Computing Capital Requirements with Guarantees

Patrick Cheridito and Moritz Weiss

RiskLab, ETH Zurich

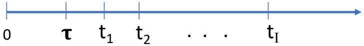
Insurance Data Science Conference, Stockholm

June 18, 2024

Asset-Liability Risk

0 = current time τ = risk horizon (e.g. 1 year)

Asset-Liability Risk



0 = current time $\quad \tau =$ risk horizon (e.g. 1 year)

- All relevant information at time τ is given by a random vector $X = (X_1, \ldots, X_d)$

Asset-Liability Risk

0 = current time τ = risk horizon (e.g. 1 year)

- All relevant information at time τ is given by a random vector $X = (X_1, \ldots, X_d)$
- Portfolio value at time τ:

$$V = v(X) + \mathbb{E}^{\mathbb{Q}} \left[\sum_{i=1}^{I} \frac{N_\tau}{N_{t_i}} C_{t_i} \,\middle|\, X \right]$$

Asset-Liability Risk

0 = current time τ = risk horizon (e.g. 1 year)

- All relevant information at time τ is given by a random vector $X = (X_1, \dots, X_d)$
- Portfolio value at time τ:

$$V = v(X) + \mathbb{E}^{\mathbb{Q}}\left[\sum_{i=1}^{I} \frac{N_\tau}{N_{t_i}} C_{t_i} \,\middle|\, X\right], \text{ where}$$

$v : \mathbb{R}^d \to \mathbb{R}$, $\mathbb{Q}$ is a pricing measure, $0 < \tau < t_1 < \dots < t_I$, C_{t_i} are future cash flows, N_{t_i} is a numeraire process

Asset-Liability Risk

0 = current time τ = risk horizon (e.g. 1 year)

- All relevant information at time τ is given by a random vector $X = (X_1, \ldots, X_d)$
- Portfolio value at time τ:

$$V = v(X) + \mathbb{E}^{\mathbb{Q}}\left[\sum_{i=1}^{I} \frac{N_\tau}{N_{t_i}} C_{t_i} \,\middle|\, X\right], \text{ where}$$

$v : \mathbb{R}^d \to \mathbb{R}$, $\mathbb{Q}$ is a pricing measure, $0 < \tau < t_1 < \ldots < t_I$, C_{t_i} are future cash flows, N_{t_i} is a numeraire process

- Loss at time τ: $L = -V = \mathbb{E}^{\mathbb{Q}}[Y \mid X]$

Asset-Liability Risk

0 = current time τ = risk horizon (e.g. 1 year)

- All relevant information at time τ is given by a random vector $X = (X_1, \ldots, X_d)$
- Portfolio value at time τ:

$$V = v(X) + \mathbb{E}^{\mathbb{Q}}\left[\sum_{i=1}^{I} \frac{N_\tau}{N_{t_i}} C_{t_i} \,\middle|\, X\right], \text{ where}$$

$v : \mathbb{R}^d \to \mathbb{R}$, $\mathbb{Q}$ is a pricing measure, $0 < \tau < t_1 < \ldots < t_I$, C_{t_i} are future cash flows, N_{t_i} is a numeraire process

- Loss at time τ: $L = -V = \mathbb{E}^{\mathbb{Q}}[Y \mid X]$, where

$$Y = -v(X) - \sum_{i=1}^{I} \frac{N_\tau}{N_{t_i}} C_{t_i}$$

Asset-Liability Risk

0 = current time τ = risk horizon (e.g. 1 year)

- All relevant information at time τ is given by a random vector $X = (X_1, \ldots, X_d)$
- Portfolio value at time τ:

$$V = v(X) + \mathbb{E}^{\mathbb{Q}}\left[\sum_{i=1}^{I} \frac{N_\tau}{N_{t_i}} C_{t_i} \,\middle|\, X\right],$$ where

$v : \mathbb{R}^d \to \mathbb{R}$, $\mathbb{Q}$ is a pricing measure, $0 < \tau < t_1 < \ldots < t_I$, C_{t_i} are future cash flows, N_{t_i} is a numeraire process

- Loss at time τ: $L = -V = \mathbb{E}^{\mathbb{Q}}[Y \mid X]$, where

$$Y = -v(X) - \sum_{i=1}^{I} \frac{N_\tau}{N_{t_i}} C_{t_i}$$

- We are interested in *Value-at-Risk* (for $\alpha = 99.5\%$) and *Expected Shortfall* (for $\alpha = 99\%$)

Asset-Liability Risk

0 = current time $\quad \tau$ = risk horizon (e.g. 1 year)

- All relevant information at time τ is given by a random vector $X = (X_1, \dots, X_d)$
- Portfolio value at time τ:

$$V = v(X) + \mathbb{E}^{\mathbb{Q}}\left[\sum_{i=1}^{I} \frac{N_\tau}{N_{t_i}} C_{t_i} \;\middle|\; X\right], \text{ where}$$

$v : \mathbb{R}^d \to \mathbb{R}$, $\mathbb{Q}$ is a pricing measure, $0 < \tau < t_1 < \dots < t_I$, C_{t_i} are future cash flows, N_{t_i} is a numeraire process

- Loss at time τ: $L = -V = \mathbb{E}^{\mathbb{Q}}[Y \mid X]$, where

$$Y = -v(X) - \sum_{i=1}^{I} \frac{N_\tau}{N_{t_i}} C_{t_i}$$

- We are interested in *Value-at-Risk* (for $\alpha = 99.5\%$) and *Expected Shortfall* (for $\alpha = 99\%$)

$$\mathrm{VaR}_\alpha(L) := \min\{x \in \mathbb{R} : \mathbb{P}[L \le x] \ge \alpha\} \text{ and } \mathrm{ES}_\alpha(L) := \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_u(L)du$$

Asset-Liability Risk

0 = current time $\tau =$ risk horizon (e.g. 1 year)

- All relevant information at time τ is given by a random vector $X = (X_1, \ldots, X_d)$
- Portfolio value at time τ:

$$V = v(X) + \mathbb{E}^{\mathbb{Q}}\left[\sum_{i=1}^{I} \frac{N_\tau}{N_{t_i}} C_{t_i} \,\middle|\, X \right], \text{ where}$$

$v : \mathbb{R}^d \to \mathbb{R}$, $\mathbb{Q}$ is a pricing measure, $0 < \tau < t_1 < \ldots < t_I$, C_{t_i} are future cash flows, N_{t_i} is a numeraire process

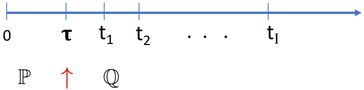
- Loss at time τ: $L = -V = \mathbb{E}^{\mathbb{Q}}[Y \mid X]$, where

$$Y = -v(X) - \sum_{i=1}^{I} \frac{N_\tau}{N_{t_i}} C_{t_i}$$

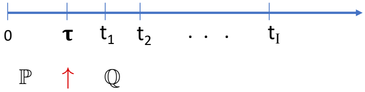
- We are interested in *Value-at-Risk* (for $\alpha = 99.5\%$) and *Expected Shortfall* (for $\alpha = 99\%$)

$$\mathrm{VaR}_\alpha(L) := \min\{x \in \mathbb{R} : \mathbb{P}[L \le x] \ge \alpha\} \text{ and } \mathrm{ES}_\alpha(L) := \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_u(L)du \approx \mathbb{E}[L \mid L \ge \mathrm{VaR}_\alpha(L)]$$

Pasting together the *real world measure* $\mathbb{P}$ **and the** *pricing measure* $\mathbb{Q}$



Pasting together the *real world measure* $\mathbb{P}$ **and the** *pricing measure* $\mathbb{Q}$

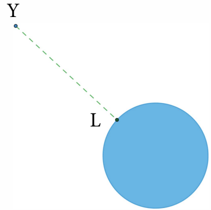


$$\mathbb{P} \otimes \mathbb{Q}[A] = \int_{\mathbb{R}^d} \mathbb{Q}[A \mid X = x] \pi(dx), \quad A \in \mathcal{F},$$

where π is the distribution of $X = (X_1, \ldots, X_d)$

Conditional Expectations as Minimizing Functions

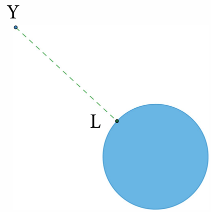
- $L = \mathbb{E}^{\mathbb{P}\otimes\mathbb{Q}}[Y \mid X]$ is the L^2-projection of Y on $L^2(\mathcal{F}^X)$



Conditional Expectations as Minimizing Functions

- $L = \mathbb{E}^{\mathbb{P}\otimes\mathbb{Q}}[Y \mid X]$ is the L^2-projection of Y on $L^2(\mathcal{F}^X)$

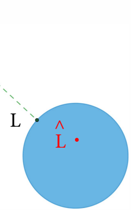
$$\mathbb{E}^{\mathbb{P}\otimes\mathbb{Q}}\left[(Y-L)^2\right] = \min_{f\,:\,\mathbb{R}^d\to\mathbb{R}} \mathbb{E}^{\mathbb{P}\otimes\mathbb{Q}}\left[(Y-f(X))^2\right]$$



Conditional Expectations as Minimizing Functions

- $L = \mathbb{E}^{\mathbb{P}\otimes\mathbb{Q}}[Y \mid X]$ is the L^2-projection of Y on $L^2(\mathcal{F}^X)$

$$\mathbb{E}^{\mathbb{P}\otimes\mathbb{Q}}\left[(Y-L)^2\right] = \min_{f\,:\,\mathbb{R}^d\to\mathbb{R}} \mathbb{E}^{\mathbb{P}\otimes\mathbb{Q}}\left[(Y-f(X))^2\right]$$



- **Least Squares Monte Carlo**

simulate (X^j, Y^j) and solve $\min_{f\in\mathcal{S}} \frac{1}{J}\sum_{j=1}^{J}\left(Y^j - f(X^j)\right)^2$

over a subfamily $\mathcal{S}$ of all Borel functions $f\colon \mathbb{R}^d \to \mathbb{R}$.

Least Squares Regression

- linear regression on polynomials Longstaff and Schwartz (2001), Ha and Bauer (2021)

Least Squares Regression

- linear regression on polynomials Longstaff and Schwartz (2001), Ha and Bauer (2021)
- regression trees Boudabsa and Filipović (2022)

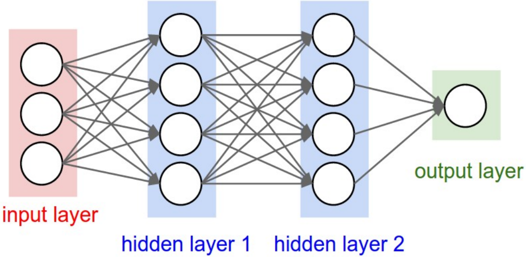
Least Squares Regression

- linear regression on polynomials Longstaff and Schwartz (2001), Ha and Bauer (2021)
- regression trees Boudabsa and Filipović (2022)
- neural network regression Kohlen et al (2010), Fiore et al. (2018), Cheridito et al. (2020)

Here, we minimize

$$\theta \mapsto \frac{1}{J} \sum_{j=1}^{J} \left(Y^j - f_\theta(X^j)\right)^2$$

over a set of *neural networks* $f_\theta : \mathbb{R}^d \to \mathbb{R}, \quad \theta \in \mathbb{R}^q$



Monte Carlo Estimation of VaR and ES

- Let $X^1, \ldots, X^n$ be *independent $\mathbb{P}$-simulations* of $X = (X_1, \ldots, X_d)$

Monte Carlo Estimation of VaR and ES

- Let $X^1, \ldots, X^n$ be *independent $\mathbb{P}$-simulations* of $X = (X_1, \ldots, X_d)$

- Denote by $X^{(1)}, \ldots, X^{(n)}$ the *reordered sample* so that

$$L^{(1)} = f_\theta(X^{(1)}) \geq \ldots \geq L^{(n)} = f_\theta(X^{(n)})$$

Monte Carlo Estimation of VaR and ES

- Let $X^1, \ldots, X^n$ be *independent $\mathbb{P}$-simulations* of $X = (X_1, \ldots, X_d)$

- Denote by $X^{(1)}, \ldots, X^{(n)}$ the *reordered sample* so that

$$L^{(1)} = f_\theta(X^{(1)}) \geq \ldots \geq L^{(n)} = f_\theta(X^{(n)})$$

- Apply *VaR_α* and *ES_α* to the *empirical measure* $$\frac{1}{n} \sum_{i=1}^{n} \delta_{L^{(i)}}$$

Monte Carlo Estimation of VaR and ES

- Let $X^1, \dots, X^n$ be *independent $\mathbb{P}$-simulations* of $X = (X_1, \dots, X_d)$

- Denote by $X^{(1)}, \dots, X^{(n)}$ the *reordered sample* so that

$$L^{(1)} = f_\theta(X^{(1)}) \geq \ldots \geq L^{(n)} = f_\theta(X^{(n)})$$

- Apply *VaR_α* and *ES_α* to the *empirical measure* $\frac{1}{n}\sum_{i=1}^{n} \delta_{L^{(i)}}$

$$\rightsquigarrow \qquad \widehat{\mathrm{VaR}}_\alpha(n) = L^{(j)} \quad \text{and} \quad \widehat{\mathrm{ES}}_\alpha(n) = \frac{1}{1-\alpha}\sum_{i=1}^{j-1}\frac{L^{(i)}}{n} + \left(1 - \frac{j-1}{(1-\alpha)n}\right)L^{(j)},$$

where

$$j = \min\{i \in \{1, \dots, n\} : i/n > 1-\alpha\}$$

Monte Carlo Estimation of VaR and ES

- Let $X^1, \ldots, X^n$ be *independent $\mathbb{P}$-simulations* of $X = (X_1, \ldots, X_d)$

- Denote by $X^{(1)}, \ldots, X^{(n)}$ the *reordered sample* so that

$$L^{(1)} = f_\theta(X^{(1)}) \geq \ldots \geq L^{(n)} = f_\theta(X^{(n)})$$

- Apply *VaR_α* and *ES_α* to the *empirical measure* $\frac{1}{n} \sum_{i=1}^{n} \delta_{L^{(i)}}$

$$\rightsquigarrow \qquad \widehat{\mathrm{VaR}}_\alpha(n) = L^{(j)} \quad \text{and} \quad \widehat{\mathrm{ES}}_\alpha(n) = \frac{1}{1-\alpha} \sum_{i=1}^{j-1} \frac{L^{(i)}}{n} + \left(1 - \frac{j-1}{(1-\alpha)n}\right) L^{(j)},$$

where

$$j = \min \{i \in \{1, \ldots, n\} : i/n > 1 - \alpha\}$$

there exist convergence rates

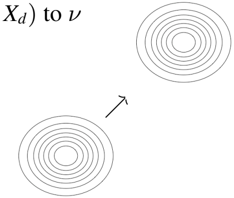
see. e.g., David and Nagaraja (2003) and Zwingmann and Holzmann (2016)

Importance Sampling

- *Sample more frequently from the tail of L* when estimating VaR_α and ES_α

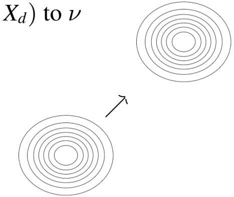
Importance Sampling

- *Sample more frequently from the tail of L* when estimating VaR_α and ES_α
- by shifting the original distribution of $X = (X_1, \ldots, X_d)$ to ν

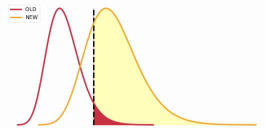


Importance Sampling

- *Sample more frequently from the tail of L* when estimating VaR_α and ES_α
- by shifting the original distribution of $X = (X_1, \ldots, X_d)$ to ν

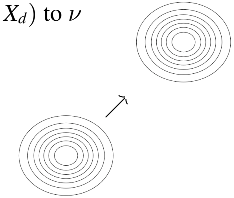


- so that $L = f(X)$ has more weight in the tail

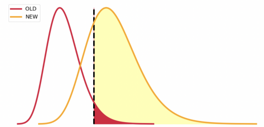


Importance Sampling

- *Sample more frequently from the tail of L* when estimating VaR_α and ES_α
- by shifting the original distribution of $X = (X_1, \ldots, X_d)$ to ν



- so that $L = f(X)$ has more weight in the tail



- note that $f = \arg\min_{\varphi\,:\,\mathbb{R}^d \to \mathbb{R}} \mathbb{E}^{\mathbb{P}^\nu \otimes \mathbb{Q}}\left[(Y - \varphi(X))^2\right]$

Example 1: Portfolio of Call and Put Options

- stocks $dS_t^i = \mu_i S_t^i dt + \sigma_i S_t^i dW_t^{\mathbb{P},i} = r S_t^i dt + \sigma_i S_t^i dW_t^{\mathbb{Q},i}, \quad i = 1, \ldots, 20,$

Example 1: Portfolio of Call and Put Options

- stocks $dS_t^i = \mu_i S_t^i dt + \sigma_i S_t^i dW_t^{\mathbb{P},i} = rS_t^i dt + \sigma_i S_t^i dW_t^{\mathbb{Q},i}, \quad i = 1, \ldots, 20,$
- call options $(S_T^i - K)^+, \quad i = 1, \ldots 10,$

Example 1: Portfolio of Call and Put Options

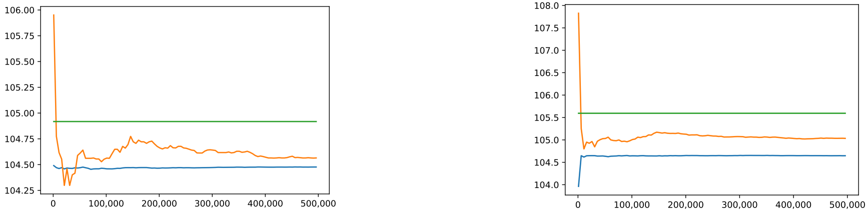
- stocks $dS_t^i = \mu_i S_t^i dt + \sigma_i S_t^i dW_t^{\mathbb{P},i} = r S_t^i dt + \sigma_i S_t^i dW_t^{\mathbb{Q},i}, \quad i = 1, \ldots, 20,$
- call options $(S_T^i - K)^+, \quad i = 1, \ldots 10,$ put options $(K - S_T^i)^+, \quad i = 11, \ldots 20,$

Example 1: Portfolio of Call and Put Options

- stocks $dS_t^i = \mu_i S_t^i dt + \sigma_i S_t^i dW_t^{\mathbb{P},i} = r S_t^i dt + \sigma_i S_t^i dW_t^{\mathbb{Q},i}, \quad i = 1, \ldots, 20,$
- call options $(S_T^i - K)^+, \quad i = 1, \ldots 10,$ put options $(K - S_T^i)^+, \quad i = 11, \ldots 20,$
- time-τ loss $L = e^{-r(T-\tau)}\, \mathbb{E}^{\mathbb{Q}} \left[\sum_{i=1}^{10} (S_T^i - K)^+ + \sum_{i=11}^{20} (K - S_T^i)^+ \,\middle|\, S_\tau^1, \ldots, S_\tau^{20} \right]$

Example 1: Portfolio of Call and Put Options

- stocks $dS_t^i = \mu_i S_t^i dt + \sigma_i S_t^i dW_t^{\mathbb{P},i} = r S_t^i dt + \sigma_i S_t^i dW_t^{\mathbb{Q},i}, \quad i = 1, \dots, 20,$
- call options $(S_T^i - K)^+, \quad i = 1, \dots 10,$ put options $(K - S_T^i)^+, \quad i = 11, \dots 20,$
- time-τ loss $L = e^{-r(T-\tau)}\, \mathbb{E}^{\mathbb{Q}}\left[\sum_{i=1}^{10} (S_T^i - K)^+ + \sum_{i=11}^{20} (K - S_T^i)^+ \;\middle|\; S_\tau^1, \dots, S_\tau^{20} \right]$



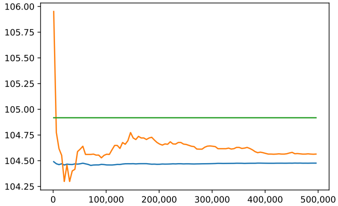
empirical 99.5%-VaR

empirical 99%-ES

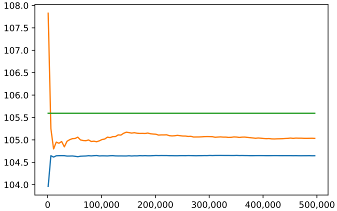
without IS with IS reference values obtained from Black–Scholes

Example 1: Portfolio of Call and Put Options

- stocks $dS^i_t = \mu_i S^i_t dt + \sigma_i S^i_t dW^{\mathbb{P},i}_t = r S^i_t dt + \sigma_i S^i_t dW^{\mathbb{Q},i}_t, \quad i = 1, \dots, 20,$
- call options $(S^i_T - K)^+, \quad i = 1, \dots 10,$ put options $(K - S^i_T)^+, \quad i = 11, \dots 20,$
- time-τ loss $L = e^{-r(T-\tau)} \, \mathbb{E}^{\mathbb{Q}} \left[\sum_{i=1}^{10} (S^i_T - K)^+ + \sum_{i=11}^{20} (K - S^i_T)^+ \;\middle|\; S^1_\tau, \dots, S^{20}_\tau \right]$



empirical 99.5%-VaR



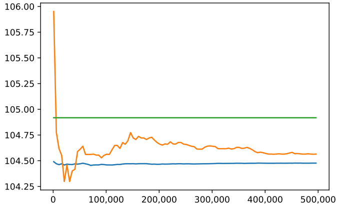
empirical 99%-ES

without IS with IS reference values obtained from Black–Scholes

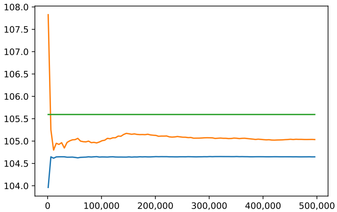
Approximation Error I: $L = f(X) \approx \hat{L} = f_\theta(X)$

Example 1: Portfolio of Call and Put Options

- stocks $dS_t^i = \mu_i S_t^i dt + \sigma_i S_t^i dW_t^{\mathbb{P},i} = r S_t^i dt + \sigma_i S_t^i dW_t^{\mathbb{Q},i}, \quad i = 1, \dots, 20,$
- call options $(S_T^i - K)^+, \quad i = 1, \dots 10,$ put options $(K - S_T^i)^+, \quad i = 11, \dots 20,$
- time-τ loss $L = e^{-r(T-\tau)}\, \mathbb{E}^{\mathbb{Q}} \left[\sum_{i=1}^{10} (S_T^i - K)^+ + \sum_{i=11}^{20} (K - S_T^i)^+ \,\middle|\, S_\tau^1, \dots, S_\tau^{20} \right]$



empirical 99.5%-VaR



empirical 99%-ES

without IS with IS reference values obtained from Black–Scholes

Approximation Error I: $L = f(X) \approx \hat{L} = f_\theta(X)$

Approximation Error II: $\mathrm{ES}_\alpha(\hat{L}) \approx \widehat{\mathrm{ES}}_\alpha(\hat{L})$

Example 2: Variable Annuity with GMIB (Guaranteed Minimum Income Benefit)

Example 2: Variable Annuity with GMIB (Guaranteed Minimum Income Benefit)

(Ha and Bauer, 2021)

value of the annuity at time T : $\max\{S_T = e^{q_T}\ ,\ b\, a_{x+T}(T)\}$

Example 2: Variable Annuity with GMIB (Guaranteed Minimum Income Benefit)

(Ha and Bauer, 2021)

value of the annuity at time T : $\max\{S_T = e^{q_T} \ , \ b\, a_{x+T}(T)\}$

$$L = \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_\tau^T r_s + \mu_{x+s} ds} \max\{e^{q_T}, b\, a_{x+T}(T)\} \mid q_\tau,\ r_\tau,\ \mu_{x+\tau}\right]$$

where $a_{x+T}(T) =$ time-T value of a life-time annuity and

$q_\tau =$ log-stock index, $r_\tau =$ interest rate, $\mu_{x+\tau} =$ mortality rate

Example 2: Variable Annuity with GMIB (Guaranteed Minimum Income Benefit)

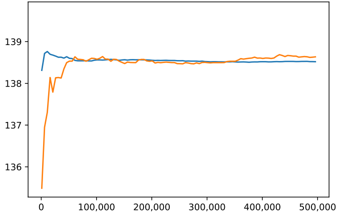
(Ha and Bauer, 2021)

value of the annuity at time T : $\max\left\{S_T = e^{q_T}\ ,\ b\, a_{x+T}(T)\right\}$

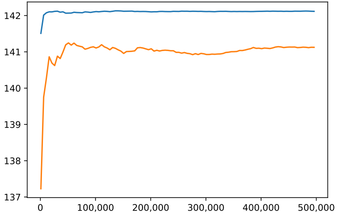
$$L = \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_\tau^T r_s + \mu_{x+s} ds} \max\left\{e^{q_T}, b\, a_{x+T}(T)\right\} \mid q_\tau,\, r_\tau,\, \mu_{x+\tau}\right]$$

where $a_{x+T}(T) =$ time-T value of a life-time annuity and

$q_\tau =$ log-stock index, $r_\tau =$ interest rate, $\mu_{x+\tau} =$ mortality rate



empirical 99.5%-VaR



empirical 99%-ES

without IS with IS

Example 2: Variable Annuity with GMIB (Guaranteed Minimum Income Benefit)

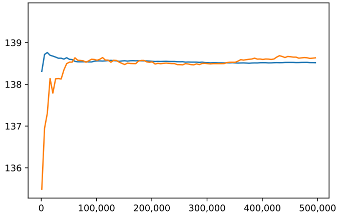
(Ha and Bauer, 2021)

value of the annuity at time T : $\max\left\{S_T = e^{q_T}\ ,\ b\,a_{x+T}(T)\right\}$

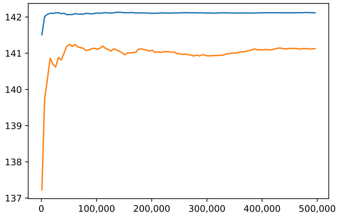
$$L = \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_\tau^T r_s + \mu_{x+s} ds} \max\left\{e^{q_T}, b\,a_{x+T}(T)\right\} \mid q_\tau,\ r_\tau,\ \mu_{x+\tau}\right]$$

where $a_{x+T}(T) =$ time-T value of a life-time annuity and

$q_\tau =$ log-stock index, $r_\tau =$ interest rate, $\mu_{x+\tau} =$ mortality rate



empirical 99.5%-VaR



empirical 99%-ES

without IS with IS

Approximation Error I: $L = f(X) \approx \hat{L} = f_\theta(X)$ *black box!*

Example 2: Variable Annuity with GMIB (Guaranteed Minimum Income Benefit)

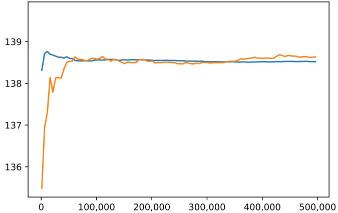
(Ha and Bauer, 2021)

value of the annuity at time T : $\max\left\{S_T = e^{q_T}\ ,\ b\, a_{x+T}(T)\right\}$

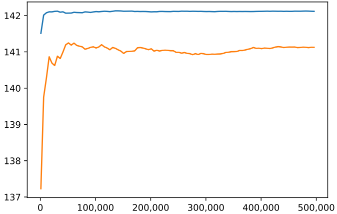
$$L = \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_\tau^T r_s+\mu_{x+s}ds}\max\left\{e^{q_T}, b\, a_{x+T}(T)\right\} \mid q_\tau,\, r_\tau,\, \mu_{x+\tau}\right]$$

where $a_{x+T}(T) =$ time-T value of a life-time annuity and

$q_\tau =$ log-stock index, $r_\tau =$ interest rate, $\mu_{x+\tau} =$ mortality rate



empirical 99.5%-VaR



empirical 99%-ES

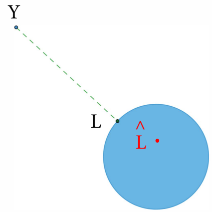
without IS with IS

Approximation Error I: $L = f(X) \approx \hat{L} = f_\theta(X)$ *black box!*

Approximation Error II: $\mathrm{ES}_\alpha(\hat{L}) \approx \widehat{\mathrm{ES}}_\alpha(\hat{L})$ *well understood*

Computation of Conditional Expectations with Guarantees

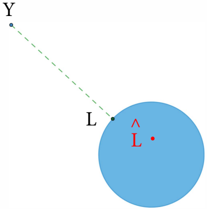
Goal   Derive an *alternative representation* of the minimal L^2-distance



Computation of Conditional Expectations with Guarantees

Goal Derive an *alternative representation* of the minimal L^2-distance

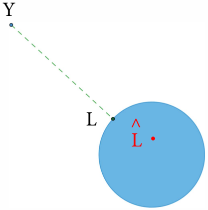
$$\min_{f\,:\,\mathbb{R}^d \to \mathbb{R}} \|Y - f(X)\|_2 = \|Y - \mathbb{E}[Y \mid X]\|_2$$



Computation of Conditional Expectations with Guarantees

Goal   Derive an *alternative representation* of the minimal L^2-distance

$$\min_{f : \mathbb{R}^d \to \mathbb{R}} \|Y - f(X)\|_2 = \|Y - \mathbb{E}[Y \mid X]\|_2$$

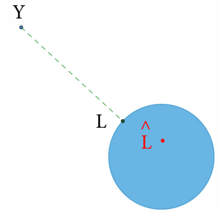


Assumption Y is of the form $Y = h(X, V)$ for a known function $h : \mathbb{R}^{d+k} \to \mathbb{R}$ and a k-dim random vector V independent of X

Computation of Conditional Expectations with Guarantees

Goal Derive an *alternative representation* of the minimal L^2-distance

$$\min_{f:\mathbb{R}^d\to\mathbb{R}} \|Y - f(X)\|_2 = \|Y - \mathbb{E}[Y \mid X]\|_2$$



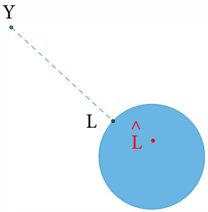
Assumption Y is of the form $Y = h(X,V)$ for a known function $h:\mathbb{R}^{d+k}\to\mathbb{R}$ and a k-dim random vector V independent of X

Define $Z = h(X,\tilde{V})$ for an independent copy $\tilde{V}$ of V

Computation of Conditional Expectations with Guarantees

Goal   Derive an *alternative representation* of the minimal L^2-distance

$$\min_{f:\mathbb{R}^d\to\mathbb{R}} \|Y - f(X)\|_2 = \|Y - \mathbb{E}[Y \mid X]\|_2$$



Assumption Y is of the form $Y = h(X,V)$ for a known function $h : \mathbb{R}^{d+k} \to \mathbb{R}$ and a k-dim random vector V independent of X

Define $Z = h(X, \tilde{V})$ for an independent copy $\tilde{V}$ of V

Theorem $\|Y - \mathbb{E}[Y \mid X]\|_2^2 = \mathbb{E}[Y(Y - Z)]$

- For any *candidate regression function* $\hat{f}\colon \mathbb{R}^d \to \mathbb{R}$ one has

$$\left\|Y - \hat{f}(X)\right\|^2_{L^2(\mathbb{P})} \;\geq\; \|Y - \mathbb{E}[Y \mid X]\|^2_{L^2(\mathbb{P})} \;=\; \mathbb{E}[Y(Y-Z)]$$

- For any *candidate regression function* $\hat{f}\colon \mathbb{R}^d \to \mathbb{R}$ one has

$$\left\| Y - \hat{f}(X) \right\|_{L^2(\mathbb{P})}^2 \geq \| Y - \mathbb{E}[Y \mid X] \|_{L^2(\mathbb{P})}^2 = \mathbb{E}[Y(Y - Z)]$$

- By *Pythagoras* $\| Y - \mathbb{E}[Y \mid X] \|_{L^2(\mathbb{P})}^2 + \left\| \mathbb{E}[Y \mid X] - \hat{f}(X) \right\|_{L^2(\mathbb{P})}^2 = \left\| Y - \hat{f}(X) \right\|_{L^2(\mathbb{P})}^2$

- For any *candidate regression function* $\hat{f}\colon \mathbb{R}^d \to \mathbb{R}$ one has

$$\left\|Y - \hat{f}(X)\right\|^2_{L^2(\mathbb{P})} \geq \|Y - \mathbb{E}[Y \mid X]\|^2_{L^2(\mathbb{P})} = \mathbb{E}[Y(Y-Z)]$$

- By *Pythagoras* $$\|Y - \mathbb{E}[Y \mid X]\|^2_{L^2(\mathbb{P})} + \left\|\mathbb{E}[Y \mid X] - \hat{f}(X)\right\|^2_{L^2(\mathbb{P})} = \left\|Y - \hat{f}(X)\right\|^2_{L^2(\mathbb{P})}$$

- Therefore $$\left\|\mathbb{E}[Y \mid X] - \hat{f}(X)\right\|^2_{L^2(\mathbb{P})} = \left\|Y - \hat{f}(X)\right\|^2_{L^2(\mathbb{P})} - \mathbb{E}[Y(Y-Z)]$$

- For any *candidate regression function* $\hat{f}\colon \mathbb{R}^d \to \mathbb{R}$ one has

$$\left\| Y - \hat{f}(X) \right\|^2_{L^2(\mathbb{P})} \;\geq\; \| Y - \mathbb{E}[Y \mid X] \|^2_{L^2(\mathbb{P})} \;=\; \mathbb{E}[Y(Y-Z)]$$

- By *Pythagoras* $$\| Y - \mathbb{E}[Y \mid X] \|^2_{L^2(\mathbb{P})} + \left\| \mathbb{E}[Y \mid X] - \hat{f}(X) \right\|^2_{L^2(\mathbb{P})} = \left\| Y - \hat{f}(X) \right\|^2_{L^2(\mathbb{P})}$$

- Therefore $$\left\| \mathbb{E}[Y \mid X] - \hat{f}(X) \right\|^2_{L^2(\mathbb{P})} \;=\; \left\| Y - \hat{f}(X) \right\|^2_{L^2(\mathbb{P})} - \mathbb{E}[Y(Y-Z)]$$

$\uparrow$ $\nwarrow$ $\nearrow$

L^2-approximation error *can be estimated*

- For any *candidate regression function* $\hat{f}\colon \mathbb{R}^d \to \mathbb{R}$ one has

$$\left\|Y - \hat{f}(X)\right\|^2_{L^2(\mathbb{P})} \geq \|Y - \mathbb{E}[Y \mid X]\|^2_{L^2(\mathbb{P})} = \mathbb{E}[Y(Y-Z)]$$

- By *Pythagoras* $$\|Y - \mathbb{E}[Y \mid X]\|^2_{L^2(\mathbb{P})} + \left\|\mathbb{E}[Y \mid X] - \hat{f}(X)\right\|^2_{L^2(\mathbb{P})} = \left\|Y - \hat{f}(X)\right\|^2_{L^2(\mathbb{P})}$$

- Therefore $$\left\|\mathbb{E}[Y \mid X] - \hat{f}(X)\right\|^2_{L^2(\mathbb{P})} = \left\|Y - \hat{f}(X)\right\|^2_{L^2(\mathbb{P})} - \mathbb{E}[Y(Y-Z)]$$

↑ ↖ ↗

L^2-approximation error *can be estimated*

- ES$_\alpha$ is *L^2- Lipschitz-continuous:* $$\left|\mathrm{ES}_\alpha(f(X)) - \mathrm{ES}_\alpha(\hat{f}(X)\right| \leq \frac{1}{1-\alpha} \|f(X) - \hat{f}(X)\|_{L^2(\mathbb{P})}$$

- For any *candidate regression function* $\hat{f}\colon \mathbb{R}^d \to \mathbb{R}$ one has

$$\left\|Y - \hat{f}(X)\right\|^2_{L^2(\mathbb{P})} \geq \|Y - \mathbb{E}[Y \mid X]\|^2_{L^2(\mathbb{P})} = \mathbb{E}[Y(Y-Z)]$$

- By *Pythagoras* $\|Y - \mathbb{E}[Y \mid X]\|^2_{L^2(\mathbb{P})} + \left\|\mathbb{E}[Y \mid X] - \hat{f}(X)\right\|^2_{L^2(\mathbb{P})} = \left\|Y - \hat{f}(X)\right\|^2_{L^2(\mathbb{P})}$

- Therefore $\left\|\mathbb{E}[Y \mid X] - \hat{f}(X)\right\|^2_{L^2(\mathbb{P})} = \left\|Y - \hat{f}(X)\right\|^2_{L^2(\mathbb{P})} - \mathbb{E}[Y(Y-Z)]$

↑ ↖ ↗

L^2-approximation error *can be estimated*

- ES$_\alpha$ is L^2- *Lipschitz-continuous:* $\left|\mathrm{ES}_\alpha(f(X)) - \mathrm{ES}_\alpha(\hat{f}(X)\right| \leq \frac{1}{1-\alpha}\|f(X) - \hat{f}(X)\|_{L^2(\mathbb{P})}$

- Problem: for $\alpha = 0.99$, $\frac{1}{1-\alpha} = 100$

Guarantees for Numerical Approximations

- *Numerical approximation* $L = f(X) \approx \hat{L} = \hat{f}(X)$

Guarantees for Numerical Approximations

- *Numerical approximation* $L = f(X) \approx \hat{L} = \hat{f}(X)$
- Denote $\mathbb{P}_\alpha = \frac{1}{1-\alpha} 1_{\{L \geq \mathrm{Var}_\alpha(L)\}} \cdot \mathbb{P}, \qquad \hat{\mathbb{P}}_\alpha = \frac{1}{1-\alpha} 1_{\{\hat{L} \geq \mathrm{Var}_\alpha(\hat{L})\}} \cdot \mathbb{P}$

Guarantees for Numerical Approximations

- *Numerical approximation* $L = f(X) \approx \hat{L} = \hat{f}(X)$
- Denote $\mathbb{P}_\alpha = \frac{1}{1-\alpha} 1_{\{L \geq \mathrm{Var}_\alpha(L)\}} \cdot \mathbb{P}, \qquad \hat{\mathbb{P}}_\alpha = \frac{1}{1-\alpha} 1_{\{\hat{L} \geq \mathrm{Var}_\alpha(\hat{L})\}} \cdot \mathbb{P}$

Theorem $\left|\mathrm{ES}_\alpha(L) - \mathrm{ES}_\alpha(\hat{L})\right| \leq \left\|L - \hat{L}\right\|_{L^2(\mathbb{P}_\alpha)} \vee \left\|L - \hat{L}\right\|_{L^2(\hat{\mathbb{P}}_\alpha)}$

Guarantees for Numerical Approximations

- *Numerical approximation* $L = f(X) \approx \hat{L} = \hat{f}(X)$
- Denote $\mathbb{P}_\alpha = \frac{1}{1-\alpha} 1_{\{L \geq \mathrm{Var}_\alpha(L)\}} \cdot \mathbb{P}, \qquad \hat{\mathbb{P}}_\alpha = \frac{1}{1-\alpha} 1_{\{\hat{L} \geq \mathrm{Var}_\alpha(\hat{L})\}} \cdot \mathbb{P}$

Theorem $$\left|\mathrm{ES}_\alpha(L) - \mathrm{ES}_\alpha(\hat{L})\right| \leq \left\|L - \hat{L}\right\|_{L^2(\mathbb{P}_\alpha)} \vee \underset{\substack{\uparrow \\ \text{can be estimated}}}{\left\|L - \hat{L}\right\|_{L^2(\hat{\mathbb{P}}_\alpha)}}$$

Guarantees for Numerical Approximations

- *Numerical approximation* $L = f(X) \approx \hat{L} = \hat{f}(X)$

- Denote $\mathbb{P}_\alpha = \frac{1}{1-\alpha} 1_{\{L \geq \mathrm{Var}_\alpha(L)\}} \cdot \mathbb{P}, \qquad \hat{\mathbb{P}}_\alpha = \frac{1}{1-\alpha} 1_{\left\{\hat{L} \geq \mathrm{Var}_\alpha(\hat{L})\right\}} \cdot \mathbb{P}$

Theorem $\left|\mathrm{ES}_\alpha(L) - \mathrm{ES}_\alpha(\hat{L})\right| \leq \left\|L - \hat{L}\right\|_{L^2(\mathbb{P}_\alpha)} \vee \underset{\substack{\uparrow \\ \text{can be estimated}}}{\left\|L - \hat{L}\right\|_{L^2(\hat{\mathbb{P}}_\alpha)}}$

- *We assume* $\left\|L - \hat{L}\right\|_{L^2(\mathbb{P}_\alpha)} \approx \left\|L - \hat{L}\right\|_{L^2(\hat{\mathbb{P}}_\alpha)}$

Numerical Results

	$\widehat{\mathrm{ES}}_\alpha(\hat{L})$	error	relative error
Option Portfolio	104.6	$\pm$ 1.5	$\pm$ 1.4 %
Variable Annuity	142.0	$\pm$ 1.7	$\pm$ 1.2 %

Thank You!