Importance Sampling by High-Dimensional Embedding

C.-H. Sean Han
Dept. of Quantitative Finance
Natl. Tsing-Hua University, Taiwan

UCSB

Sep. 26, 2014
Outline

- High-dimensional embedding IS
- Mathematical Formulation as an Entropy Minimization Problem
- FORM IS as a special case
Start From an Example:
Portfolio Default Probability Estimation

\[ p = \mathbb{E}\{\mathbb{I}(f(X) < c)\}. \]

E.X. the value of a portfolio \( f(X) = \sum_{i=1}^{n} \alpha_i X_i \in \mathbb{R} \), where \( X \) is defined under \( \{\Omega, \mathcal{F}, P\} \).

Importance sampling can be efficient for rare event simulation by

\[ p = \tilde{\mathbb{E}} \left\{ \mathbb{I}(f(X) < c) \frac{dP}{d\tilde{P}} \right\}. \]
Review of IS: Variance Minimization

1. Asymptotic Approximation
   - perturbation method: Fournie-Lebuchoux-Touzi(97), Fouque-H(04),...
   - large deviation principle: Glasserman-Heidelberger-Shababuddin(99), Guasoni-Robertson(08),...

2. Adaptive Scheme
   - cross (relative) entropy: Rubinstein-Kroese(04),...
   - fixed point: Arouna(03), Joudain-Lelong(09),...
3. **Mixed Scheme**
   - Cannamela, Garnier, and Iooss (08), Dingec and Hormann (11),...

4. **Alternatives**
   FORM IS* (or called Design Point IS):
   Melchers (99), De and Tamarchenko (01),...

*Thanks to Bernard Lapeyre
High-Dimensional Embedding IS

A new procedures with three phases:

P1. embeds the evaluation problem into a high-dimensional space.
P2. estimates the occurrence of tail events 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.
Procedure of Importance Sampling by High-Dimensional Embedding

\[ \mathbb{E}\{\mathbb{I}(\bar{X} \leq \bar{C})\} = \mathbb{E}\left\{ \mathbb{I}(\bar{X} \leq C) \frac{dP}{dP(C)} \right\} \]

P2: I.S.

P1: Embedding

\[ \mathbb{E}\{\mathbb{I}(f(\bar{X}) \leq c)\} = \tilde{\mathbb{E}}\{\mathbb{I}(f(\bar{X}) \leq c) \frac{dP}{dP} \} \]

I.S.

P3: Projection
Phase 2: Lower Tail Probability

\[ P = \mathbb{P}(\tau_1 \leq T, \tau_2 \leq T, \ldots, \tau_n \leq T) \]

- **Gaussian Copula**

\[ \{\tau_i \leq T\} = \{Z_i \leq c_i := \Phi^{-1}(F_i(T))\}. \]
\[ P = \mathbb{E}\{\mathbb{I}(Z \leq C)\}, Z = (Z_1, \ldots, Z_n)^T \sim N(0, \Sigma) \]

- **Student's t Copula**

\[ \{\tau_i \leq T\} = \{S_i \leq c_i := t_{\nu}^{-1}(F_i(T'))\}. \]
\[ P = \mathbb{E}\{\mathbb{I}(S \leq C)\}, \]
\[ S = (S_1, \ldots, S_n)^T = \frac{Z}{\sqrt{Y(\nu)/\nu}} \sim T_{\Sigma, \nu} \]
Importance Sampling: Exponential Twist

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

where

\[ f_\mu(z) = \frac{\exp(\mu \cdot z) f(z)}{M_Z(\mu)}. \]

Cherny and Maslov (04) proved that the parameter vector \( \mu \) can be solved from an entropy minimization problem with a linear constraint.
Asymptotic Zero Variance: Gaussian

\[ P_1(\alpha) = \tilde{\mathbb{E}} \{ I(X < \sqrt{\alpha} C) \exp(2\sqrt{\alpha} C^T \Sigma^{-1} X + \alpha C^T \Sigma^{-1} C) \}, \]

where \( X \sim \mathcal{N}(\sqrt{\alpha} C, \Sigma) \) under \( \tilde{\mathbb{P}} \).

**Theorem:** Efficient Importance sampling

\[ \lim_{\alpha \to \infty} \frac{1}{\alpha} \log P_2(\alpha) = 2 \quad \lim_{\alpha \to \infty} \frac{1}{\alpha} \log P_1(\alpha) = -C^T \Sigma^{-1} C. \]
Gaussian Tail Probability Estimation: IS vs. mvncdf.m

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

\(c = -2, \rho = 0.5\), and the total number of simulations=75000. Averaged CPU time: 1.47E-01, 1.50E-01, 1.75E-01, not including dimensions beyond 25.
Asymptotic Zero Variance: Student t

\[ P_{1,IS}^t(C, \nu) = \mathbb{E} \left[ I(\tilde{X} < 0) \frac{(1 + \frac{C^T \Sigma^{-1} C}{\nu})^{-\nu/2}}{\exp(-\Sigma^{-1} C \tilde{X})} \right] \]

where \( \tilde{X} = \frac{Y}{\nu} \left( \frac{Z}{\sqrt{Y/\nu}} - C \right) \).

**Thm:** Efficient Importance sampling scheme.

\[
\lim_{\alpha \to \infty} \frac{1}{\ln \alpha} \ln P_{2,IS}^t(\sqrt{\alpha} C, \nu) = 2 \times \lim_{\alpha \to \infty} \frac{1}{\ln \alpha} \ln P_{1}^t(\sqrt{\alpha} C, \nu) = -\nu
\]
Student $t$ Tail Probability Estimation:

IS vs. mvtcdf.m

<table>
<thead>
<tr>
<th>n</th>
<th>Basic MC</th>
<th>IS</th>
<th>mvtcdf.m</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SE</td>
<td>Mean</td>
</tr>
<tr>
<td>5</td>
<td>2.25E-03</td>
<td>1.73E-04</td>
<td>1.91E-03</td>
</tr>
<tr>
<td>10</td>
<td>2.93E-04</td>
<td>6.25E-05</td>
<td>2.84E-04</td>
</tr>
<tr>
<td>15</td>
<td>5.33E-05</td>
<td>2.67E-05</td>
<td>8.97E-05</td>
</tr>
<tr>
<td>20</td>
<td>2.67E-05</td>
<td>1.89E-05</td>
<td>2.94E-05</td>
</tr>
<tr>
<td>25</td>
<td>4.00E-05</td>
<td>2.31E-05</td>
<td>1.71E-05</td>
</tr>
<tr>
<td>30</td>
<td>-</td>
<td>-</td>
<td>1.07E-05</td>
</tr>
<tr>
<td>50</td>
<td>-</td>
<td>-</td>
<td>1.76E-06</td>
</tr>
<tr>
<td>75</td>
<td>-</td>
<td>-</td>
<td>6.64E-07</td>
</tr>
<tr>
<td>100</td>
<td>-</td>
<td>-</td>
<td>2.54E-07</td>
</tr>
</tbody>
</table>

$c = -2$, $\rho = 0.5$, degree of freedom is 3, and the total number of simulation is 75000.

Averaged CPU time: 1.60E-01, 3.87E-01, 2.34E-01, not including dimensions beyond 25.
Remark on GPU Computing

Graphics Processing Unit (GPU) is a parallel computing device.

Two costs of using GPU:
1. Data transfer is time consuming.
2. CUDA is a low-level programming language.
GPU Computing for previous examples

$5 \times$ speed up by Matlab parallel toolbox for the previous example.

$40 \times$ speed up by CUDA C on GPU.
Summary on Phase 2

Asymptotically optimal IS is obtained for multivariate normal or student t distributions.

GPU computing enabled.

This efficient estimator can be characterized by a relative entropy minimization problem with a linear constraint.
Procedure of Importance Sampling by High-Dimensional Embedding

\[ \mathbb{E}\{\mathbb{I}(\bar{X} \leq \bar{C})\} = \mathbb{E}\left\{ \mathbb{I}(X \leq C) \frac{dP}{dP(C)} \right\} \]

P1: Embedding

P2: I.S.

P3: Projection

\[ \mathbb{E}\{\mathbb{I}(f(\bar{X}) \leq c)\} = \mathbb{E}\{\mathbb{I}(f(X) \leq c) \frac{dP}{dP}\} \]

I.S.
Phase 3: Projection by Parametrization

Further consider a relative entropy minimization problem with **nonlinear** constraint as such

\[
\inf_{\mu(C) \in \mathbb{R}^n} \left[ \mathbb{H}(\tilde{P}_{\mu(C)} | P) \right],
\]

s.t. \( f(C) = c. \)

E.X., in the case of multivariate normal, the objective function is just \( C^T C. \)
Mathematical Formulation for Embedding IS

\[
\inf_{\tilde{P} \in \mathcal{E}} \mathbb{H}(\tilde{P} | P) \text{s.t.} E\tilde{P} [X] \in D, \quad (1)
\]

where \( \mathcal{E} = \{ Q : Q << P \} \) and
\( D = \{ C \in \mathbb{R}^n : f(C) = c \} \).

This method can be modified to be less dependent on \( f \).

Variational analysis by Altun and Smola (06) and Koyejo and Ghosh (13).
FORM IS

- used for structural reliability in engineering
- Value-at-Risk estimation, US pattered by De and Tamarchenko.
FORM IS as an Embedding IS

**Thm** 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.
Portfolio Default Probability Estimation

\[ p = \mathbb{E}\{\mathbbm{1}(X < c)\}, \]

where \( X = \sum_{i=1}^{n} \alpha_{i}X_{i}. \)
Example 1: Normal Case

Suppose \( X \sim \mathcal{N}(0, \Sigma) \) under \( P \) and \( X \sim \mathcal{N}(\mu, \Sigma) \) under \( \tilde{P} \), then

\[
H(\tilde{P}|P) = \frac{1}{2} \mu^T \Sigma^{-1} \mu.
\]

Thus, we consider

\[
\min \mu^T \Sigma^{-1} \mu \quad \text{subject to } \alpha^T \mu = C.
\]

The optimal solution is

\[
\mu^* = \frac{C}{\alpha^T \Sigma \alpha} \Sigma \alpha.
\]
Example 2: Mixed Portfolio - Normal and Student $t$

Consider \( \min H(\tilde{P}|P) \)

subject to
\[
\tilde{P} \left[ a^T A_W W + b^T A_N N \frac{1}{\sqrt{Y/\nu}} \right] = C,
\]

where the equivalent probability measure \( \tilde{P} \) with Radon-Nikodym derivative

\[
\frac{d\tilde{P}}{dP} = \exp \left( \mu^T_W W + \mu^T_N N - \frac{1}{2} \mu^T_W \mu_W - \frac{1}{2} \mu^T_N \mu_N \right).
\]
Example 2: Case of Mixed Portfolio - Cont.

That is, \( \min \frac{1}{2} \left( \mu_W^T \mu_W + \mu_N^T \mu_N \right) \) subject to

\[
a^T A_W \mu_W + \sqrt{\frac{\nu}{2}} \frac{\Gamma((\nu - 1)/2)}{\Gamma(\nu/2)} b^T A_N \mu_N = C.
\]

The solution is obtained as

\[
\begin{pmatrix}
\mu_W^*
\mu_N^*
\end{pmatrix}
= \frac{C K}{K^T K},
\]

where

\[
K = \begin{pmatrix}
a^T A_W \\
\sqrt{\frac{\nu}{2}} \frac{\Gamma((\nu - 1)/2)}{\Gamma(\nu/2)} b^T A_N
\end{pmatrix}.
\]
Example 3: Case of Basket GBMs

\[
\min H(\tilde{\mathbb{P}}|P) = \mathbb{E} \left[ \ln \frac{d\tilde{\mathbb{P}}}{d\mathbb{P}} \right] = \frac{1}{2} \left( h_1^2 + h_2^2 + \cdots + h_n^2 \right) T
\]

subject to

\[
c = \alpha_1 \mathbb{E}[S_{1T}] + \alpha_2 \mathbb{E}[S_{2T}] + \cdots + \alpha_n \mathbb{E}[S_{nT}]
\]

\[
= \alpha_1 S_{10} \exp \left( (\mu_1 - \sigma_1 h_1)T \right)
\]

\[
+ \alpha_2 S_{20} \exp \left( (\mu_2 - \sigma_2 \rho_{12} h_1 - \sigma_2 \sqrt{1 - \rho_{12}^2} h_2)T \right)
\]

\[
+ \cdots + \alpha_n S_{n0} \times \exp \left( (\mu_n - \sigma_n (\rho_{1n} h_1 + \cdots + \rho_{n-1n} h_{n-1} + \sqrt{1 - \rho_{1n}^2 - \cdots - \rho_{n-1n}^2} h_n))T \right)
\]
Conclusion

- Embedding IS can be suitable for solving high dimensional integral problems.

- FORM (Design point) IS can be understood as a special case of Embedding IS.

- Suitable for GPU parallel computing