Importance Sampling by High-Dimensional Embedding

Chuan-Hsiang Han [1]   Erik Che-hsiu Cheng [2]

[1]Department of Quantitative Finance, National Tsing-Hua University, Taiwan

[2]Department of Mathematics, National Taiwan University, Taiwan

Motivation

Figure 1: In engineering, probability of failure can measure **system performance**.

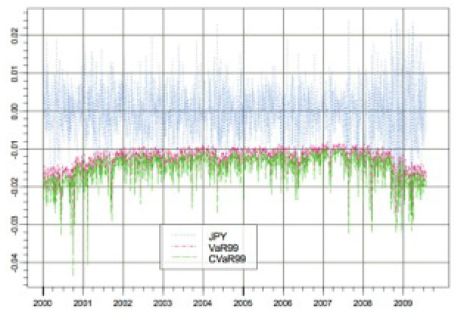


Figure 2: In finance, probability of default can measure **downside risk**.

Common problem: How to efficiently compute

$$\mathbb{P}(g(X) < c) = \mathbb{E}\{\mathbb{I}(g(X) < c)\},$$

where $X \sim \mathcal{N}(\vec{0}, \Sigma)$ and $c \in \Re$.

Methodology: High-Dimensional Embedding IS

A new procedure with three phases:

P1. embeds the evaluation problem into a high-dimensional space.

P2. estimates lower tail probabilities by **efficient** importance samplings.

P3. projects those associated probability measures with some marginal condition.

This procedure leads to an **entropy minimization problem with constraints**, well studied in the field of **machine learning**.

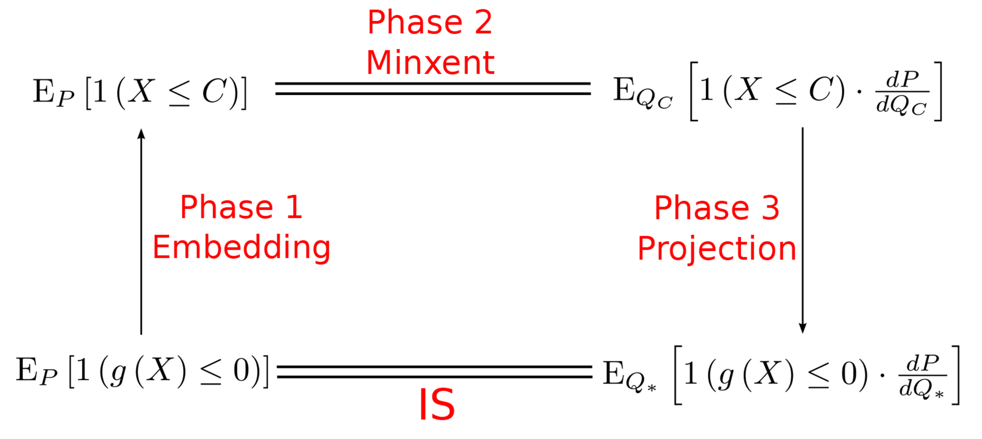


Figure 3: Three phases of high-dimensional embedding IS

Phase 2: Lower Tail Probability

- Gaussian Copula It's joint default probability is defined by

$$P = \mathbb{E}\{\mathbb{I}(Z \leq C)\}, \; Z = (Z_1, \cdots, Z_n)^T \sim N(0, \Sigma)$$

- Importance Sampling (**IS**) by Exponential Twist:

$$P_1 = \mathbb{E}_\mu \left\{ \mathbb{I}(Z < C) \cdot \frac{f(Z)}{f_\mu(Z)} \right\}$$

where

$f_\mu(z) = \dfrac{\exp(\mu \cdot z) f(z)}{M_Z(\mu)}$, and $M_Z(\mu)$ denotes m.g.f. of Z.

Asymptotic Zero Variance: Gaussian

$$P_1(\alpha) = \tilde{\mathbb{E}} \left\{ \mathbb{I}\left(X < \sqrt{\alpha} C\right) \exp\left(-\sqrt{\alpha} C^T \Sigma^{-1} X + \frac{\alpha}{2} C^T \Sigma^{-1} C \right) \right\}$$

where $X \sim \mathcal{N}(\sqrt{\alpha}C, \Sigma)$ under $\tilde{\mathbb{P}}$. $P_2(\alpha)$ denotes the second moment of the IS estimator with scale α.

Theorem

Efficient (Asymptotically Optimal) Importance Sampling by Large Deviation

$$\lim_{\alpha \to \infty} \frac{1}{\alpha} \log P_2(\alpha) = 2 \lim_{\alpha \to \infty} \frac{1}{\alpha} \log P_1(\alpha) = -C^T \Sigma^{-1} C.$$

Note: $variance(\alpha) = P_2(\alpha) - P_1^2(\alpha) \approx 0$ for α large enough.

Embedding IS: Constrained Relative Entropy Minimization Problem

$$\inf_{\tilde{P} \in \mathcal{E}} \left[\mathbb{H}\left(\tilde{P} \middle| P\right) \text{ s.t. } E_{\tilde{P}}[\mathbf{X}] \in D \right], \qquad (1)$$

where $\mathcal{E} = \{Q : Q \ll P\}$, $\mathbb{H}(\cdot \mid \cdot)$ denotes the relative entropy (Kullback-Leibler divergence), and $D = \left\{ C \in R^n : g(\vec{C}) = c \right\}$.

Variational analysis by Altun and Smola [1] and Koyejo and Ghosh [4].

Numerical Results

Gaussian Tail Probability Estimation

EIS vs. **Matlab code-mvncdf.m**

	Basic MC		EIS		mvncdf.m	
n	*Mean*	*SE*	*Mean*	*SE*	*Mean*	*SE*
5	2.67E-05	1.89E-05	1.37E-05	3.31E-07	1.40E-05	4.31E-08
10	-	-	2.12E-07	1.08E-08	1.98E-07	4.41E-09
15	-	-	1.59E-08	1.30E-09	1.49E-08	5.20E-10
20	-	-	2.41E-09	2.64E-10	2.11E-09	1.18E-10
25	-	-	5.64E-10	6.76E-11	5.39E-10	9.07E-11
30	-	-	2.01E-10	4.03E-11	-	-
50	-	-	3.84E-12	1.53E-12	-	-
75	-	-	4.18E-13	1.83E-13	-	-
100	-	-	6.99E-14	3.41E-14	-	-

EIS stands for efficient importance sampling. The total number of simulations=75,000. Averaged CPU time: 1.47E-01(s), 1.50E-01 (s), 1.75E-01 (s), not including dimensions beyond 25.

Table 1: Result of Monte Carlo Simulation under multivariate normal distribution (sample size = 2M, d=40.)[3]

		GPU	CPU	Speed up
BMC	Mean	1.50e-6	1.50e-6	
	SE	8.7e-7	8.7e-7	
	Time	0.05 (s)	1.77 (s)	X36
EIS	Mean	2.00e-6	2.01e-6	
	SE	1.40e-8	1.39e-8	
	Time	0.10 (s)	1.77 (s)	X18
Accuracy	Variance Reduction Ratio	X3861	X3918	

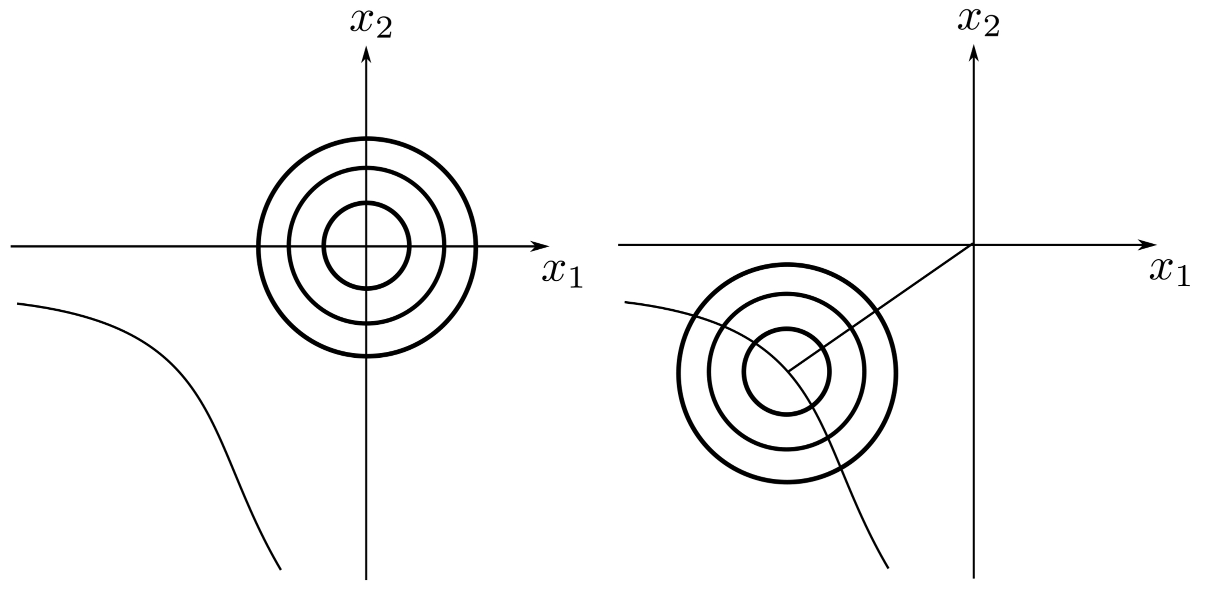
BMC stands for the basic Monte Carlo method. EIS stands for efficient importance sampling.

Total Reduction in time and variance:

(GPU+EIS) / (CPU+BMC) ≈ 68000 faster!

Application I: Risk Management. FORM IS

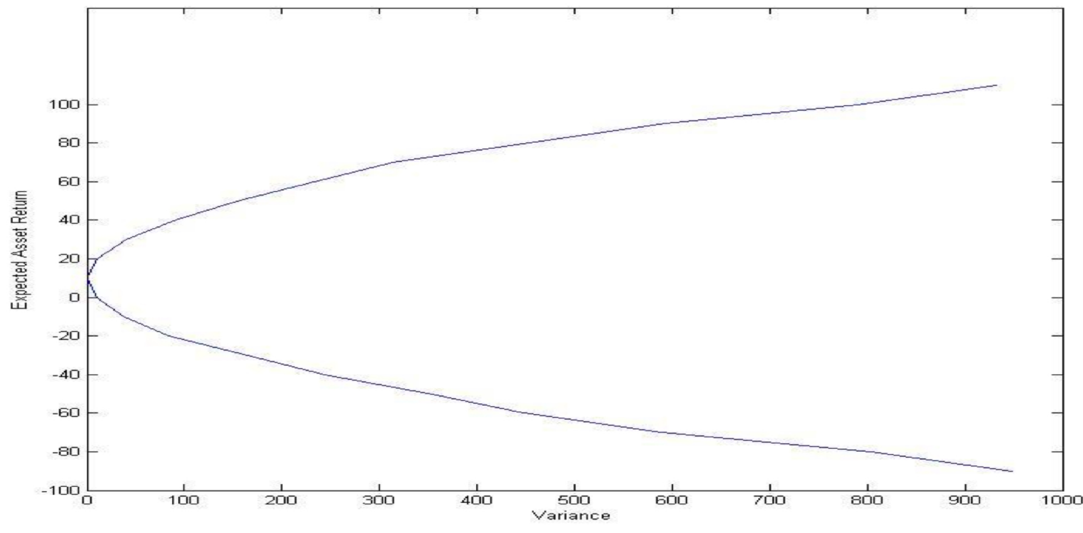
- used for structural reliability in engineering.
- Value-at-Risk estimation, **US patterned** by De and Tamarchenko[2].
- In the next theorem, we prove that FORM IS is a special case of our proposed embedding IS.



Theorem: FORM IS as Embedding IS

Let X be a standard multivariate normal defined on $(\Omega, \mathcal{F}, P)$ and Q^* denotes the inf-argument of problem (1). Then Q^* is a multivariate normal distribution with independent components centered at a design point.

Application II: Optimal Investment. Efficient Frontier of 100 Stocks



Contribution and Impact

- A new class of IS algorithms is proposed. Advantages are (1) scalability, (2) asymptotic optimality, (3) GPU parallel computing.
- FORM (Design point) IS, a US patterned method for VaR estimation, is a special case of embedding IS.
- Background theory links deeply to machine learning.

References

[1] Y. Altun and A. J. Smola. *Unifying divergence minimization and statistical inference via convex duality.* COLT, p. 139-153. 2006.

[2] R. De and T. Tamarchenko. *System and method for determining Value-at-Risk using FORM/SORM.* Patent US 20030061152 A1.

[3] C.-H. Han and Y.-T. Lin. *Accelerated variance reduction methods on GPU.* Submitted.

[4] O. Koyejo and J. Ghosh. *A Representation Approach for Relative Entropy Minimization with Expectation Constraints.* ICML Workshop on Divergences and Divergence Learning (WDDL), 2013.