_B/\Delta-1} (\dot{z}(j\Delta) + Y_{j\Delta}) \cdot X_{(j+1)\Delta} + \sum_{j=0}^{T_B/\Delta-1} \frac{\|\dot{z}(j\Delta) + Y_{j\Delta}\|_2^2}{2}
$$

Interpreting the Change-of-Measure

- **Importance Sampling** Q**:** W_t is i.i.d. N(0,1) under Q

$$
\begin{aligned}
Y_{t+\Delta} - Y_t &= \dot{z}(t)\,\Delta + \Delta W_{t+\Delta} \implies \text{Fluid } dy_t \approx \dot{z}(t)\,dt \\
Y_0 &= (1,1), \\
X_{t+\Delta} &= W_{t+\Delta} + (\dot{z}(t) + Y_t).
\end{aligned}
$$

- **Likelihood ratio** dP/dQ**:**

$$
L_0 = \exp\left(- \sum_{j=0}^{T_B/\Delta - 1} (\dot{z}(j\Delta) + Y_{j\Delta}) \cdot X_{(j+1)\Delta} + \sum_{j=0}^{T_B/\Delta - 1} \frac{\|\dot{z}(j\Delta) + Y_{j\Delta}\|_2^2}{2}\right.
$$

- **Importance sampling estimator:**

$$
\text{IS Estimator} = L_0 \times I(T_B < T_A)
$$

Agenda

- Introduction
- A Simple Random Walk Example
- Systematic Approach
- **Testing Efficiency**
- Counter-examples, Heavy-tails and Beyond

Another Interpretation

- **Importance Sampling** Q**:** W_t is Brownian motion under Q

$$
\begin{aligned}
Y_{t+\Delta} - Y_t &= Y_t \Delta + \Delta W_{t+\Delta} \Longrightarrow \text{Fluid } dy(t) \approx y(t)\, dt, \\
Y_0 &= \Delta (1,1)^T, \\
X_{t+\Delta} &= W_{t+\Delta} + 2Y_t.
\end{aligned}
$$

Another Interpretation

- **Importance Sampling** Q**:** W_t is Brownian motion under Q

$$
\begin{aligned}
Y_{t+\Delta} - Y_t &= Y_t \Delta + \Delta W_{t+\Delta} \Longrightarrow \text{Fluid } dy(t) \approx y(t)\, dt, \\
Y_0 &= \Delta (1,1)^T, \\
X_{t+\Delta} &= W_{t+\Delta} + 2Y_t.
\end{aligned}
$$

- **Likelihood ratio** dP/dQ**:**

$$
L = \exp\left(- \sum_{j=0}^{T_B/\Delta - 1} 2Y_{(j\Delta)} \cdot X_{(j+1)\Delta} + \sum_{j=0}^{T_B/\Delta - 1} 2 \left\| Y_{(j\Delta)} \right\|_2^2 \right)
$$

Another Interpretation

- **Importance Sampling** Q**:** W_t is Brownian motion under Q

$$\begin{aligned} Y_{t+\Delta} - Y_t &= Y_t \Delta + \Delta W_{t+\Delta} \Longrightarrow \text{Fluid } dy(t) \approx y(t)\, dt, \\ Y_0 &= \Delta (1,1)^T, \\ X_{t+\Delta} &= W_{t+\Delta} + 2Y_t. \end{aligned}$$

- **Likelihood ratio** dP/dQ**:**

$$L = \exp\left(- \sum_{j=0}^{T_B/\Delta - 1} 2Y_{(j\Delta)} \cdot X_{(j+1)\Delta} + \sum_{j=0}^{T_B/\Delta - 1} 2 \left\| Y_{(j\Delta)} \right\|_2^2 \right)$$

- **Importance sampling estimator:**

$$\text{IS Estimator} = L \times I\left(T_B < T_A\right)$$

Asymptotic Optimality

- **First note:** $Y_{t+\Delta} - Y_t = -\Delta Y_t + \Delta X_{t+\Delta} \Longrightarrow$

$$\frac{\left\| Y_{(j+1)\Delta} \right\|_2^2 - \| Y_{j\Delta} \|_2^2}{\Delta} = -2 \| Y_{j\Delta} \|_2^2 + 2 Y_{(j\Delta)} \cdot X_{(j+1)\Delta} + \Delta \left\| X_{(j+1)\Delta} - Y_{j\Delta} \right\|_2^2 .$$

Asymptotic Optimality

- **First note:** $Y_{t+\Delta} - Y_t = -\Delta Y_t + \Delta X_{t+\Delta} \Longrightarrow$

$$\frac{\left\| Y_{(j+1)\Delta} \right\|_2^2 - \| Y_{j\Delta} \|_2^2}{\Delta} = -2 \| Y_{j\Delta} \|_2^2 + 2 Y_{(j\Delta)} \cdot X_{(j+1)\Delta} + \Delta \left\| X_{(j+1)\Delta} - Y_{j\Delta} \right\|_2^2 .$$

- **Get likelihood ratio representation:**

$$L = \exp\left(- \sum_{j=0}^{T_B/\Delta - 1} 2 Y_{(j\Delta)} \cdot X_{(j+1)\Delta} + \sum_{j=0}^{T_B/\Delta - 1} 2 \left\| Y_{(j\Delta)} \right\|_2^2 \right)$$

$$= \exp\left(- \| Y_{T_B} \|_2^2 / \Delta + \| y_0 \|_2^2 / \Delta + \Delta \sum_{j=0}^{T_B/\Delta - 1} \left\| W_{(j+1)\Delta} + Y_{j\Delta} \right\|_2^2 \right)$$

Asymptotic Optimality

- **Bound Likelihood ratio:**

$$L \leq \exp\left(-\inf_{y \in B} \|y\|_2^2 / \Delta + \|y_0\|_2^2 / \Delta + O_p(1)\right)$$

Asymptotic Optimality

- **Bound Likelihood ratio:**

$$L \leq \exp\left(-\inf_{y\in B}\|y\|_2^2/\Delta + \|y_0\|_2^2/\Delta + O_p(1)\right)$$

- **Second moment of estimator:**

$$E^Q[L^2 \times I(T_B < T_A)] = \exp\left(-2\left(\inf_{y\in B}\|y\|_2^2 - \|y_0\|_2^2\right)/\Delta + o(1/\Delta)\right)$$

Asymptotic Optimality

- **Bound Likelihood ratio:**

$$L \leq \exp\left(-\inf_{y\in B} \|y\|_2^2 / \Delta + \|y_0\|_2^2 / \Delta + O_p(1)\right)$$

- **Second moment of estimator:**

$$E^Q[L^2 \times I(T_B < T_A)] = \exp\left(-2\left(\inf_{y\in B} \|y\|_2^2 - \|y_0\|_2^2\right) / \Delta + o(1/\Delta)\right)$$

- **Asymptotic Optimality follows since:**

$$P_{y_0}(T_B < T_A) = \exp\left(-\left(\inf_{y\in B} \|y\|_2^2 - \|y_0\|_2^2\right) / \Delta + o(1/\Delta)\right).$$

Agenda

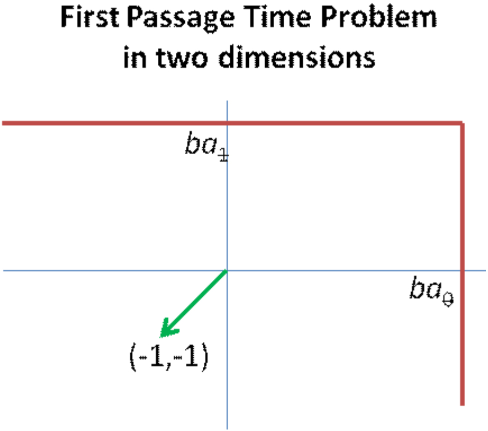
- Introduction
- A Simple Random Walk Example
- Systematic Approach
- Efficiency Argument
- **Counter-examples, Heavy-tails and Beyond**

Counter-examples

- **Model:** $Y_t = (-1,-1)t + X_t$; X_t is Brownian motion & $B_b = \{(x,y) : x \geq a_0 b \text{ or } y \geq a_1 b\}$

Counter-examples

- **Model:** $Y_t = (-1,-1)t + X_t$; X_t is Brownian motion & $B_b = \{(x,y) : x \geq a_0 b \text{ or } y \geq a_1 b\}$
- **Estimate:** $u(b) = P(T_{B_b} < \infty)$ as $b \nearrow \infty$



Counter-examples

- **Known result:** As $b \nearrow \infty$

$$u(b) = \exp(-2b\min\{a_0, a_1\})(1 + o(1)).$$

Counter-examples

- **Known result:** As $b \nearrow \infty$

$$u(b) = \exp(-2b\min\{a_0, a_1\})(1 + o(1)).$$

- **Optimal Path:**

$$y(t) = \begin{cases} t \times (1, -1) & \text{if} \quad a_0 < a_1 \\ t \times (-1, 1) & \text{if} \quad a_1 < a_0 \end{cases}.$$

Counter-examples

- **Known result:** As $b \nearrow \infty$

$$u(b) = \exp\left(-2b\min\{a_0, a_1\}\right)(1 + o(1)).$$

- **Optimal Path:**

$$y(t) = \begin{cases} t \times (1, -1) & \text{if } a_0 < a_1 \\ t \times (-1, 1) & \text{if } a_1 < a_0 \end{cases}.$$

- **Let us assume** $a_0 < a_1$**...**

Counter-examples

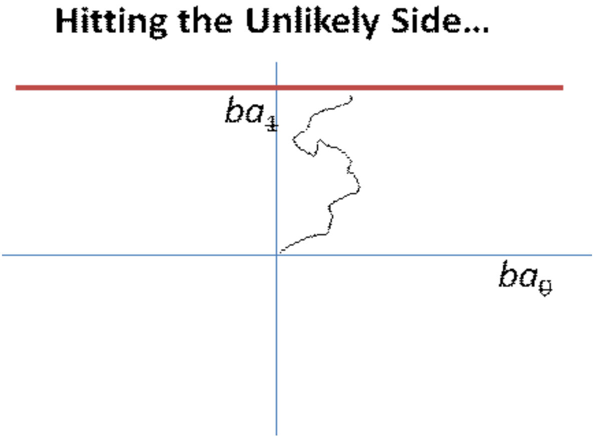
- **Change-of-measure Q_0:** Brownian Motion with drift $(1, -1)$.

Second Moment of Estimator:

$$\begin{aligned} \text{2nd Moment} &= E^{Q}[\exp\left(-(4,0)\cdot X_{T_{B_b}}\right) I\left(T_{B_b} < \infty\right)] \\ &= E[\exp\left(-(2,0)\, X_{T_{B_b}}\right) I\left(T_{B_b} < \infty\right)]. \end{aligned}$$

Counter-examples

- **What happens if a path exits $A_b = \{(x, y) : y \geq a_1 b\}$?**



Counter-examples

- **Lower bound:**

$$\begin{aligned} \text{2nd Mnt} &\geq E\left[\exp\left(-2X^{(1)}_{T_{A_b}}\right) I\left(T_{B_b} < \infty, T_{A_b} = T_{B_b}\right)\right] \\ &= E^{Q_1}\left[\exp\left(-2X^{(1)}_{T_{A_b}} - 2a_1 b\right) I\left(T_{A_b} = T_{B_b}\right)\right], \end{aligned}$$

where Q_1 is change-of-measure yielding Brownian Motion with drift $(-1, 1)$.

Counter-examples

- **Lower bound:**

$$\begin{aligned} \text{2nd Mnt} &\geq E\left[\exp\left(-2X^{(1)}_{T_{A_b}}\right) I\left(T_{B_b} < \infty, T_{A_b} = T_{B_b}\right)\right] \\ &= E^{Q_1}\left[\exp\left(-2X^{(1)}_{T_{A_b}} - 2a_1 b\right) I\left(T_{A_b} = T_{B_b}\right)\right], \end{aligned}$$

  where Q_1 is change-of-measure yielding Brownian Motion with drift $(-1, 1)$.

- Note that $X^{(1)}_{T_{A_b}} \approx -a_0 a_1 b$ and $T_{A_b} = T_{B_b}$ under Q_1

Counter-examples

- **Lower bound:**

$$\begin{aligned} \text{2nd Mnt} &\geq E\left[\exp\left(-2X^{(1)}_{T_{A_b}}\right) I\left(T_{B_b} < \infty, T_{A_b} = T_{B_b}\right)\right] \\ &= E^{Q_1}\left[\exp\left(-2X^{(1)}_{T_{A_b}} - 2a_1 b\right) I\left(T_{A_b} = T_{B_b}\right)\right], \end{aligned}$$

where Q_1 is change-of-measure yielding Brownian Motion with drift $(-1, 1)$.

- Note that $X^{(1)}_{T_{A_b}} \approx -a_0 a_1 b$ and $T_{A_b} = T_{B_b}$ under Q_1
- **Consequently:** 2nd Moment at least roughly

$$\exp\left(-2a_1 b\left(1 - a_0\right)\right)$$

Can easily pick $a_0 < a_1$ to **break asymptotic optimality and even get HUGE variance**!!

Heavy-tails...

- **Model: Random walk...**

$$S_n = X_1 + ... + X_n,$$

X_i's i.i.d. heavy tailed

Heavy-tails...

- **Model: Random walk...**

$$S_n = X_1 + \ldots + X_n,$$

X_i's i.i.d. heavy tailed

- **Typical example: Regular variation**

$$P\left(X_i > t\right) = L\left(t\right) t^{-\alpha}$$

for $\alpha > 0$ and $L\left(t\beta\right)/L\left(t\right) \longrightarrow 1$ as $t \to \infty$ for $\beta > 0$ (e.g. $L\left(t\right) = \log(1+t)$).

Heavy-tails...

- **Model: Random walk...**

$$S_n = X_1 + ... + X_n,$$

X_i's i.i.d. heavy tailed

- **Typical example: Regular variation**

$$P(X_i > t) = L(t)\, t^{-\alpha}$$

for $\alpha > 0$ and $L(t\beta)/L(t) \longrightarrow 1$ as $t \to \infty$ for $\beta > 0$ (e.g. $L(t) = \log(1 + t)$).

- Note $E \exp(\theta X_1) = \infty$ for $\theta > 0$.

Heavy-tails...

- **Consider estimating:** $P\left(T_b < \infty\right),\ T_b = \inf\{n \geq 0 : S_n > b\}$

Heavy-tails...

- **Consider estimating:** $P(T_b < \infty)$, $T_b = \inf\{n \geq 0 : S_n > b\}$
- X_i's i.i.d. reg. varying & $EX_i < 0$

Heavy-tails...

- **Consider estimating:** $P(T_b < \infty)$, $T_b = \inf\{n \geq 0 : S_n > b\}$
- X_i's i.i.d. reg. varying & $EX_i < 0$
- Turns out that

$$P(X_j \leq y | T_b < \infty) \longrightarrow P(X_j \leq y)$$

Heavy-tails...

- **Consider estimating:** $P(T_b < \infty)$, $T_b = \inf\{n \geq 0 : S_n > b\}$
- X_i's i.i.d. reg. varying & $EX_i < 0$
- Turns out that

$$P(X_j \leq y \,|\, T_b < \infty) \longrightarrow P(X_j \leq y)$$

- **No clear way to mimic zero variance change-of-measure!**

Heavy-tails...

- **Consider estimating:** $P(T_b < \infty)$, $T_b = \inf\{n \geq 0 : S_n > b\}$
- X_i's i.i.d. reg. varying & $EX_i < 0$
- Turns out that

$$P\left(X_j \leq y \middle| T_b < \infty\right) \longrightarrow P\left(X_j \leq y\right)$$

- **No clear way to mimic zero variance change-of-measure!**
- **No clear way to apply the systematic approach!**

Solution? State-dependent Importance Sampling

As we shall see these examples can be addressed using

$$\textbf{State} - \textbf{dependent importance sampling}$$

which we will study in the second part of this lecture...

Summary and Conclusions

- Importance sampling is a useful techinque for rare-event estimation

Summary and Conclusions

- Importance sampling is a useful techinque for rare-event estimation
- Zero-variance importance sampler = conditional distribution given rare event

Summary and Conclusions

- Importance sampling is a useful techinque for rare-event estimation
- Zero-variance importance sampler = conditional distribution given rare event
- Light-tailed problems: Fundamental family given by exponential tiltings

Summary and Conclusions

- Importance sampling is a useful techinque for rare-event estimation
- Zero-variance importance sampler = conditional distribution given rare event
- Light-tailed problems: Fundamental family given by exponential tiltings
- Large deviations helps approximate the conditional distribution given rare event

Summary and Conclusions

- Importance sampling is a useful techinque for rare-event estimation
- Zero-variance importance sampler = conditional distribution given rare event
- Light-tailed problems: Fundamental family given by exponential tiltings
- Large deviations helps approximate the conditional distribution given rare event
- Optimal path in large deviations dictates tilting (several ways of interepreting)

Summary and Conclusions

- Importance sampling is a useful techinque for rare-event estimation
- Zero-variance importance sampler = conditional distribution given rare event
- Light-tailed problems: Fundamental family given by exponential tiltings
- Large deviations helps approximate the conditional distribution given rare event
- Optimal path in large deviations dictates tilting (several ways of interepreting)
- Approach fails (badly!) in non-convex problems and in heavy-tailed situations