Improving the Convergence of Iterative Importance Sampling for Computing Upper and Lower Expectations

Thomas Fetz
Unit for Engineering Mathematics
University of Innsbruck, Austria
Thomas.Fetz@uibk.ac.at

11th International Symposium on Imprecise Probabilities: Theories and Applications
July 3–6, 2019, Ghent, Belgium
Problem statement

Given:
- Function \( h : A \subseteq \mathbb{R}^d \rightarrow \mathbb{R} : x \rightarrow h(x) \) which is \textbf{expensive} to evaluate.
- Family \( \{f_t\}_{t \in \mathcal{T}} \) of density functions \( f_t \).

Aim:
- Computation of \textbf{lower} / \textbf{upper} expectations = solving optimisation problems

\[
\theta^* = \min_{t \in \mathcal{T}} \theta(t), \quad \theta^* = \max_{t \in \mathcal{T}} \theta(t), \quad \text{objective function} \quad \theta(t) = \int_A h(x) f_t(x) \, dx.
\]
- In engineering: \textbf{Lower} / \textbf{upper probabilities of failure} for

\[
h(x) = \mathbb{1}_{g(x) < 0}(x) \quad (g \text{ limit state function, } g(x) < 0 \text{ means failure}).
\]

Method:
- \textbf{Efficient approximation / estimate} \( \hat{\theta} \) of the objective function \( \theta \)
  using \textbf{importance sampling / reweighting} techniques.
- \textbf{Fixed point iterations} to improve estimates \( \hat{\theta}^* \) of the lower expectation \( \hat{\theta}^* \).
- \textbf{Weighted combinations of previous results of fixed point iteration} to improve convergence.
Simple numerical example

- **Function h:**
  \[ h(x) = 1_D(x), \quad D = (-\infty, -2) \cup [2, \infty). \]

- **Family of density functions:**
  \[ f_t \sim \mathcal{N}(\mu(t), \sigma^2(t)), \quad t \in \mathcal{T} = [-7, 7] \]
  mean \( \mu(t) = t \), variance \( \sigma^2(t) = 4 \).

- **Exact result:** \( \theta_* = 0.3173 \) at \( t_* = 0 \).

---

**Monte Carlo simulation**

1. **Set of random numbers**
   \[ \Omega_t = \{ U_1, U_2, \ldots, U_n \}, \quad U_i = (V_i, W_i) \]
   \( V_i, W_i \sim \text{Uniform}(0, 1) \)

2. **Sample points**
   \[ x_t(U_i) = \mu(t) + \sigma(t) \cdot \sqrt{-2 \ln V_i \cdot \cos(2\pi W_i)} \sim \mathcal{N}(\mu(t), \sigma^2(t)) \]

3. **Monte Carlo simulation w.r.t. \( \Omega_t \):**
   \[ \hat{\theta}_\Omega_t(t) = \frac{1}{n} \sum_{i=1}^{n} 1_D(x_t(U_i)) \]
   estimate of \( \theta(t) \).

---

**Diagram:**
- Graph of \( f_t \) and \( D \) with labeled axes.
- Graph of \( \theta \) with marked value at \( t_* = 0 \).
Example & Monte Carlo simulation

Simple numerical example

- **Function** $h$: 
  
  $$h(x) = \mathbb{1}_D(x), \quad D = (-\infty, -2] \cup [2, \infty).$$

- **Family of density functions**: 
  
  $$f_t \sim \mathcal{N}(\mu(t), \sigma^2(t)), \quad t \in \mathcal{T} = [-7, 7]$$ 

  mean $\mu(t) = t$, variance $\sigma^2(t) = 4$.

- **Exact result**: $\theta^* = 0.3173$ at $t^* = 0$.

Monte Carlo simulation, three steps

1. **Set of random numbers** $\Omega_t = \{U_1, U_2, \ldots, U_n\}$, $U_i = (V_i, W_i)$, $V_i, W_i \sim \mathcal{U}([0, 1])$ i.i.d.

2. **Sample points** $x_t(U_i) = x_t(V_i, W_i) = \mu(t) + \sigma(t) \cdot \sqrt{-2\ln V_i} \cdot \cos(2\pi W_i) \sim \mathcal{N}(\mu(t), \sigma^2(t))$.
   
   (transformation of $\Omega_t$, Box-Muller)

3. **Monte Carlo simulation w.r.t.** $\Omega_t$: 
   
   $$\hat{\theta}_{\Omega_t}(t) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_D(x_t(V_i, W_i)), \quad \text{estimate of } \theta(t).$$

   (expensive evaluation of $h$)
Approach 1: Different sets of random numbers for each parameter value $t$

- **Different** sets of $n$ random numbers $U_i = (V_i, W_i)$ for each parameter value $t$.

- **Different** sets of $n$ sample points $x_i(V_i, W_i)$ for each parameter value $t$.

- **Noisy** estimate (approximation) $\hat{\theta}_{\Omega(\cdot)}$ of $\theta$ using MC.

- **Difficult** to solve the optimisation problem.
- **Expensive** because of the evaluation of expensive $h$ for different sets of sample points in the optimisation algorithm.
Approach 2: One single set of random numbers for all parameter values $t$

One single set of $n$ random numbers $U_i = (V_i, W_i)$ for all parameter values $t$.

Different sets of $n$ sample points $x_t(V_i, W_i)$ for each parameter value $t$.

Estimate $\hat{\theta}_\Omega$ of $\theta$ is a step function.

- Difficult to solve the optimisation problem.
- Expensive because of the evaluation of expensive $h$ for different sets of sample points in the optimisation algorithm.
Approach 3: One single set of random numbers & One single set of sample points

\[ \int_A 1_D(x) f_t(x) \, dx = \int_A 1_D(x) \frac{f_t(x)}{f_s^R(x)} f_s^R(x) \, dx. \]

One single set of random numbers \( U_i = (V_i, W_i) \) for all \( t \).

One single set of sample points \( x_t(V_i, W_i) \) for all \( t \).

Reweighting the sample points!

Estimate \( \hat{\theta}_{\Omega,s=3} \) is cheap to evaluate and continuous. Easy to solve the optimisation problem.

Importance sampling / reweighting

- **Importance sampling density** \( f_s^R \) for density \( f_s, s \in \mathcal{T} \).
  (here in the example: \( f_s^R := f_s \) and \( s = 3 \))

- **Weights** \( w_{st}(x) = f_t(x)/f_s^R(x) \).
  (classical: \( w_s(x) = f_s(x)/f_s^R(x) \))

- **MC**: \( \hat{\theta}_{\Omega,s}(t) = \frac{1}{n} \sum_{i=1}^{n} 1_D(x_s(V_i, W_i)) \cdot w_{st}(x_s(V_i, W_i)). \)
  (classical importance sampling if \( s = t \))

- **Normalisation**: Divide by \( \sum_{i=1}^{n} w_{st}(x_s(V_i, W_i)) \).

Bad approximation \( \hat{\theta}_{\Omega,s}(t) \) of \( \hat{\theta}_{\Omega}(t) \) and \( \theta(t) \) for \( t \) far from \( s \)!
Function $\hat{\theta}_\Omega(s,t) := \hat{\theta}_{\Omega,s}(t)$ for all $s \in \mathcal{T}$. 

Fixed point problem!
Improvement of $\hat{\theta}_{\Omega}^*$ by fixed point iteration

$$s^{(k+1)} = \tau_{\Omega}(s^{(k)}) = \arg\min_{t \in \mathcal{T}} \hat{\theta}_{\Omega,s^{(k)}}(t)$$

**Improving convergence of fixed point iteration**

1. Increased reweighting sample coverage, e.g. increased variance.
2. Design point method (for engineering problems).
3. Weighted combination of previous results of iteration, complementary to (1+2).
Weighted combination of previous results of fixed point iteration

Observations
- Estimates $\hat{\theta}_{\Omega,s^{(i)}}(t)$ are **bad for $t$ far from $s^{(i)}$** → wrong minimum at $\tau_{\star\Omega}(s^{(i)})$ far from $s^{(i)}$ → leading away from fixed point → circling.
- Exact function $\vartheta$ is **constant** in $s$-direction.

Idea
- **Weighted combination** of previous results $\hat{\theta}_{\Omega,s^{(i)}}$.
- **High** weights for $t$ **close to** $s^{(i)}$ (good estimates).
- **Low** weights for $t$ **far from** $s^{(i)}$ (bad estimates).

New iteration scheme using weighted combination

$$s^{(k+1)} = \tau_{\star\Omega}^{(k)}(s^{(k)}, \ldots, s^{(1)}) = \arg\min_{t \in \mathcal{T}} \sum_{i=1}^{k} \phi^{(k)}_{s^{(i)}}(t) \cdot \hat{\theta}_{\Omega,s^{(i)}}(t), \quad \phi^{(k)}_{s^{(i)}}(t) \geq 0 \quad \text{and} \quad \sum_{i=1}^{k} \phi^{(k)}_{s^{(i)}}(t) = 1.$$
Weighted combination of previous results of fixed point iteration

Combination of $\hat{\theta}_{\Omega,s^{(1)}}$ and $\hat{\theta}_{\Omega,s^{(2)}}$ resulting in $\hat{\theta}_{\Omega,s^{(2)}}^{(2)}$, $\hat{\theta}_\Omega$.

Updated $\hat{\varphi}_\Omega^{(k+1)}(s,t) = \varphi_s^{(k+1)}(t) \cdot \hat{\theta}_{\Omega,s}(t) + \sum_{i=1}^k \varphi_s^{(k+1)}(t) \cdot \hat{\theta}_{\Omega,s(i)}(t)$ and iteration path.
Please visit my poster for more details and numerical examples!

Thank you for your attention!


