Risk simulation with optimally stratified importance sampling

W. Hörmann and İ. Başoğlu
Bogazici University Istanbul
Industrial Engineering Department

June 12, 2014

Some Risk Measures

- Tail loss probability for given threshold level, τ:
  $\Pr(Loss \geq \tau) = E\left[\mathbb{1}\left\{Loss \geq \tau\right\}\right]$
- $VaR(\alpha)$: Value-at-risk, $1-\alpha$ quantile of the loss distribution.
- Conditional excess: $E\left[Loss | Loss \geq \tau\right]$
- Conditional value-at-risk: $CVaR(\alpha)$: $E\left[Loss | Loss \geq VaR(\alpha)\right]$

Some Risk Measures

- Tail loss probability for given threshold level, τ:
  $\Pr(Loss \geq \tau) = E\left[\mathbb{1}\left\{Loss \geq \tau\right\}\right]$
- $VaR(\alpha)$: Value-at-risk, $1 - \alpha$ quantile of the loss distribution.
- Conditional excess: $E\left[Loss | Loss \geq \tau\right]$
- Conditional value-at-risk: $CVaR(\alpha)$: $E\left[Loss | Loss \geq VaR(\alpha)\right]$

Some Risk Measures

- Tail loss probability for given threshold level, τ:
  $\Pr(Loss \geq \tau) = E\left[\mathbb{1}\left\{Loss \geq \tau\right\}\right]$
- $VaR(\alpha)$: Value-at-risk, $1 - \alpha$ quantile of the loss distribution.
- Conditional excess: $E\left[Loss | Loss \geq \tau\right]$
- Conditional value-at-risk: $CVaR(\alpha)$: $E\left[Loss | Loss \geq VaR(\alpha)\right]$

Some Risk Measures

- Tail loss probability for given threshold level, τ:
  $\Pr(\mathit{Loss} \geq \tau) = E\left[\mathbb{1}\left\{\mathit{Loss} \geq \tau\right\}\right]$
- $\mathit{VaR}(\alpha)$: Value-at-risk, $1-\alpha$ quantile of the loss distribution.
- Conditional excess: $E\left[\mathit{Loss} \mid \mathit{Loss} \geq \tau\right]$
- Conditional value-at-risk: $\mathit{CVaR}(\alpha)$: $E\left[\mathit{Loss} \mid \mathit{Loss} \geq \mathit{VaR}(\alpha)\right]$

t-copula for modeling log-returns

- For modeling dependence among log-returns the importance of copulae is stressed (Frey and McNeil, 2001).
- t-copula models have a good fit to the joint distribution of log returns. (for example Mashal et al. 2003; Kole et al. 2007).

$$Return\,(\mathbf{T}) = \sum_{d=1}^{D} w_d e^{c_d G_d^{-1}(F_\nu(T_d))}$$

$$c_d = \sqrt{\frac{\sigma_d^2}{252}\frac{1}{\mathrm{var}_d}}$$

$$\boldsymbol{T} = (T_1, \ldots, T_D)' = \frac{\boldsymbol{LZ}}{\sqrt{Y/\nu}}$$

Our Objectives

For a linear asset portfolio of moderate size (2 to 10 assets):

- Efficient estimation of a single tail loss probability, $\Pr(Loss \geq \tau)$
- Efficient estimation of a single conditional excess $E\left[Loss | Loss \geq \tau\right]$.
- Efficient estimation of multiple tail loss probabilities or multiple conditional excesses in a single simulation. (Important to calculate VaR and CVaR.)

Our Objectives

For a linear asset portfolio of moderate size (2 to 10 assets):

- Efficient estimation of a single tail loss probability, $\Pr(Loss \geq \tau)$
- Efficient estimation of a single conditional excess $E[Loss|Loss \geq \tau]$.
- Efficient estimation of multiple tail loss probabilities or multiple conditional excesses in a single simulation. (Important to calculate VaR and CVaR.)

Our Objectives

For a linear asset portfolio of moderate size (2 to 10 assets):

- Efficient estimation of a single tail loss probability, $\Pr(Loss \geq \tau)$
- Efficient estimation of a single conditional excess $E[Loss | Loss \geq \tau]$.
- Efficient estimation of multiple tail loss probabilities or multiple conditional excesses in a single simulation.
  (Important to calculate VaR and CVaR.)

General Principles of Monte Carlo Simulation:

- Estimation of an expectation: $x = E_f[q(\boldsymbol{X})]$
  - $\boldsymbol{X} \in \mathbb{R}^D$ and $\boldsymbol{X}$ has density $f(.)$,
  - $q : \mathbb{R}^D \to \mathbb{R}$ is the "simulation function",
  - $E_f[q^2(\boldsymbol{X})] < \infty$.
- $x = E_f[q(\boldsymbol{X})] = \int_{\boldsymbol{x} \in \mathbb{R}^D} q(\boldsymbol{x}) f(\boldsymbol{x})\, d\boldsymbol{x}$
- The naive estimator: $\hat{x}_{NV} = N^{-1} \sum_{n=1}^{N} q(\boldsymbol{X}^n)$
  - Generate iid sample $\boldsymbol{X}^1, \ldots, \boldsymbol{X}^N$ from density $f(.)$.
  - Central Limit Theorem: $\frac{\hat{x}_{NV} - x}{\sigma / N} \to N(0, 1)$.
  - Error bound: $\hat{x}_{NV} \pm \Phi^{-1}(\alpha/2)\, \sigma \Big/ \sqrt{N}$.
- To get more precise results: Variance Reduction Methods
  - Look for new simulation function $q(.)$ with the same expectation and smaller variance

General Principles of Monte Carlo Simulation:

- Estimation of an expectation: $x = E_f\left[q\left(\boldsymbol{X}\right)\right]$
  - $\boldsymbol{X} \in \mathbb{R}^D$ and $\boldsymbol{X}$ has density $f\left(.\right)$,
  - $q : \mathbb{R}^D \to \mathbb{R}$ is the "simulation function",
  - $E_f\left[q^2\left(\boldsymbol{X}\right)\right] < \infty$.
- $x = E_f\left[q\left(\boldsymbol{X}\right)\right] = \int_{\boldsymbol{x} \in \mathbb{R}^D} q\left(\boldsymbol{x}\right) f\left(\boldsymbol{x}\right) d\boldsymbol{x}$
- The naive estimator: $\hat{x}_{NV} = N^{-1} \sum_{n=1}^{N} q\left(\boldsymbol{X}^n\right)$
  - Generate iid sample $\boldsymbol{X}^1, \ldots, \boldsymbol{X}^N$ from density $f\left(.\right)$.
  - Central Limit Theorem: $\frac{\hat{x}_{NV} - x}{\sigma / N} \to N\left(0, 1\right)$.
  - Error bound: $\hat{x}_{NV} \pm \Phi^{-1}\left(\alpha/2\right)\sigma \big/ \sqrt{N}$.
- To get more precise results: Variance Reduction Methods
  - Look for new simulation function $q(.)$ with the same expectation and smaller variance

General Principles of Monte Carlo Simulation:

- Estimation of an expectation: $x = E_f\left[q\left(\boldsymbol{X}\right)\right]$
  - $\boldsymbol{X} \in \mathbb{R}^D$ and $\boldsymbol{X}$ has density $f\left(.\right)$,
  - $q : \mathbb{R}^D \to \mathbb{R}$ is the "simulation function",
  - $E_f\left[q^2\left(\boldsymbol{X}\right)\right] < \infty$.
- $x = E_f\left[q\left(\boldsymbol{X}\right)\right] = \int_{\boldsymbol{x}\in\mathbb{R}^D} q\left(\boldsymbol{x}\right) f\left(\boldsymbol{x}\right) d\boldsymbol{x}$
- The naive estimator: $\hat{x}_{NV} = N^{-1}\sum_{n=1}^{N} q\left(\boldsymbol{X}^n\right)$
  - Generate iid sample $\boldsymbol{X}^1, \ldots, \boldsymbol{X}^N$ from density $f\left(.\right)$.
  - Central Limit Theorem: $\frac{\hat{x}_{NV} - x}{\sigma / N} \to N\left(0, 1\right)$.
  - Error bound: $\hat{x}_{NV} \pm \Phi^{-1}\left(\alpha/2\right)\sigma \Big/ \sqrt{N}$.
- To get more precise results: Variance Reduction Methods
  - Look for new simulation function $q(.)$ with the same expectation and smaller variance

General Principles of Monte Carlo Simulation:

- Estimation of an expectation: $x = E_f\left[q\left(\boldsymbol{X}\right)\right]$
  - $\boldsymbol{X} \in \mathbb{R}^D$ and $\boldsymbol{X}$ has density $f\left(.\right)$,
  - $q : \mathbb{R}^D \to \mathbb{R}$ is the "simulation function",
  - $E_f\left[q^2\left(\boldsymbol{X}\right)\right] < \infty$.
- $x = E_f\left[q\left(\boldsymbol{X}\right)\right] = \int_{\boldsymbol{x}\in\mathbb{R}^D} q\left(\boldsymbol{x}\right) f\left(\boldsymbol{x}\right) d\boldsymbol{x}$
- The naive estimator: $\hat{x}_{NV} = N^{-1} \sum_{n=1}^{N} q\left(\boldsymbol{X}^n\right)$
  - Generate iid sample $\boldsymbol{X}^1, \ldots, \boldsymbol{X}^N$ from density $f\left(.\right)$.
  - Central Limit Theorem: $\frac{\hat{x}_{NV} - x}{\sigma / N} \to N\left(0, 1\right)$.
  - Error bound: $\hat{x}_{NV} \pm \Phi^{-1}\left(\alpha/2\right)\sigma \Big/ \sqrt{N}$.
- To get more precise results: Variance Reduction Methods
  - Look for new simulation function $q(.)$ with the same expectation and smaller variance

General Principles of Monte Carlo Simulation:

- Estimation of an expectation: $x = E_f[q(\boldsymbol{X})]$
  - $\boldsymbol{X} \in \mathbb{R}^D$ and $\boldsymbol{X}$ has density $f(.)$,
  - $q : \mathbb{R}^D \to \mathbb{R}$ is the "simulation function",
  - $E_f\left[q^2(\boldsymbol{X})\right] < \infty$.
- $x = E_f[q(\boldsymbol{X})] = \displaystyle\int_{\boldsymbol{x} \in \mathbb{R}^D} q(\boldsymbol{x}) f(\boldsymbol{x})\, d\boldsymbol{x}$
- The naive estimator: $\hat{x}_{NV} = N^{-1} \sum_{n=1}^{N} q(\boldsymbol{X}^n)$
  - Generate iid sample $\boldsymbol{X}^1, \ldots, \boldsymbol{X}^N$ from density $f(.)$.
  - Central Limit Theorem: $\dfrac{\hat{x}_{NV} - x}{\sigma / N} \to N(0, 1)$.
  - Error bound: $\hat{x}_{NV} \pm \Phi^{-1}(\alpha/2)\, \sigma \big/ \sqrt{N}$.
- To get more precise results: Variance Reduction Methods
  - Look for new simulation function $q(.)$ with the same expectation and smaller variance

General Principles of Monte Carlo Simulation:

- Estimation of an expectation: $x = E_f[q(\boldsymbol{X})]$
  - $\boldsymbol{X} \in \mathbb{R}^D$ and $\boldsymbol{X}$ has density $f(.)$,
  - $q : \mathbb{R}^D \to \mathbb{R}$ is the "simulation function",
  - $E_f\left[q^2(\boldsymbol{X})\right] < \infty$.
- $x = E_f[q(\boldsymbol{X})] = \int_{\boldsymbol{x} \in \mathbb{R}^D} q(\boldsymbol{x}) f(\boldsymbol{x})\, d\boldsymbol{x}$
- The naive estimator: $\hat{x}_{NV} = N^{-1} \sum_{n=1}^{N} q(\boldsymbol{X}^n)$
  - Generate iid sample $\boldsymbol{X}^1, \ldots, \boldsymbol{X}^N$ from density $f(.)$.
  - Central Limit Theorem: $\dfrac{\hat{x}_{NV} - x}{\sigma / N} \to N(0, 1)$.
  - Error bound: $\hat{x}_{NV} \pm \Phi^{-1}(\alpha/2)\, \sigma \Big/ \sqrt{N}$.
- To get more precise results: Variance Reduction Methods
  - Look for new simulation function $q(.)$ with the same expectation and smaller variance

General Principles of Monte Carlo Simulation:

- Estimation of an expectation: $x = E_f[q(\boldsymbol{X})]$
  - $\boldsymbol{X} \in \mathbb{R}^D$ and $\boldsymbol{X}$ has density $f(.)$,
  - $q : \mathbb{R}^D \to \mathbb{R}$ is the "simulation function",
  - $E_f\left[q^2(\boldsymbol{X})\right] < \infty$.
- $x = E_f[q(\boldsymbol{X})] = \int_{\boldsymbol{x} \in \mathbb{R}^D} q(\boldsymbol{x}) f(\boldsymbol{x})\, d\boldsymbol{x}$
- The naive estimator: $\hat{x}_{NV} = N^{-1} \sum_{n=1}^{N} q(\boldsymbol{X}^n)$
  - Generate iid sample $\boldsymbol{X}^1, \ldots, \boldsymbol{X}^N$ from density $f(.)$.
  - Central Limit Theorem: $\frac{\hat{x}_{NV} - x}{\sigma / N} \to N(0, 1)$.
  - Error bound: $\hat{x}_{NV} \pm \Phi^{-1}(\alpha/2)\, \sigma \Big/ \sqrt{N}$.
- To get more precise results: Variance Reduction Methods
  - Look for new simulation function $q(.)$ with the same expectation and smaller variance

General Principles of Monte Carlo Simulation:

- Estimation of an expectation: $x = E_f\left[q\left(\boldsymbol{X}\right)\right]$
  - $\boldsymbol{X} \in \mathbb{R}^D$ and $\boldsymbol{X}$ has density $f(.)$,
  - $q : \mathbb{R}^D \to \mathbb{R}$ is the "simulation function",
  - $E_f\left[q^2\left(\boldsymbol{X}\right)\right] < \infty$.
- $x = E_f\left[q\left(\boldsymbol{X}\right)\right] = \int_{\boldsymbol{x}\in\mathbb{R}^D} q\left(\boldsymbol{x}\right) f\left(\boldsymbol{x}\right) d\boldsymbol{x}$
- The naive estimator: $\hat{x}_{NV} = N^{-1} \sum_{n=1}^{N} q\left(\boldsymbol{X}^n\right)$
  - Generate iid sample $\boldsymbol{X}^1, \ldots, \boldsymbol{X}^N$ from density $f(.)$.
  - Central Limit Theorem: $\dfrac{\hat{x}_{NV} - x}{\sigma / N} \to N(0, 1)$.
  - Error bound: $\hat{x}_{NV} \pm \Phi^{-1}(\alpha/2)\,\sigma \Big/ \sqrt{N}$.
- To get more precise results: Variance Reduction Methods
  - Look for new simulation function $q(.)$ with the same expectation and smaller variance

General Principles of Monte Carlo Simulation:

- Estimation of an expectation: $x = E_f\left[q\left(\boldsymbol{X}\right)\right]$
  - $\boldsymbol{X} \in \mathbb{R}^D$ and $\boldsymbol{X}$ has density $f\left(.\right)$,
  - $q : \mathbb{R}^D \to \mathbb{R}$ is the "simulation function",
  - $E_f\left[q^2\left(\boldsymbol{X}\right)\right] < \infty$.
- $x = E_f\left[q\left(\boldsymbol{X}\right)\right] = \int_{\boldsymbol{x}\in\mathbb{R}^D} q\left(\boldsymbol{x}\right) f\left(\boldsymbol{x}\right) d\boldsymbol{x}$
- The naive estimator: $\hat{x}_{NV} = N^{-1}\sum_{n=1}^{N} q\left(\boldsymbol{X}^n\right)$
  - Generate iid sample $\boldsymbol{X}^1, \ldots, \boldsymbol{X}^N$ from density $f\left(.\right)$.
  - Central Limit Theorem: $\dfrac{\hat{x}_{NV} - x}{\sigma/N} \to N\left(0, 1\right)$.
  - Error bound: $\hat{x}_{NV} \pm \Phi^{-1}\left(\alpha/2\right)\sigma\Big/\sqrt{N}$.
- To get more precise results: Variance Reduction Methods
  - Look for new simulation function $q(.)$ with the same expectation and smaller variance

General Principles of Monte Carlo Simulation:

- Estimation of an expectation: $x = E_f\left[q\left(\boldsymbol{X}\right)\right]$
  - $\boldsymbol{X} \in \mathbb{R}^D$ and $\boldsymbol{X}$ has density $f\left(.\right)$,
  - $q : \mathbb{R}^D \to \mathbb{R}$ is the "simulation function",
  - $E_f\left[q^2\left(\boldsymbol{X}\right)\right] < \infty$.
- $x = E_f\left[q\left(\boldsymbol{X}\right)\right] = \int_{\boldsymbol{x}\in\mathbb{R}^D} q\left(\boldsymbol{x}\right) f\left(\boldsymbol{x}\right) d\boldsymbol{x}$
- The naive estimator: $\hat{x}_{NV} = N^{-1} \sum_{n=1}^{N} q\left(\boldsymbol{X}^n\right)$
  - Generate iid sample $\boldsymbol{X}^1, \ldots, \boldsymbol{X}^N$ from density $f\left(.\right)$.
  - Central Limit Theorem: $\dfrac{\hat{x}_{NV} - x}{\sigma / N} \to N\left(0, 1\right)$.
  - Error bound: $\hat{x}_{NV} \pm \Phi^{-1}\left(\alpha/2\right) \sigma \Big/ \sqrt{N}$.
- To get more precise results: Variance Reduction Methods
  - Look for new simulation function $q(.)$ with the same expectation and smaller variance

General Principles of Monte Carlo Simulation:

- Estimation of an expectation: $x = E_f\left[q\left(\boldsymbol{X}\right)\right]$
  - $\boldsymbol{X} \in \mathbb{R}^D$ and $\boldsymbol{X}$ has density $f\left(.\right)$,
  - $q : \mathbb{R}^D \to \mathbb{R}$ is the "simulation function",
  - $E_f\left[q^2\left(\boldsymbol{X}\right)\right] < \infty$.
- $x = E_f\left[q\left(\boldsymbol{X}\right)\right] = \int_{\boldsymbol{x}\in\mathbb{R}^D} q\left(\boldsymbol{x}\right) f\left(\boldsymbol{x}\right) d\boldsymbol{x}$
- The naive estimator: $\hat{x}_{NV} = N^{-1} \sum_{n=1}^{N} q\left(\boldsymbol{X}^n\right)$
  - Generate iid sample $\boldsymbol{X}^1, \ldots, \boldsymbol{X}^N$ from density $f\left(.\right)$.
  - Central Limit Theorem: $\dfrac{\hat{x}_{NV} - x}{\sigma/N} \to N\left(0, 1\right)$.
  - Error bound: $\hat{x}_{NV} \pm \Phi^{-1}\left(\alpha/2\right)\sigma \Big/ \sqrt{N}$.
- To get more precise results: Variance Reduction Methods
  - Look for new simulation function $q(.)$ with the same expectation and smaller variance

Importance Sampling (IS)

IS is a frequently used method for rare event situations

- $$\begin{aligned} E_f\left[q\left(\boldsymbol{X}\right)\right] &= \int_{\boldsymbol{x}\in\mathbb{R}^D} q\left(\boldsymbol{x}\right) f\left(\boldsymbol{x}\right) d\boldsymbol{x} \\ &= \int_{\boldsymbol{x}\in\mathbb{R}^D} q\left(\boldsymbol{x}\right) \frac{f\left(\boldsymbol{x}\right)}{f_{IS}\left(\boldsymbol{x}\right)} f_{IS}\left(\boldsymbol{x}\right) d\boldsymbol{x} \\ &= \int_{\boldsymbol{x}\in\mathbb{R}^D} q\left(\boldsymbol{x}\right) \rho\left(\boldsymbol{x}\right) f_{IS}\left(\boldsymbol{x}\right) d\boldsymbol{x} \\ &= E_{f_{IS}}\left[q\left(\boldsymbol{X}\right) \rho\left(\boldsymbol{X}\right)\right]. \end{aligned}$$
- Generate iid sample $\boldsymbol{X}^1, \ldots, \boldsymbol{X}^n$ from density $f_{IS}(.)$,
- Evaluate $\hat{x}_{IS} = N^{-1} \sum_{i=1}^{n} q\left(\boldsymbol{X}^i\right) \rho\left(\boldsymbol{X}^i\right)$.
- Regularity conditions are necessary to prove that the estimate is unbiased

Importance Sampling (IS)

IS is a frequently used method for rare event situations

- $$\begin{aligned} E_f[q(\boldsymbol{X})] &= \int_{\boldsymbol{x}\in\mathbb{R}^D} q(\boldsymbol{x}) f(\boldsymbol{x})\, d\boldsymbol{x} \\ &= \int_{\boldsymbol{x}\in\mathbb{R}^D} q(\boldsymbol{x}) \frac{f(\boldsymbol{x})}{f_{IS}(\boldsymbol{x})} f_{IS}(\boldsymbol{x})\, d\boldsymbol{x} \\ &= \int_{\boldsymbol{x}\in\mathbb{R}^D} q(\boldsymbol{x}) \rho(\boldsymbol{x}) f_{IS}(\boldsymbol{x})\, d\boldsymbol{x} \\ &= E_{f_{IS}}[q(\boldsymbol{X}) \rho(\boldsymbol{X})]. \end{aligned}$$
- Generate iid sample $\boldsymbol{X}^1, \ldots, \boldsymbol{X}^n$ from density $f_{IS}(.)$,
- Evaluate $\hat{x}_{IS} = N^{-1} \sum_{i=1}^{n} q\left(\boldsymbol{X}^i\right) \rho\left(\boldsymbol{X}^i\right)$.
- Regularity conditions are necessary to prove that the estimate is unbiased

Importance Sampling (IS)

IS is a frequently used method for rare event situations

- $$\begin{aligned} E_f[q(\boldsymbol{X})] &= \int_{\boldsymbol{x}\in\mathbb{R}^D} q(\boldsymbol{x}) f(\boldsymbol{x})\, d\boldsymbol{x} \\ &= \int_{\boldsymbol{x}\in\mathbb{R}^D} q(\boldsymbol{x}) \frac{f(\boldsymbol{x})}{f_{IS}(\boldsymbol{x})} f_{IS}(\boldsymbol{x})\, d\boldsymbol{x} \\ &= \int_{\boldsymbol{x}\in\mathbb{R}^D} q(\boldsymbol{x}) \rho(\boldsymbol{x}) f_{IS}(\boldsymbol{x})\, d\boldsymbol{x} \\ &= E_{f_{IS}}[q(\boldsymbol{X}) \rho(\boldsymbol{X})]. \end{aligned}$$
- Generate iid sample $\boldsymbol{X}^1, \ldots, \boldsymbol{X}^n$ from density $f_{IS}(.)$,
- Evaluate $\hat{x}_{IS} = N^{-1} \sum_{i=1}^{n} q\left(\boldsymbol{X}^i\right) \rho\left(\boldsymbol{X}^i\right)$.
- Regularity conditions are necessary to prove that the estimate is unbiased

Importance Sampling (IS)

IS is a frequently used method for rare event situations

- $$\begin{aligned} E_f\left[q\left(\boldsymbol{X}\right)\right] &= \int_{\boldsymbol{x}\in\mathbb{R}^D} q\left(\boldsymbol{x}\right) f\left(\boldsymbol{x}\right) d\boldsymbol{x} \\ &= \int_{\boldsymbol{x}\in\mathbb{R}^D} q\left(\boldsymbol{x}\right) \frac{f\left(\boldsymbol{x}\right)}{f_{IS}\left(\boldsymbol{x}\right)} f_{IS}\left(\boldsymbol{x}\right) d\boldsymbol{x} \\ &= \int_{\boldsymbol{x}\in\mathbb{R}^D} q\left(\boldsymbol{x}\right) \rho\left(\boldsymbol{x}\right) f_{IS}\left(\boldsymbol{x}\right) d\boldsymbol{x} \\ &= E_{f_{IS}}\left[q\left(\boldsymbol{X}\right)\rho\left(\boldsymbol{X}\right)\right]. \end{aligned}$$
- Generate iid sample $\boldsymbol{X}^1, \ldots, \boldsymbol{X}^n$ from density $f_{IS}(.)$,
- Evaluate $\hat{x}_{IS} = N^{-1}\sum_{i=1}^{n} q\left(\boldsymbol{X}^i\right)\rho\left(\boldsymbol{X}^i\right)$.
- Regularity conditions are necessary to prove that the estimate is unbiased

Importance Sampling (IS)

IS is a frequently used method for rare event situations

- $$\begin{aligned} E_f\left[q\left(\boldsymbol{X}\right)\right] &= \int_{\boldsymbol{x}\in\mathbb{R}^D} q\left(\boldsymbol{x}\right) f\left(\boldsymbol{x}\right) d\boldsymbol{x} \\ &= \int_{\boldsymbol{x}\in\mathbb{R}^D} q\left(\boldsymbol{x}\right) \frac{f\left(\boldsymbol{x}\right)}{f_{IS}\left(\boldsymbol{x}\right)} f_{IS}\left(\boldsymbol{x}\right) d\boldsymbol{x} \\ &= \int_{\boldsymbol{x}\in\mathbb{R}^D} q\left(\boldsymbol{x}\right) \rho\left(\boldsymbol{x}\right) f_{IS}\left(\boldsymbol{x}\right) d\boldsymbol{x} \\ &= E_{f_{IS}}\left[q\left(\boldsymbol{X}\right)\rho\left(\boldsymbol{X}\right)\right]. \end{aligned}$$
- Generate iid sample $\boldsymbol{X}^1, \ldots, \boldsymbol{X}^n$ from density $f_{IS}(.)$,
- Evaluate $\hat{x}_{IS} = N^{-1}\sum_{i=1}^{n} q\left(\boldsymbol{X}^i\right)\rho\left(\boldsymbol{X}^i\right)$.
- Regularity conditions are necessary to prove that the estimate is unbiased

Importance Sampling (IS)

IS is a frequently used method for rare event situations

- $$\begin{aligned} E_f\left[q\left(\boldsymbol{X}\right)\right] &= \int_{\boldsymbol{x}\in\mathbb{R}^D} q\left(\boldsymbol{x}\right) f\left(\boldsymbol{x}\right) d\boldsymbol{x} \\ &= \int_{\boldsymbol{x}\in\mathbb{R}^D} q\left(\boldsymbol{x}\right) \frac{f\left(\boldsymbol{x}\right)}{f_{IS}\left(\boldsymbol{x}\right)} f_{IS}\left(\boldsymbol{x}\right) d\boldsymbol{x} \\ &= \int_{\boldsymbol{x}\in\mathbb{R}^D} q\left(\boldsymbol{x}\right) \rho\left(\boldsymbol{x}\right) f_{IS}\left(\boldsymbol{x}\right) d\boldsymbol{x} \\ &= E_{f_{IS}}\left[q\left(\boldsymbol{X}\right) \rho\left(\boldsymbol{X}\right)\right]. \end{aligned}$$
- Generate iid sample $\boldsymbol{X}^1, \ldots, \boldsymbol{X}^n$ from density $f_{IS}(.)$,
- Evaluate $\hat{x}_{IS} = N^{-1} \sum_{i=1}^{n} q\left(\boldsymbol{X}^i\right) \rho\left(\boldsymbol{X}^i\right)$.
- Regularity conditions are necessary to prove that the estimate is unbiased

Importance Sampling (IS)

IS is a frequently used method for rare event situations

- $$\begin{aligned} E_f\left[q\left(\boldsymbol{X}\right)\right] &= \int_{\boldsymbol{x}\in\mathbb{R}^D} q\left(\boldsymbol{x}\right) f\left(\boldsymbol{x}\right) d\boldsymbol{x} \\ &= \int_{\boldsymbol{x}\in\mathbb{R}^D} q\left(\boldsymbol{x}\right) \frac{f\left(\boldsymbol{x}\right)}{f_{IS}\left(\boldsymbol{x}\right)} f_{IS}\left(\boldsymbol{x}\right) d\boldsymbol{x} \\ &= \int_{\boldsymbol{x}\in\mathbb{R}^D} q\left(\boldsymbol{x}\right) \rho\left(\boldsymbol{x}\right) f_{IS}\left(\boldsymbol{x}\right) d\boldsymbol{x} \\ &= E_{f_{IS}}\left[q\left(\boldsymbol{X}\right)\rho\left(\boldsymbol{X}\right)\right]. \end{aligned}$$
- Generate iid sample $\boldsymbol{X}^1, \ldots, \boldsymbol{X}^n$ from density $f_{IS}(.)$,
- Evaluate $\hat{x}_{IS} = N^{-1}\sum_{i=1}^{n} q\left(\boldsymbol{X}^i\right)\rho\left(\boldsymbol{X}^i\right)$.
- Regularity conditions are necessary to prove that the estimate is unbiased

Importance Sampling cont.

- $V\left[\hat{x}_{IS}\right]$ is minimal for $f_{IS}^{*}(\boldsymbol{x}) = \frac{|q(\boldsymbol{x})|\, f(\boldsymbol{x})}{\int_{\boldsymbol{x}\in\mathbb{R}^D} |q(\boldsymbol{x})|\; f(\boldsymbol{x})\, d\boldsymbol{x}}$
  That density is unknown for relevant applications.
- In practice an IS density is typically taken from a parametric family (often the same as $f(.)$). The parameters are selected such that the IS density imitates $|q(\boldsymbol{x})\, f(\boldsymbol{x})|$.
- The cross entropy method is a general approach to select the parameters of the IS density.
- Often the variance reduction reached with IS decreases fast with the dimension of the problem.

Importance Sampling cont.

- $V\left[\hat{x}_{IS}\right]$ is minimal for $f_{IS}^{*}\left(\boldsymbol{x}\right) = \dfrac{\left|q\left(\boldsymbol{x}\right)\right| f\left(\boldsymbol{x}\right)}{\int_{\boldsymbol{x}\in\mathbb{R}^D} \left|q\left(\boldsymbol{x}\right)\right|\ f\left(\boldsymbol{x}\right) d\boldsymbol{x}}$
  That density is unknown for relevant applications.
- In practice an IS density is typically taken from a parametric family (often the same as $f(.)$). The parameters are selected such that the IS density imitates $\left|q\left(\boldsymbol{x}\right) f\left(\boldsymbol{x}\right)\right|$.
- The cross entropy method is a general approach to select the parameters of the IS density.
- Often the variance reduction reached with IS decreases fast with the dimension of the problem.

Importance Sampling cont.

- $V\left[\hat{x}_{IS}\right]$ is minimal for $f^*_{IS}\left(\boldsymbol{x}\right) = \frac{\left|q\left(\boldsymbol{x}\right)\right| f\left(\boldsymbol{x}\right)}{\int_{\boldsymbol{x}\in\mathbb{R}^D} \left|q\left(\boldsymbol{x}\right)\right| \ f\left(\boldsymbol{x}\right) d\boldsymbol{x}}$
  That density is unknown for relevant applications.
- In practice an IS density is typically taken from a parametric family (often the same as $f(.)$). The parameters are selected such that the IS density imitates $\left|q\left(\boldsymbol{x}\right) f\left(\boldsymbol{x}\right)\right|$.
- The cross entropy method is a general approach to select the parameters of the IS density.
- Often the variance reduction reached with IS decreases fast with the dimension of the problem.

Importance Sampling cont.

- $V[\hat{x}_{IS}]$ is minimal for $f^*_{IS}(\boldsymbol{x}) = \dfrac{|q(\boldsymbol{x})|\, f(\boldsymbol{x})}{\int_{\boldsymbol{x}\in\mathbb{R}^D} |q(\boldsymbol{x})|\, f(\boldsymbol{x})\, d\boldsymbol{x}}$
  That density is unknown for relevant applications.

- In practice an IS density is typically taken from a parametric family (often the same as $f(.)$). The parameters are selected such that the IS density imitates $|q(\boldsymbol{x})\, f(\boldsymbol{x})|$.
- The cross entropy method is a general approach to select the parameters of the IS density.

- Often the variance reduction reached with IS decreases fast with the dimension of the problem.

IS Algo for Tailloss Probabilities

Sak, WH and Leydold (2010) use IS for the iid normal input Z and the chi-square random variate Y.

- Problem: Even for heuristic approach necessary to find a good direction for Z. It depends on the threshold τ. Thus a numeric optimization is required.
- Disadvantage of IS: Selection of the IS density and its parameters difficult.

IS Algo for Tailloss Probabilities

Sak, WH and Leydold (2010) use IS for the iid normal input Z and the chi-square random variate Y.

- Problem: Even for heuristic approach necessary to find a good direction for Z. It depends on the threshold τ. Thus a numeric optimization is required.
- Disadvantage of IS: Selection of the IS density and its parameters difficult.

Stratified Sampling

- Let ξ_i, $i = 1, \ldots, I$ be a partition of $\mathbb{R}^D$ into I strata,
- Assume $p_i = \Pr\{\boldsymbol{X} \in \xi_i\}$ are known for $i = 1, \ldots, I$.
- Let $\boldsymbol{X}_i$ be the random vector that follows the conditional distribution of $\boldsymbol{X}$ given $\boldsymbol{X} \in \xi_i$.

$$x = E_f\left[q\left(\boldsymbol{X}\right)\right] = \sum_{i=1}^{I} p_i E_f\left[q\left(\boldsymbol{X}\right)|X \in \xi_i\right] = \sum_{i=1}^{I} p_i E_f\left[q\left(\boldsymbol{X}_i\right)\right]$$

- The stratified estimator: $\hat{x}_{STRS} = \sum_{i=1}^{I} p_i N_i^{-1} \sum_{n=1}^{N_i} q\left(\boldsymbol{X}_i^n\right)$
  - N_i replications in stratum i, $N = \sum_{i=1}^{I} N_i$
  - Generate iid sample $\boldsymbol{X}_i^1, \ldots, \boldsymbol{X}_i^{N_i}$ in each stratum.
  - How should we select N_i, the samplesize for each stratum?
    standard stratification uses proportional allocation
    possible generalisation to QMC

Stratified Sampling

- Let ξ_i, $i = 1, \ldots, I$ be a partition of $\mathbb{R}^D$ into I strata,
- Assume $p_i = \Pr\{\boldsymbol{X} \in \xi_i\}$ are known for $i = 1, \ldots, I$.
- Let $\boldsymbol{X}_i$ be the random vector that follows the conditional distribution of $\boldsymbol{X}$ given $\boldsymbol{X} \in \xi_i$.

$$x = E_f[q(\boldsymbol{X})] = \sum_{i=1}^{I} p_i E_f[q(\boldsymbol{X}) | X \in \xi_i] = \sum_{i=1}^{I} p_i E_f[q(\boldsymbol{X}_i)]$$

- The stratified estimator: $\hat{x}_{STRS} = \sum_{i=1}^{I} p_i N_i^{-1} \sum_{n=1}^{N_i} q(\boldsymbol{X}_i^n)$
  - N_i replications in stratum i, $N = \sum_{i=1}^{I} N_i$
  - Generate iid sample $\boldsymbol{X}_i^1, \ldots, \boldsymbol{X}_i^{N_i}$ in each stratum.
  - How should we select N_i, the samplesize for each stratum? standard stratification uses proportional allocation possible generalisation to QMC

Stratified Sampling

- Let ξ_i, $i = 1, \ldots, I$ be a partition of $\mathbb{R}^D$ into I strata,
- Assume $p_i = \Pr\{\boldsymbol{X} \in \xi_i\}$ are known for $i = 1, \ldots, I$.
- Let $\boldsymbol{X}_i$ be the random vector that follows the conditional distribution of $\boldsymbol{X}$ given $\boldsymbol{X} \in \xi_i$.

$$x = E_f\left[q\left(\boldsymbol{X}\right)\right] = \sum_{i=1}^{I} p_i E_f\left[q\left(\boldsymbol{X}\right) | X \in \xi_i\right] = \sum_{i=1}^{I} p_i E_f\left[q\left(\boldsymbol{X}_i\right)\right]$$

- The stratified estimator: $\hat{x}_{STRS} = \sum_{i=1}^{I} p_i N_i^{-1} \sum_{n=1}^{N_i} q\left(\boldsymbol{X}_i^n\right)$
  - N_i replications in stratum i, $N = \sum_{i=1}^{I} N_i$
  - Generate iid sample $\boldsymbol{X}_i^1, \ldots, \boldsymbol{X}_i^{N_i}$ in each stratum.
  - How should we select N_i, the samplesize for each stratum? standard stratification uses proportional allocation possible generalisation to QMC

Stratified Sampling

- Let ξ_i, $i = 1, \ldots, I$ be a partition of $\mathbb{R}^D$ into I strata,
- Assume $p_i = \Pr\{\boldsymbol{X} \in \xi_i\}$ are known for $i = 1, \ldots, I$.
- Let $\boldsymbol{X}_i$ be the random vector that follows the conditional distribution of $\boldsymbol{X}$ given $\boldsymbol{X} \in \xi_i$.

$$x = E_f\left[q\left(\boldsymbol{X}\right)\right] = \sum_{i=1}^{I} p_i E_f\left[q\left(\boldsymbol{X}\right)|X \in \xi_i\right] = \sum_{i=1}^{I} p_i E_f\left[q\left(\boldsymbol{X}_i\right)\right]$$

- The stratified estimator: $\hat{x}_{STRS} = \sum_{i=1}^{I} p_i N_i^{-1} \sum_{n=1}^{N_i} q\left(\boldsymbol{X}_i^n\right)$
  - N_i replications in stratum i, $N = \sum_{i=1}^{I} N_i$
  - Generate iid sample $\boldsymbol{X}_i^1, \ldots, \boldsymbol{X}_i^{N_i}$ in each stratum.
  - How should we select N_i, the samplesize for each stratum? standard stratification uses proportional allocation possible generalisation to QMC

Stratified Sampling

- Let ξ_i, $i = 1, \ldots, I$ be a partition of $\mathbb{R}^D$ into I strata,
- Assume $p_i = \Pr\{\boldsymbol{X} \in \xi_i\}$ are known for $i = 1, \ldots, I$.
- Let $\boldsymbol{X}_i$ be the random vector that follows the conditional distribution of $\boldsymbol{X}$ given $\boldsymbol{X} \in \xi_i$.

$$x = E_f[q(\boldsymbol{X})] = \sum_{i=1}^{I} p_i E_f[q(\boldsymbol{X}) | X \in \xi_i] = \sum_{i=1}^{I} p_i E_f[q(\boldsymbol{X}_i)]$$

- The stratified estimator: $\hat{x}_{STRS} = \sum_{i=1}^{I} p_i N_i^{-1} \sum_{n=1}^{N_i} q(\boldsymbol{X}_i^n)$
  - N_i replications in stratum i, $N = \sum_{i=1}^{I} N_i$
  - Generate iid sample $\boldsymbol{X}_i^1, \ldots, \boldsymbol{X}_i^{N_i}$ in each stratum.
  - How should we select N_i, the samplesize for each stratum?
    standard stratification uses proportional allocation
    possible generalisation to QMC

Stratified Sampling

- Let ξ_i, $i = 1, \ldots, I$ be a partition of $\mathbb{R}^D$ into I strata,
- Assume $p_i = \Pr\{\boldsymbol{X} \in \xi_i\}$ are known for $i = 1, \ldots, I$.
- Let $\boldsymbol{X}_i$ be the random vector that follows the conditional distribution of $\boldsymbol{X}$ given $\boldsymbol{X} \in \xi_i$.

$$x = E_f[q(\boldsymbol{X})] = \sum_{i=1}^{I} p_i E_f[q(\boldsymbol{X}) | X \in \xi_i] = \sum_{i=1}^{I} p_i E_f[q(\boldsymbol{X}_i)]$$

- The stratified estimator: $\hat{x}_{STRS} = \sum_{i=1}^{I} p_i N_i^{-1} \sum_{n=1}^{N_i} q(\boldsymbol{X}_i^n)$
  - N_i replications in stratum i, $N = \sum_{i=1}^{I} N_i$
  - Generate iid sample $\boldsymbol{X}_i^1, \ldots, \boldsymbol{X}_i^{N_i}$ in each stratum.
  - How should we select N_i, the samplesize for each stratum? standard stratification uses proportional allocation possible generalisation to QMC

Stratified Sampling

- Let ξ_i, $i = 1, \ldots, I$ be a partition of $\mathbb{R}^D$ into I strata,
- Assume $p_i = \Pr\{\boldsymbol{X} \in \xi_i\}$ are known for $i = 1, \ldots, I$.
- Let $\boldsymbol{X}_i$ be the random vector that follows the conditional distribution of $\boldsymbol{X}$ given $\boldsymbol{X} \in \xi_i$.

$$x = E_f\left[q\left(\boldsymbol{X}\right)\right] = \sum_{i=1}^{I} p_i E_f\left[q\left(\boldsymbol{X}\right)|X \in \xi_i\right] = \sum_{i=1}^{I} p_i E_f\left[q\left(\boldsymbol{X}_i\right)\right]$$

- The stratified estimator: $\hat{x}_{STRS} = \sum_{i=1}^{I} p_i N_i^{-1} \sum_{n=1}^{N_i} q\left(\boldsymbol{X}_i^n\right)$
  - N_i replications in stratum i, $N = \sum_{i=1}^{I} N_i$
  - Generate iid sample $\boldsymbol{X}_i^1, \ldots, \boldsymbol{X}_i^{N_i}$ in each stratum.
  - How should we select N_i, the samplesize for each stratum? standard stratification uses proportional allocation possible generalisation to QMC

Stratified Sampling

- Let ξ_i, $i = 1, \ldots, I$ be a partition of $\mathbb{R}^D$ into I strata,
- Assume $p_i = \Pr\{\boldsymbol{X} \in \xi_i\}$ are known for $i = 1, \ldots, I$.
- Let $\boldsymbol{X}_i$ be the random vector that follows the conditional distribution of $\boldsymbol{X}$ given $\boldsymbol{X} \in \xi_i$.

$$x = E_f[q(\boldsymbol{X})] = \sum_{i=1}^{I} p_i E_f[q(\boldsymbol{X}) | X \in \xi_i] = \sum_{i=1}^{I} p_i E_f[q(\boldsymbol{X}_i)]$$

- The stratified estimator: $\hat{x}_{STRS} = \sum_{i=1}^{I} p_i N_i^{-1} \sum_{n=1}^{N_i} q(\boldsymbol{X}_i^n)$
  - N_i replications in stratum i, $N = \sum_{i=1}^{I} N_i$
  - Generate iid sample $\boldsymbol{X}_i^1, \ldots, \boldsymbol{X}_i^{N_i}$ in each stratum.
  - How should we select N_i, the samplesize for each stratum?
    standard stratification uses proportional allocation
    possible generalisation to QMC

Stratified Sampling with optimal allocation

- Define allocation fractions $\pi_i = N_i / N$

$$\hat{x}_{STRS} = \sum_{i=1}^{I} p_i N_i^{-1} \sum_{n=1}^{N_i} q\left(\boldsymbol{X}_i^n\right) = N^{-1} \sum_{i=1}^{I} p_i \pi_i^{-1} \sum_{n=1}^{\pi_i N} q\left(\boldsymbol{X}_i^n\right)$$

- Let $\sigma_i^2 = V\left[q\left(\boldsymbol{X}\right) | \boldsymbol{X} \in \xi_i\right] = V\left[q\left(\boldsymbol{X}_i\right)\right]$, then:

$$V\left[\hat{x}_{STRS}\right] = N^{-1} \sum_{i=1}^{I} p_i^2 \sigma_i^2 \pi_i^{-1} \geq N^{-1} \left(\sum_{i=1}^{I} p_i \sigma_i\right)^2$$

- Optimal allocation fractions: $\pi_i^* = \frac{p_i \sigma_i}{\sum_{k=1}^{I} p_k \sigma_k}, i = 1, \ldots, I.$
  - Use a pilot sample allocated proportional to p_i
  - Estimate conditional standard deviations $\hat{\sigma}_i$
  - Calculate optimal allocation fractions $\pi^* = (\pi_1^*, \ldots, \pi_I^*)$
  - Use optimal allocation in the main run

Stratified Sampling with optimal allocation

- Define allocation fractions $\pi_i = N_i / N$

$$\hat{x}_{STRS} = \sum_{i=1}^{I} p_i N_i^{-1} \sum_{n=1}^{N_i} q\left(\boldsymbol{X}_i^n\right) = N^{-1} \sum_{i=1}^{I} p_i \pi_i^{-1} \sum_{n=1}^{\pi_i N} q\left(\boldsymbol{X}_i^n\right)$$

- Let $\sigma_i^2 = V\left[q\left(\boldsymbol{X}\right) | \boldsymbol{X} \in \xi_i\right] = V\left[q\left(\boldsymbol{X}_i\right)\right]$, then:

$$V\left[\hat{x}_{STRS}\right] = N^{-1} \sum_{i=1}^{I} p_i^2 \sigma_i^2 \pi_i^{-1} \geq N^{-1} \left(\sum_{i=1}^{I} p_i \sigma_i\right)^2$$

- Optimal allocation fractions: $\pi_i^* = \frac{p_i \sigma_i}{\sum_{k=1}^{I} p_k \sigma_k}, i = 1, \ldots, I.$
  - Use a pilot sample allocated proportional to p_i;
  - Estimate conditional standard deviations $\hat{\sigma}_i$;
  - Calculate optimal allocation fractions $\pi^* = \left(\pi_1, \ldots, \pi_I\right)'$;
  - Use optimal allocation in the main run.

Stratified Sampling with optimal allocation

- Define allocation fractions $\pi_i = N_i/N$

$$\hat{x}_{STRS} = \sum_{i=1}^{I} p_i N_i^{-1} \sum_{n=1}^{N_i} q\left(\boldsymbol{X}_i^n\right) = N^{-1} \sum_{i=1}^{I} p_i \pi_i^{-1} \sum_{n=1}^{\pi_i N} q\left(\boldsymbol{X}_i^n\right)$$

- Let $\sigma_i^2 = V\left[q\left(\boldsymbol{X}\right) | \boldsymbol{X} \in \xi_i\right] = V\left[q\left(\boldsymbol{X}_i\right)\right]$, then:

$$V\left[\hat{x}_{STRS}\right] = N^{-1} \sum\nolimits_{i=1}^{I} p_i^2 \sigma_i^2 \pi_i^{-1} \geq N^{-1} \left(\sum\nolimits_{i=1}^{I} p_i \sigma_i\right)^2$$

- Optimal allocation fractions: $\pi_i^* = \dfrac{p_i \sigma_i}{\sum_{k=1}^{I} p_k \sigma_k}$, $i = 1, \ldots, I$.
  - Use a pilot sample allocated proportional to p_i.
  - Estimate conditional standard deviations $\hat{\sigma}_i$,
  - Calculate optimal allocation fractions $\pi^* = \left(\pi_1, \ldots, \pi_I\right)^T$,
  - Use optimal allocation in the main run.

Stratified Sampling with optimal allocation

- Define allocation fractions $\pi_i = N_i / N$

$$\hat{x}_{STRS} = \sum_{i=1}^{I} p_i N_i^{-1} \sum_{n=1}^{N_i} q\left(\boldsymbol{X}_i^n\right) = N^{-1} \sum_{i=1}^{I} p_i \pi_i^{-1} \sum_{n=1}^{\pi_i N} q\left(\boldsymbol{X}_i^n\right)$$

- Let $\sigma_i^2 = V\left[q\left(\boldsymbol{X}\right) | \boldsymbol{X} \in \xi_i\right] = V\left[q\left(\boldsymbol{X}_i\right)\right]$, then:

$$V\left[\hat{x}_{STRS}\right] = N^{-1} \sum_{i=1}^{I} p_i^2 \sigma_i^2 \pi_i^{-1} \geq N^{-1} \left(\sum_{i=1}^{I} p_i \sigma_i\right)^2$$

- Optimal allocation fractions: $\pi_i^* = \frac{p_i \sigma_i}{\sum_{k=1}^{I} p_k \sigma_k}$, $i = 1, \ldots, I$.
  - Use a pilot sample allocated proportional to p_i,
  - Estimate conditional standard deviations $\hat{\sigma}_i$,
  - Calculate optimal allocation fractions $\pi^* = (\pi_1, \ldots, \pi_I)'$,
  - Use optimal allocation in the main run.

Stratified Sampling with optimal allocation

- Define allocation fractions $\pi_i = N_i/N$

$$\hat{x}_{STRS} = \sum_{i=1}^{I} p_i N_i^{-1} \sum_{n=1}^{N_i} q\left(\boldsymbol{X}_i^n\right) = N^{-1} \sum_{i=1}^{I} p_i \pi_i^{-1} \sum_{n=1}^{\pi_i N} q\left(\boldsymbol{X}_i^n\right)$$

- Let $\sigma_i^2 = V\left[q\left(\boldsymbol{X}\right) | \boldsymbol{X} \in \xi_i\right] = V\left[q\left(\boldsymbol{X}_i\right)\right]$, then:

$$V\left[\hat{x}_{STRS}\right] = N^{-1} \sum\nolimits_{i=1}^{I} p_i^2 \sigma_i^2 \pi_i^{-1} \geq N^{-1} \left(\sum\nolimits_{i=1}^{I} p_i \sigma_i\right)^2$$

- Optimal allocation fractions: $\pi_i^* = \dfrac{p_i \sigma_i}{\sum_{k=1}^{I} p_k \sigma_k}, \; i = 1, \ldots, I.$
  - Use a pilot sample allocated proportional to p_i,
  - Estimate conditional standard deviations $\hat{\sigma}_i$,
  - Calculate optimal allocation fractions $\pi^* = (\pi_1, \ldots, \pi_I)'$,
  - Use optimal allocation in the main run.

Stratified Sampling with optimal allocation

- Define allocation fractions $\pi_i = N_i/N$

$$\hat{x}_{STRS} = \sum_{i=1}^{I} p_i N_i^{-1} \sum_{n=1}^{N_i} q\left(\boldsymbol{X}_i^n\right) = N^{-1} \sum_{i=1}^{I} p_i \pi_i^{-1} \sum_{n=1}^{\pi_i N} q\left(\boldsymbol{X}_i^n\right)$$

- Let $\sigma_i^2 = V\left[q\left(\boldsymbol{X}\right)|\boldsymbol{X} \in \xi_i\right] = V\left[q\left(\boldsymbol{X}_i\right)\right]$, then:

$$V\left[\hat{x}_{STRS}\right] = N^{-1} \sum\nolimits_{i=1}^{I} p_i^2 \sigma_i^2 \pi_i^{-1} \geq N^{-1} \left(\sum\nolimits_{i=1}^{I} p_i \sigma_i\right)^2$$

- Optimal allocation fractions: $\pi_i^* = \dfrac{p_i \sigma_i}{\sum_{k=1}^{I} p_k \sigma_k}$, $i = 1, \ldots, I$.
    - Use a pilot sample allocated proportional to p_i,
    - Estimate conditional standard deviations $\hat{\sigma}_i$,
    - Calculate optimal allocation fractions $\boldsymbol{\pi}^* = (\pi_1, \ldots, \pi_I)'$,
    - Use optimal allocation in the main run.

Stratified Sampling with optimal allocation

- Define allocation fractions $\pi_i = N_i / N$

$$\hat{x}_{STRS} = \sum_{i=1}^{I} p_i N_i^{-1} \sum_{n=1}^{N_i} q\left(\boldsymbol{X}_i^n\right) = N^{-1} \sum_{i=1}^{I} p_i \pi_i^{-1} \sum_{n=1}^{\pi_i N} q\left(\boldsymbol{X}_i^n\right)$$

- Let $\sigma_i^2 = V\left[q\left(\boldsymbol{X}\right) | \boldsymbol{X} \in \xi_i\right] = V\left[q\left(\boldsymbol{X}_i\right)\right]$, then:

$$V\left[\hat{x}_{STRS}\right] = N^{-1} \sum_{i=1}^{I} p_i^2 \sigma_i^2 \pi_i^{-1} \geq N^{-1} \left(\sum_{i=1}^{I} p_i \sigma_i\right)^2$$

- Optimal allocation fractions: $\pi_i^* = \frac{p_i \sigma_i}{\sum_{k=1}^{I} p_k \sigma_k}$, $i = 1, \ldots, I$.
  - Use a pilot sample allocated proportional to p_i,
  - Estimate conditional standard deviations $\hat{\sigma}_i$,
  - Calculate optimal allocation fractions $\boldsymbol{\pi}^* = (\pi_1, \ldots, \pi_I)'$,
  - Use optimal allocation in the main run.

Stratified Sampling with optimal allocation

- Define allocation fractions $\pi_i = N_i / N$

$$\hat{x}_{STRS} = \sum_{i=1}^{I} p_i N_i^{-1} \sum_{n=1}^{N_i} q\left(\boldsymbol{X}_i^n\right) = N^{-1} \sum_{i=1}^{I} p_i \pi_i^{-1} \sum_{n=1}^{\pi_i N} q\left(\boldsymbol{X}_i^n\right)$$

- Let $\sigma_i^2 = V\left[q\left(\boldsymbol{X}\right) | \boldsymbol{X} \in \xi_i\right] = V\left[q\left(\boldsymbol{X}_i\right)\right]$, then:

$$V\left[\hat{x}_{STRS}\right] = N^{-1} \sum_{i=1}^{I} p_i^2 \sigma_i^2 \pi_i^{-1} \geq N^{-1} \left(\sum_{i=1}^{I} p_i \sigma_i\right)^2$$

- Optimal allocation fractions: $\pi_i^* = \frac{p_i \sigma_i}{\sum_{k=1}^{I} p_k \sigma_k}$, $i = 1, \ldots, I$.
  - Use a pilot sample allocated proportional to p_i,
  - Estimate conditional standard deviations $\hat{\sigma}_i$,
  - Calculate optimal allocation fractions $\boldsymbol{\pi}^* = \left(\pi_1, \ldots, \pi_I\right)'$,
  - Use optimal allocation in the main run.

Adaptive Optimal Allocation Algorithm

- The Adaptive Optimal Allocation (AOA) algorithm terminates in K iterations.
- The total sample size N is divided between iterations with a non-decreasing order (e.g., $K = 3$, $0.1N, 0.4N, 0.5N$).
- In the first iteration, the sample is allocated proportional to stratum probabilities p_i.
  - In each iteration, the conditional standard deviations, σ_i, $i = 1, \ldots, I$, are estimated using all previous drawings.
  - The estimates $\hat{\sigma}_i$ determines the allocation fractions for the next iteration.

Étoré and Jourdain (2010) show that the stratified estimator of AOA is unbiased and asymptotically normal.

Adaptive Optimal Allocation Algorithm

- The Adaptive Optimal Allocation (AOA) algorithm terminates in K iterations.
- The total sample size N is divided between iterations with a non-decreasing order (e.g., $K = 3$, $0.1N, 0.4N, 0.5N$).
- In the first iteration, the sample is allocated proportional to stratum probabilities p_i.
  - In each iteration, the conditional standard deviations, σ_i, $i = 1, \ldots, I$, are estimated using all previous drawings.
  - The estimates $\hat{\sigma}_i$ determines the allocation fractions for the next iteration.

Étoré and Jourdain (2010) show that the stratified estimator of AOA is unbiased and asymptotically normal.

Adaptive Optimal Allocation Algorithm

- The Adaptive Optimal Allocation (AOA) algorithm terminates in K iterations.
- The total sample size N is divided between iterations with a non-decreasing order (e.g., $K = 3$, $0.1N, 0.4N, 0.5N$).
- In the first iteration, the sample is allocated proportional to stratum probabilities p_i.
  - In each iteration, the conditional standard deviations, σ_i, $i = 1, \ldots, I$, are estimated using all previous drawings.
  - The estimates $\hat{\sigma}_i$ determines the allocation fractions for the next iteration.

Étoré and Jourdain (2010) show that the stratified estimator of AOA is unbiased and asymptotically normal.

Adaptive Optimal Allocation Algorithm

- The Adaptive Optimal Allocation (AOA) algorithm terminates in K iterations.
- The total sample size N is divided between iterations with a non-decreasing order (e.g., $K = 3$, $0.1N, 0.4N, 0.5N$).
- In the first iteration, the sample is allocated proportional to stratum probabilities p_i.
  - In each iteration, the conditional standard deviations, σ_i, $i = 1, \ldots, I$, are estimated using all previous drawings.
  - The estimates $\hat{\sigma}_i$ determines the allocation fractions for the next iteration.

Étoré and Jourdain (2010) show that the stratified estimator of AOA is unbiased and asymptotically normal.

Adaptive Optimal Allocation Algorithm

- The Adaptive Optimal Allocation (AOA) algorithm terminates in K iterations.
- The total sample size N is divided between iterations with a non-decreasing order (e.g., $K = 3$, $0.1N, 0.4N, 0.5N$).
- In the first iteration, the sample is allocated proportional to stratum probabilities p_i.
  - In each iteration, the conditional standard deviations, σ_i, $i = 1, \ldots, I$, are estimated using all previous drawings.
  - The estimates $\hat{\sigma}_i$ determines the allocation fractions for the next iteration.

Étoré and Jourdain (2010) show that the stratified estimator of AOA is unbiased and asymptotically normal.

Adaptive Optimal Allocation Algorithm

- The Adaptive Optimal Allocation (AOA) algorithm terminates in K iterations.
- The total sample size N is divided between iterations with a non-decreasing order (e.g., $K = 3$, $0.1N, 0.4N, 0.5N$).
- In the first iteration, the sample is allocated proportional to stratum probabilities p_i.
  - In each iteration, the conditional standard deviations, σ_i, $i = 1, \ldots, I$, are estimated using all previous drawings.
  - The estimates $\hat{\sigma}_i$ determines the allocation fractions for the next iteration.

Étoré and Jourdain (2010) show that the stratified estimator of AOA is unbiased and asymptotically normal.

IS vs Optimal Allocation Stratification (OAS)

- OAS can be interpreted as IS using the product of the original density and a step function as weight function.
- Advantage of OAS is the simple formula for the optimal allocation fractions.
- High dimensional stratification not possible in practice. Thus (like for IS) one (or two) main directions are used for most applications.
- Disadvantage of OAS: Many strata (or an adaptive strata structure) are necessary for rare event simulations.
- Is it possible and sensble to combine IS and OAS to increase the variance reduction?

IS vs Optimal Allocation Stratification (OAS)

- OAS can be interpreted as IS using the product of the original density and a step function as weight function.
- Advantage of OAS is the simple formula for the optimal allocation fractions.
- High dimensional stratification not possible in practice. Thus (like for IS) one (or two) main directions are used for most applications.
- Disadvantage of OAS: Many strata (or an adaptive strata structure) are necessary for rare event simulations.
- Is it possible and sensble to combine IS and OAS to increase the variance reduction?

IS vs Optimal Allocation Stratification (OAS)

- OAS can be interpreted as IS using the product of the original density and a step function as weight function.
- Advantage of OAS is the simple formula for the optimal allocation fractions.
- High dimensional stratification not possible in practice. Thus (like for IS) one (or two) main directions are used for most applications.
- Disadvantage of OAS: Many strata (or an adaptive strata structure) are necessary for rare event simulations.
- Is it possible and sensble to combine IS and OAS to increase the variance reduction?

IS vs Optimal Allocation Stratification (OAS)

- OAS can be interpreted as IS using the product of the original density and a step function as weight function.
- Advantage of OAS is the simple formula for the optimal allocation fractions.
- High dimensional stratification not possible in practice. Thus (like for IS) one (or two) main directions are used for most applications.
- Disadvantage of OAS: Many strata (or an adaptive strata structure) are necessary for rare event simulations.
- Is it possible and sensble to combine IS and OAS to increase the variance reduction?

IS vs Optimal Allocation Stratification (OAS)

- OAS can be interpreted as IS using the product of the original density and a step function as weight function.
- Advantage of OAS is the simple formula for the optimal allocation fractions.
- High dimensional stratification not possible in practice. Thus (like for IS) one (or two) main directions are used for most applications.
- Disadvantage of OAS: Many strata (or an adaptive strata structure) are necessary for rare event simulations.
- Is it possible and sensble to combine IS and OAS to increase the variance reduction?

Stratified Importance Sampling: Single Estimate Case

- Stratified Importance Sampling (SIS): Applies OAS to an IS algorithm.
- Numerical results for IS of (Sak et al. 2010) and SIS for *t*-copula model with generalized hyperbolic marginals (parameters estimated from NYSE data)

	$P(R<t)\approx 0.05$						$P(R<t)\approx 0.001$					
	IS			SIS			IS			SIS		
d	VR	TM	ER	VR	TM	ER	VR	TM	ER	VR	TM	ER
2	6.1	0.23	6.5	296.5	0.30	243.5	183.5	0.21	201.7	4741.2	0.28	4046.9
5	8.4	0.79	8.4	110.3	0.90	96.8	278.5	0.79	247.3	3495.7	0.91	2684.9
10	5.3	1.59	5.0	11.4	1.75	9.8	66.6	1.64	60.5	198.3	1.91	154.1

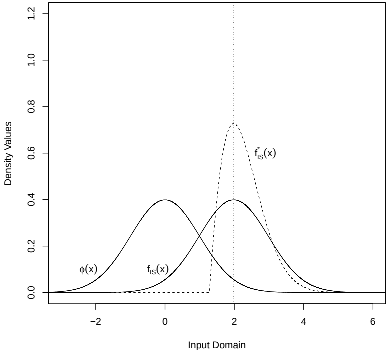
Variance reduction factors: $VR(\hat{x}) = V[\hat{x}_{NV}]/V[\hat{x}]$
Efficiency ratios: $ER(\hat{x}) = VR(\hat{x})\, TM[\hat{x}_{NV}]/TM[\hat{x}]$
TM exeuction time, $N \approx 100,000$ for all simulations.

Why works combination of IS and OAS so well?

- IS and OAS use the same direction and are very similar methods.
- We demonstrate their synergy effects for a one-dimensional example

simple example IS

$$y = E_\phi\left[q\left(Z\right)\right], \quad q\left(x\right) = \left\{e^x - 3.6\right\}^+ \quad \text{and } Z \sim N(0,1).$$



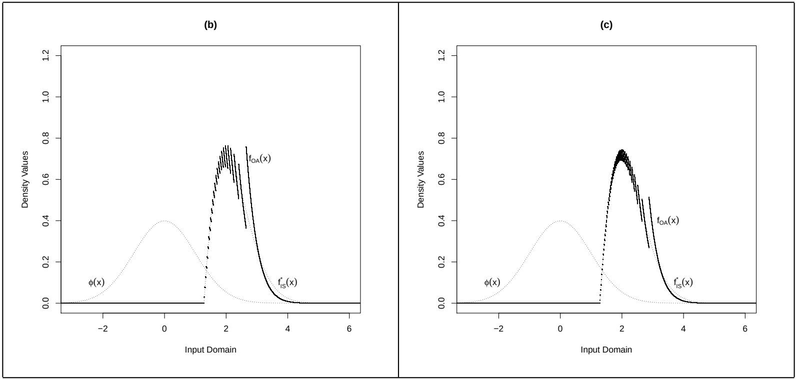
The original density $\phi\left(x\right)$,

the shifted IS density $f_{IS}\left(x\right)$ and

the optimal IS density $f_{IS}^*\left(x\right)$.

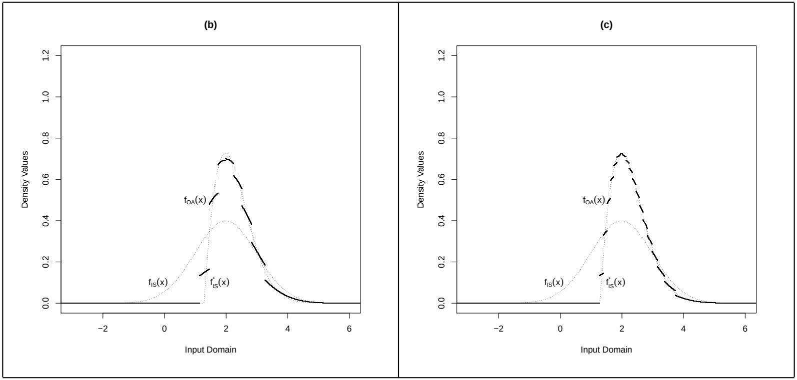
simple example with OAS

IS density corresponding to OAS with 100 and 500 strata.



simple example with SIS

IS density corresponding to SIS with 10 and 25 strata.



simple example: Variance Reduction Factors

Table: Comparison of Naive, IS, OAS 1000, and SIS 100.
exact solution 0.2815896024 .

	Estimate	Variance	VRF
Naive	0.27970	2.17E-05	1
IS	0.28192	3.77E-07	58
OAS 1000	0.28155	2.73E-09	7950
SIS 100	0.28159	8.53E-11	2.5e5

Advantages of combining IS and OAS

We have observed the following advantages when combining IS and OAS for risk simulations:

- IS helps that there are a smaller number of strata with return 0.
- Thus a smaller number of stratification intervals still leads to substantial variance reduction.
- IS helps stratification to obtain better estimates for the variances in the strata and thus better allocation fractions.

NEW IDEA

- Stratification can help to estimate many tailloss probabilities for different thresholds τ_j in a single simulation.
  That is important when estimating VaR and CVaR.

Advantages of combining IS and OAS

We have observed the following advantages when combining IS and OAS for risk simulations:

- IS helps that there are a smaller number of strata with return 0.
- Thus a smaller number of stratification intervals still leads to substantial variance reduction.
- IS helps stratification to obtain better estimates for the variances in the strata and thus better allocation fractions.

NEW IDEA

- Stratification can help to estimate many tailloss probabilities for different thresholds τ_j in a single simulation.
  That is important when estimating VaR and CVaR.

Advantages of combining IS and OAS

We have observed the following advantages when combining IS and OAS for risk simulations:

- IS helps that there are a smaller number of strata with return 0.
- Thus a smaller number of stratification intervals still leads to substantial variance reduction.
- IS helps stratification to obtain better estimates for the variances in the strata and thus better allocation fractions.

NEW IDEA

- Stratification can help to estimate many tailloss probabilities for different thresholds τ_j in a single simulation.
  That is important when estimating VaR and CVaR.

Advantages of combining IS and OAS

We have observed the following advantages when combining IS and OAS for risk simulations:

- IS helps that there are a smaller number of strata with return 0.
- Thus a smaller number of stratification intervals still leads to substantial variance reduction.
- IS helps stratification to obtain better estimates for the variances in the strata and thus better allocation fractions.

NEW IDEA

- Stratification can help to estimate many tailloss probabilities for different thresholds τ_j in a single simulation.
  That is important when estimating VaR and CVaR.

Multiple Estimates Case

- Suppose we are interested in probability estimates for distinct threshold values, τ_j, $j = 1, \ldots, J$.
- We estimate the j-th tail loss probability and its relative error with the following formula:

$$\hat{x}_j = \sum_{i=1}^{I} p_i \hat{x}_{ij}, \qquad RE\left[\hat{x}_j\right] = \frac{\Phi^{-1}(0.975)}{\hat{x}_j \sqrt{N}} \sqrt{\sum_{i=1}^{I} \frac{p_i^2 \hat{\sigma}_{ij}^2}{\pi_i}}$$

- General questions
  - How should we define the "overall error" of our J estimation problems?
  - How can we minimize that "overall error"?

Multiple Estimates Case

- Suppose we are interested in probability estimates for distinct threshold values, τ_j, $j = 1, \ldots, J$.
- We estimate the j-th tail loss probability and its relative error with the following formula:

$$\hat{x}_j = \sum_{i=1}^{I} p_i \hat{x}_{ij}, \qquad RE\left[\hat{x}_j\right] = \frac{\Phi^{-1}(0.975)}{\hat{x}_j\sqrt{N}}\sqrt{\sum_{i=1}^{I}\frac{p_i^2\hat{\sigma}_{ij}^2}{\pi_i}}$$

- General questions
  - How should we define the "overall error" of our J estimation problems?
  - How can we minimize that "overall error"?

Multiple Estimates Case

- Suppose we are interested in probability estimates for distinct threshold values, τ_j, $j = 1, \ldots, J$.
- We estimate the j-th tail loss probability and its relative error with the following formula:

$$\hat{x}_j = \sum_{i=1}^{I} p_i \hat{x}_{ij}, \qquad RE\left[\hat{x}_j\right] = \frac{\Phi^{-1}(0.975)}{\hat{x}_j \sqrt{N}} \sqrt{\sum_{i=1}^{I} \frac{p_i^2 \hat{\sigma}_{ij}^2}{\pi_i}}$$

- General questions
  - How should we define the "overall error" of our J estimation problems?
  - How can we minimize that "overall error"?

Multiple Estimates Case

- Suppose we are interested in probability estimates for distinct threshold values, τ_j, $j = 1, \ldots, J$.
- We estimate the j-th tail loss probability and its relative error with the following formula:

$$\hat{x}_j = \sum_{i=1}^{I} p_i \hat{x}_{ij}, \qquad RE\left[\hat{x}_j\right] = \frac{\Phi^{-1}(0.975)}{\hat{x}_j \sqrt{N}} \sqrt{\sum_{i=1}^{I} \frac{p_i^2 \hat{\sigma}_{ij}^2}{\pi_i}}$$

- General questions
  - How should we define the "overall error" of our J estimation problems?
  - How can we minimize that "overall error"?

Multiple Estimates Case

- Suppose we are interested in probability estimates for distinct threshold values, τ_j, $j = 1, \ldots, J$.
- We estimate the j-th tail loss probability and its relative error with the following formula:

$$\hat{x}_j = \sum_{i=1}^{I} p_i \hat{x}_{ij}, \qquad RE\left[\hat{x}_j\right] = \frac{\Phi^{-1}(0.975)}{\hat{x}_j \sqrt{N}} \sqrt{\sum_{i=1}^{I} \frac{p_i^2 \hat{\sigma}_{ij}^2}{\pi_i}}$$

- General questions
  - How should we define the "overall error" of our J estimation problems?
  - How can we minimize that "overall error"?

Definition of Overall Error

$$\hat{x}_{ij} = \text{Prob}(Loss > \tau_j | X \in \xi_i)$$
$$\hat{x}_j = \text{Prob}(Loss > \tau_j) = \sum_{i=1}^{I} p_i \hat{x}_{ij}$$
$$\hat{s}_i^{jk} = Cov(\hat{x}_{ij}, \hat{x}_{ik}), \quad j, k = 1, \ldots, J$$

For the vector $\hat{\boldsymbol{x}}$ the variance-covariance matrix $\boldsymbol{\Sigma}$ depends on the allocation fractions:

$$\Sigma_{jk}(\boldsymbol{\pi}) = N^{-1} \sum_{i=1}^{I} \pi_i^{-1} p_i^2 \hat{s}_i^{jk} \, .$$

We define an overall error function:

$$\omega(\boldsymbol{\pi}) = g(\boldsymbol{\Sigma}(\boldsymbol{\pi})) \, .$$

Relevant Overall Error functions

- $\omega_{MSE}(\pi) = \sum_{j=1}^{J} \Sigma_{jj}(\pi)$, the mean squared error of all estimates,
- $\omega_{MSR}(\pi) = \sum_{j=1}^{J} \hat{x}_j^{-2} \Sigma_{jj}(\pi)$, the mean squared relative error of all estimates,
- $\omega_{MAXE}(\pi) = \max\{j : \Sigma_{jj}(\pi)\}$, the maximum of the squared errors of all estimates,
- $\omega_{MAXR}(\pi) = \max\{j : \hat{x}_j^{-2} \Sigma_{jj}(\pi)\}$, the maximum of the squared relative errors of all estimates,

Relevant Overall Error functions

- $\omega_{MSE}(\pi) = \sum_{j=1}^{J} \Sigma_{jj}(\pi)$, the mean squared error of all estimates,
- $\omega_{MSR}(\pi) = \sum_{j=1}^{J} \hat{x}_j^{-2} \Sigma_{jj}(\pi)$, the mean squared relative error of all estimates,
- $\omega_{MAXE}(\pi) = \max\{j : \Sigma_{jj}(\pi)\}$, the maximum of the squared errors of all estimates,
- $\omega_{MAXR}(\pi) = \max\{j : \hat{x}_j^{-2} \Sigma_{jj}(\pi)\}$, the maximum of the squared relative errors of all estimates,

Relevant Overall Error functions

- $\omega_{MSE}(\pi) = \sum_{j=1}^{J} \Sigma_{jj}(\pi)$, the mean squared error of all estimates,
- $\omega_{MSR}(\pi) = \sum_{j=1}^{J} \hat{x}_j^{-2} \Sigma_{jj}(\pi)$, the mean squared relative error of all estimates,
- $\omega_{MAXE}(\pi) = \max\{j : \Sigma_{jj}(\pi)\}$, the maximum of the squared errors of all estimates,
- $\omega_{MAXR}(\pi) = \max\{j : \hat{x}_j^{-2} \Sigma_{jj}(\pi)\}$, the maximum of the squared relative errors of all estimates,

Relevant Overall Error functions

- $\omega_{MSE}(\pi) = \sum_{j=1}^{J} \Sigma_{jj}(\pi)$, the mean squared error of all estimates,
- $\omega_{MSR}(\pi) = \sum_{j=1}^{J} \hat{x}_j^{-2}\Sigma_{jj}(\pi)$, the mean squared relative error of all estimates,
- $\omega_{MAXE}(\pi) = \max\{j : \Sigma_{jj}(\pi)\}$, the maximum of the squared errors of all estimates,
- $\omega_{MAXR}(\pi) = \max\{j : \hat{x}_j^{-2}\Sigma_{jj}(\pi)\}$, the maximum of the squared relative errors of all estimates,

mean squared relative error

Using the variances defined above we get:

$$\omega_{MSR}(\pi) = \sum_{j=1}^{J} \hat{x}_j^{-2} \Sigma_{jj}(\pi) = N^{-1} \sum_{i=1}^{I} \pi_i^{-1} p_i^2 \sum_{j=1}^{J} \hat{x}_j^{-2} \hat{s}_i^{jj},$$

We know from AOS that for a single estimate:

$$V[\hat{x}_j] = N^{-1} \sum\nolimits_{i=1}^{I} \pi_i^{-1} p_i^2 \sigma_i^2 \geq N^{-1} \left(\sum\nolimits_{i=1}^{I} p_i \sigma_i \right)^2.$$

It attains its lower bound for $\pi_i^* = \dfrac{p_i \sigma_i}{\sum_{k=1}^{I} p_k \sigma_k}$, $i = 1, \ldots, I$.

We can see that $\omega_{MSR}(\pi)$ has the same structure as $V[\hat{x}_j]$.

Replacing σ_i^2 by $\displaystyle\sum_{j=1}^{J} \hat{x}_j^{-2} \hat{s}_i^{jj}$ we thus can minimize $\omega_{MSR}(\pi)$.

mean squared relative error, cont.

$$\omega_{MSR}(\boldsymbol{\pi}) \geq N^{-1}\left(\sum_{i=1}^{I} p_i \left(\sum_{j=1}^{J} \hat{x}_j^{-2} \hat{s}_i^{jj}\right)^{1/2}\right)^2$$

$\omega_{MSR}(\boldsymbol{\pi})$ attains its lower bound selecting

$$\pi_i^* = p_i \left(\sum_{j=1}^{J} \hat{x}_j^{-2} \hat{s}_i^{jj}\right)^{1/2} \Bigg/ \sum_{l=1}^{I} p_l \left(\sum_{j=1}^{J} \hat{x}_j^{-2} \hat{s}_l^{jj}\right)^{1/2}, \quad i = 1, \ldots, I.$$

This idea and the closed form solution can be generalized to all $\omega(\boldsymbol{\pi})$ that are linear functions of the s_i^{jk}.
(Theorem requires non-negativity condition. Assuming positive correlations is no problem for applications to simulation.)

Maximal relative error: Optimization Model

- New objective: Minimize the maximum relative error using the decision variables $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_I)'$.
- We denote $\hat{a}_{ij} = \hat{x}_j^{-2} p_i^2 \hat{\sigma}_{ij}^2$ and add constraints which guarantee that the π_i are positive and sum to one.

$$\begin{aligned} \min \quad & \max \left\{ j : \sum_{i=1}^{I} \frac{\hat{a}_{ij}}{\pi_i} \right\} \\ \text{s.t.} \quad & \sum_{i=1}^{I} \pi_i = 1, \\ & \pi_i \geq 0, \qquad i = 1, \ldots, I. \end{aligned} \quad \Rightarrow \quad \begin{aligned} \min \quad & \omega \\ \text{s.t.} \quad & \omega - \sum_{i=1}^{I} \frac{\hat{a}_{ij}}{\pi_i} \geq 0, \quad j = 1, \ldots, J, \\ & \sum_{i=1}^{I} \pi_i = 1, \\ & \pi_i \geq 0, \qquad i = 1, \ldots, I. \end{aligned}$$

Maximal relative error: Optimization Model

- New objective: Minimize the maximum relative error using the decision variables $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_I)'$.
- We denote $\hat{a}_{ij} = \hat{x}_j^{-2} p_i^2 \hat{\sigma}_{ij}^2$ and add constraints which guarantee that the π_i are positive and sum to one.

$$\begin{aligned} \min \quad & \max \left\{ j : \sum_{i=1}^{I} \frac{\hat{a}_{ij}}{\pi_i} \right\} \\ \text{s.t.} \quad & \sum_{i=1}^{I} \pi_i = 1, \\ & \pi_i > 0, \qquad i = 1, \ldots, I. \end{aligned} \quad \Rightarrow \quad \begin{aligned} \min \quad & \omega \\ \text{s.t.} \quad & \omega - \sum_{i=1}^{I} \frac{\hat{a}_{ij}}{\pi_i} \geq 0, \quad j = 1, \ldots, J, \\ & \sum_{i=1}^{I} \pi_i = 1, \\ & \pi_i > 0, \qquad i = 1, \ldots, I. \end{aligned}$$

Maximal relative error: Optimization Model

- New objective: Minimize the maximum relative error using the decision variables $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_I)'$.
- We denote $\hat{a}_{ij} = \hat{x}_j^{-2} p_i^2 \hat{\sigma}_{ij}^2$ and add constraints which guarantee that the π_i are positive and sum to one.

$$
\begin{aligned}
\min \quad & \max \left\{ j : \sum_{i=1}^{I} \frac{\hat{a}_{ij}}{\pi_i} \right\} \\
\text{s.t.} \quad & \sum_{i=1}^{I} \pi_i = 1, \\
& \pi_i > 0, \qquad\qquad i = 1, \ldots, I.
\end{aligned}
\quad \Rightarrow \quad
\begin{aligned}
\min \quad & \omega \\
\text{s.t.} \quad & \omega - \sum_{i=1}^{I} \frac{\hat{a}_{ij}}{\pi_i} \geq 0, \quad j = 1, \ldots, J, \\
& \sum_{i=1}^{I} \pi_i = 1, \\
& \pi_i > 0, \qquad\qquad i = 1, \ldots, I.
\end{aligned}
$$

Maximal relative error: Optimization Model

- New objective: Minimize the maximum relative error using the decision variables $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_I)'$.
- We denote $\hat{a}_{ij} = \hat{x}_j^{-2} p_i^2 \hat{\sigma}_{ij}^2$ and add constraints which guarantee that the π_i are positive and sum to one.

$$\begin{aligned} \min \quad & \max \left\{ j : \sum_{i=1}^{I} \frac{\hat{a}_{ij}}{\pi_i} \right\} \\ \text{s.t.} \quad & \sum_{i=1}^{I} \pi_i = 1, \\ & \pi_i > 0, \qquad i = 1, \ldots, I. \end{aligned} \quad \Rightarrow \quad \begin{aligned} \min \quad & \omega \\ \text{s.t.} \quad & \omega - \sum_{i=1}^{I} \frac{\hat{a}_{ij}}{\pi_i} \geq 0, \quad j = 1, \ldots, J, \\ & \sum_{i=1}^{I} \pi_i = 1, \\ & \pi_i > 0, \qquad i = 1, \ldots, I. \end{aligned}$$

The Allocation Heuristic

- For the previous optimization model, we have developed a heuristic which searches for a suboptimal solution in the convex hull of the respective optimal solutions π^j, $j = 1, \ldots, J$.

$$
\begin{aligned}
\min \quad & \omega \\
s.t. \quad & \omega - \sum_{i=1}^{I} \frac{\hat{a}_{ij}}{\pi_i} \geq 0, \quad j = 1, \ldots, J, \\
& \pi_i - \sum_{j=1}^{J} \lambda_j \pi_i^j = 0, \quad i = 1, \ldots, I, \\
& \sum_{j=1}^{J} \lambda_j = 1, \\
& \lambda_j \geq 0, \quad j = 1, \ldots, J.
\end{aligned}
$$

- The heuristic method terminates with an average 2 percent sub-optimality for our simulation instances.

The Allocation Heuristic

- For the previous optimization model, we have developed a heuristic which searches for a suboptimal solution in the convex hull of the respective optimal solutions π^j, $j = 1, \ldots, J$.

$$
\begin{aligned}
\min \quad & \omega \\
s.t. \quad & \omega - \sum_{i=1}^{I} \frac{\hat{a}_{ij}}{\pi_i} \geq 0, \quad j = 1, \ldots, J, \\
& \pi_i - \sum_{j=1}^{J} \lambda_j \pi_i^j = 0, \quad i = 1, \ldots, I, \\
& \sum_{j=1}^{J} \lambda_j = 1, \\
& \lambda_j \geq 0, \quad j = 1, \ldots, J.
\end{aligned}
$$

- The heuristic method terminates with an average 2 percent sub-optimality for our simulation instances.

Numerical Results

- We consider $D = 5$ stocks under the t-copula model with Generalized Hyperbolic marginals and $J = 10$ equidistant threshold values.
- For IS we use a mixture of two densities.
- For SIS, we simply used a single IS density selected for the threshold $\tau^* = 0.75\tau_{max} + 0.25\tau_{min}$.
- The parameters of the distributions were estimated from NYSE data.

Numerical Results

- We consider $D = 5$ stocks under the t-copula model with Generalized Hyperbolic marginals and $J = 10$ equidistant threshold values.
- For IS we use a mixture of two densities.
- For SIS, we simply used a single IS density selected for the threshold $\tau^* = 0.75\tau_{max} + 0.25\tau_{min}$.
- The parameters of the distributions were estimated from NYSE data.

Numerical Results

- We consider $D = 5$ stocks under the t-copula model with Generalized Hyperbolic marginals and $J = 10$ equidistant threshold values.
- For IS we use a mixture of two densities.
- For SIS, we simply used a single IS density selected for the threshold $\tau^* = 0.75\tau_{max} + 0.25\tau_{min}$.
- The parameters of the distributions were estimated from NYSE data.

Numerical Results

- We consider $D = 5$ stocks under the t-copula model with Generalized Hyperbolic marginals and $J = 10$ equidistant threshold values.
- For IS we use a mixture of two densities.
- For SIS, we simply used a single IS density selected for the threshold $\tau^* = 0.75\tau_{max} + 0.25\tau_{min}$.
- The parameters of the distributions were estimated from NYSE data.

Numerical Results cont.

- Following table shows the percentage relative errors and the variance reduction factors obtained under NV, IS and SIS. The execution times are 0.75, 0.95 and 1.25 seconds respectively.

		NV	IS		SIS	
τ	$\Pr(r(\mathbf{T}) \leq \tau)$	RE	RE	VR	RE	VR
0.895	0.0010	$\pm$19.89%	$\pm$2.65%	56.3	$\pm$0.77%	667.2
0.909	0.0014	$\pm$16.67%	$\pm$2.37%	49.5	$\pm$0.70%	567.1
0.917	0.0020	$\pm$14.13%	$\pm$2.12%	44.4	$\pm$0.69%	419.4
0.924	0.0028	$\pm$12.09%	$\pm$1.89%	40.9	$\pm$0.68%	316.1
0.932	0.0040	$\pm$9.89%	$\pm$1.77%	31.2	$\pm$0.66%	224.5
0.939	0.0060	$\pm$7.96%	$\pm$1.72%	21.4	$\pm$0.69%	133.1
0.947	0.0094	$\pm$6.46%	$\pm$1.82%	12.6	$\pm$0.66%	95.8
0.954	0.0154	$\pm$4.91%	$\pm$2.14%	5.3	$\pm$0.65%	57.1
0.962	0.0267	$\pm$3.75%	$\pm$2.26%	2.8	$\pm$0.64%	34.3
0.973	0.0500	$\pm$2.69%	$\pm$9.23%	0.1	$\pm$0.69%	15.2

Estimating Conditional Excess

- Simulating the conditional excess: $E\left[Loss|Loss \geq \tau\right]$ requires a ratio estimate; this makes variance reduction more difficult.
- Literature: "Use the same IS density as for tail-loss probabilities."
- The variance of the ratio estimate: $V\left[\hat{x}_1/\hat{x}_2\right] \approx \hat{x}_1^2 \hat{x}_2^{-4} \Sigma_{22}\left(\pi\right) - 2\hat{x}_1 \hat{x}_2^{-3} \Sigma_{12}\left(\pi\right) + \hat{x}_2^{-2} \Sigma_{11}\left(\pi\right) = \omega\left(\pi\right).$
- To reach optimal allocation for stratification we can use the theorem above to minimize the variance of a ratio estimate.

For Conditional Excess we obtained stratification and stratified IS algorithms with optimal allocation for:

- a single threshold
- for several thresholds minimizing the mean squared relative error.

Conclusions

- For practically relevant examples SIS (combination of IS and stratification) increases the efficiency of tail loss probability estimates under the t-copula model.
- Compared to the methods in the literature, the variance of the estimates are substantially reduced without a significant increase in the execution time.
- The SIS method allows the efficient estimation of tail loss probabilities for multiple threshold values in a single simulation.

Conclusions

- For practically relevant examples SIS (combination of IS and stratification) increases the efficiency of tail loss probability estimates under the *t*-copula model.
- Compared to the methods in the literature, the variance of the estimates are substantially reduced without a significant increase in the execution time.
- The SIS method allows the efficient estimation of tail loss probabilities for multiple threshold values in a single simulation.

Conclusions

- For practically relevant examples SIS (combination of IS and stratification) increases the efficiency of tail loss probability estimates under the t-copula model.
- Compared to the methods in the literature, the variance of the estimates are substantially reduced without a significant increase in the execution time.
- The SIS method allows the efficient estimation of tail loss probabilities for multiple threshold values in a single simulation.

Conclusions

- The mean square (relative) error of all estimates can be minimized using the variance estimates of a pilot run and a simple closed form formula for the allocation fractions.
- To minimize the maximal squared (relative) error of all estimates we have developed a fast and simple heuristic to find close to optimal allocation fractions.
- The idea to obtain optimal variance reduction for several estimates in a single simulation using optimal allocation stratification (OAS) can be used for all simulation problems for which OAS leads good variance reduction.

Conclusions

- The mean square (relative) error of all estimates can be minimized using the variance estimates of a pilot run and a simple closed form formula for the allocation fractions.
- To minimize the maximal squared (relative) error of all estimates we have developed a fast and simple heuristic to find close to optimal allocation fractions.
- The idea to obtain optimal variance reduction for several estimates in a single simulation using optimal allocation stratification (OAS) can be used for all simulation problems for which OAS leads good variance reduction.

Conclusions

- The mean square (relative) error of all estimates can be minimized using the variance estimates of a pilot run and a simple closed form formula for the allocation fractions.
- To minimize the maximal squared (relative) error of all estimates we have developed a fast and simple heuristic to find close to optimal allocation fractions.
- The idea to obtain optimal variance reduction for several estimates in a single simulation using optimal allocation stratification (OAS) can be used for all simulation problems for which OAS leads good variance reduction.

Questions?

THANK YOU !

References

- Başoğlu, İ., W. Hörmann, and H. Sak. 2013.
  "Optimally Stratified Importance Sampling for Portfolio Risk with Multiple Loss Thresholds".
  Optimization 62 (11): 1451–1471.
- Başoğlu, İ. and W. Hörmann, 2014.
  "Efficient stratified sampling implementations in multiresponse simulation".
  submitted to Winter Simulation Conference 2014.
- Etoré, P., and B. Jourdain. 2010.
  "Adaptive Optimal Allocation in Stratified Sampling Methods".
  Methodology and Computing in Applied Probability 12 (3): 335–360.
- Sak, H., and W. Hörmann. 2012.
  "Fast Simulations in Credit Risk".
  Quantitative Finance 12 (10): 1557–1569.