# Rare event estimation with PDE-based models

## Autumn School *Uncertainty Quantification for High-Dimensional Problems*

**Elisabeth Ullmann**

Department of Mathematics
Technical University of Munich

07–11 October 2024

CWI Amsterdam, The Netherlands

# Resources

This mini-course is based on many sources, in particular, the material in the references and the textbooks

- J. Morio and M. Balesdent (2015). *Estimation of rare event probabilities in complex aerospace and other systems*. Cambridge, England: Woodhead Publishing
- A. B. Owen (2013). *Monte Carlo theory, methods and examples*. https://artowen.su.domains/mc/
- R. Y. Rubinstein and D. P. Kroese (2016). *Simulation and the Monte Carlo Method*. Wiley

# Plan

**Thursday 10 Oct**

1 Introduction, Monte Carlo
2 Importance Sampling (IS), Sequential IS
3 Exercise IS
4 Cross-entropy (CE) method

**Friday 11 Oct**

5 Exercise CE
6 SuS, FORM, Line Sampling
7 Exercise
8 Advanced Topics

# Part I: Basics of rare event estimation

1. Introduction

2. Standard Monte Carlo and its limitations

# Outline

# Informal definition[1]

**A rare event**

. . . is highly unlikely or occurs very infrequently, yet its occurence has high social and economic costs.

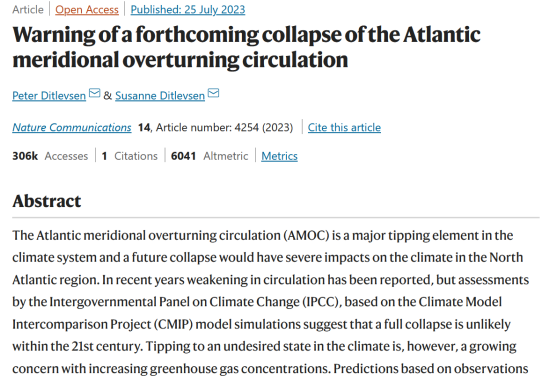
"[. . . ] on Monday 02 January 2006, the roof of the ice skating rink [in Bad Reichenhall in Southern Germany] collapsed due to construction defects following heavy snowfall. Fifteen people perished in the accident [. . . ]. Thirty-four were injured."

(Source: `https://en.wikipedia.org/wiki/Bad_Reichenhall_Ice_Rink_roof_collapse`)

[1] Figure by Chatter (left), CC BY-SA 3.0 https://creativecommons.org/licenses/by-sa/3.0, via Wikimedia Commons

# Examples of rare events

- structural failure of a bridge or building
- collapse of the Atlantic meridional overturning circulation (P. Ditlevsen & S. Ditlevsen, 2023)
- financial crash in the next year
- leak in an underground radioactive waste repository
- . . .



Article | Open Access | Published: 25 July 2023

**Warning of a forthcoming collapse of the Atlantic meridional overturning circulation**

Peter Ditlevsen & Susanne Ditlevsen

*Nature Communications* **14**, Article number: 4254 (2023) | Cite this article

**306k** Accesses | **1** Citations | **6041** Altmetric | Metrics

**Abstract**

The Atlantic meridional overturning circulation (AMOC) is a major tipping element in the climate system and a future collapse would have severe impacts on the climate in the North Atlantic region. In recent years weakening in circulation has been reported, but assessments by the Intergovernmental Panel on Climate Change (IPCC), based on the Climate Model Intercomparison Project (CMIP) model simulations suggest that a full collapse is unlikely within the 21st century. Tipping to an undesired state in the climate is, however, a growing concern with increasing greenhouse gas concentrations. Predictions based on observations

(Image source: Screenshot of webpage `https://www.nature.com/articles/s41467-023-39810-w`, accessed on 25 August 2023)

# Mathematical problem formulation

Estimate the probability of failure

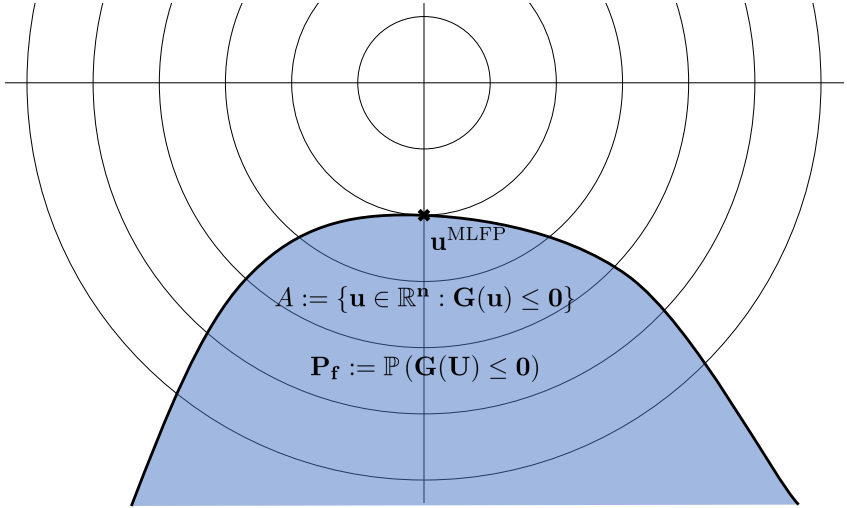
$$P_f := \mathbb{P}(G(\boldsymbol{U}) \leq 0) = \int_{\mathbb{R}^n} \underbrace{\mathcal{I}\left(G(\boldsymbol{u}) \leq 0\right)}_{\text{Indicates failure event}} \varphi_n(\boldsymbol{u}) \mathrm{d}\boldsymbol{u}$$

- $G : \mathbb{R}^n \to \mathbb{R}$ is a limit-state function (LSF)
- $\mathcal{I} : \mathbb{R} \to \{0, 1\}$ is the indicator function
- $\boldsymbol{U}$ is a $n$-dim. standard normal random vector, notation: $\boldsymbol{U} \sim N(0, I_n)$
- $\boldsymbol{u} \in \mathbb{R}^n$ is a parameter vector in the image space of $\boldsymbol{U}$
- $\varphi_n : \mathbb{R}^n \to \mathbb{R}$ is the $n$-variate standard normal density

# Terminology

- Limit-state function $G$
- Safe domain $\{\boldsymbol{u} \in \mathbb{R}^n : G(\boldsymbol{u}) > 0\}$
- Failure domain $\{\boldsymbol{u} \in \mathbb{R}^n : G(\boldsymbol{u}) \leq 0\}$
- Limit-state surface $\{\boldsymbol{u} \in \mathbb{R}^n : G(\boldsymbol{u}) = 0\}$

# Limit state function and failure domain



# Challenges

- The failure domain $A = \{\boldsymbol{u} \in \mathbb{R}^n : G(\boldsymbol{u}) \leq 0\}$ is not known in closed form.
- The failure domain $A$ is in a high-dimensional parameter space.
- The evaluation of $G$ is computationally expensive (requires PDE solve).

# Approximations of the probability of failure I

★ Sampling-based methods

- Sequential Importance Sampling, e.g. (Papaioannou, Papadimitriou, et al., 2016)
- Subset simulation (Au & Beck, 2001); Generalized splitting (Botev & Kroese, 2012)
- Moving particles method (Guyader et al., 2011)
- Reversible shaking transformations (Gobet & G. Liu, 2015)
- Multilevel/Hierarchical sampling (Ullmann & Papaioannou, 2015; Elfverson, Hellman, et al., 2016; Peherstorfer et al., 2018; Proppe, 2020; Wagner, Latz, et al., 2020; Haji-Ali et al., 2022; Elfverson, Scheichl, et al., 2022)
- ...

# Approximations of the probability of failure II

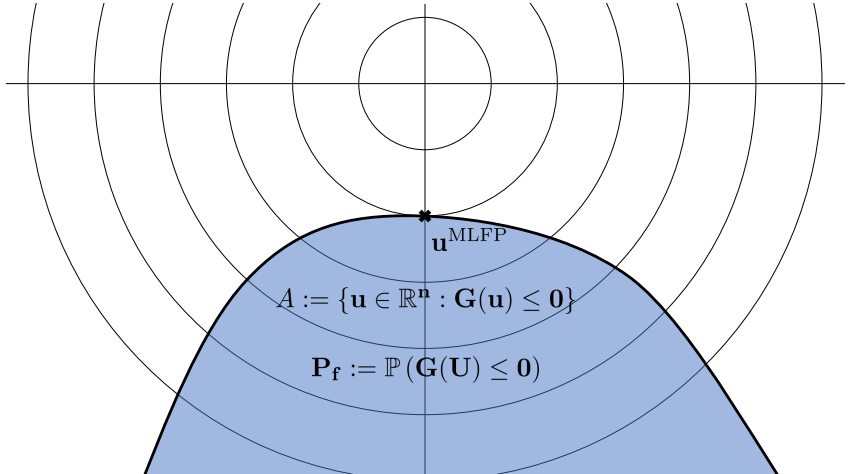
- ★ Semi-analytical methods, reliability-based methods
  - First order reliability method, FORM (Hasofer & Lind, 1974)
  - Second order reliability method, SORM (Breitung, 1984)
  - Line Sampling (Koutsourelakis et al., 2004)
  - ...
- ★ Large deviation theory-based methods, e.g. (Dembo & Zeitouni, 2010; Tong, Vanden-Eijnden, et al., 2021; Tong & Stadler, 2022)
- ★ ...

# Outline

# Standard Monte Carlo estimate

. . . throwing darts (according to the standard normal density) and counting the hits . . .

$$\widehat{P}_f^{\mathrm{MC}} := (\#\text{darts in blue set})/(\text{total number of darts})$$



# Standard Monte Carlo I

Let $H\colon \mathbb{R}^d \to \mathbb{R}$ denote a performance function of the system of interest. Define the real-valued random variable $Y := H(\boldsymbol{U})$ and assume $\mathbb{E}\left[|Y|\right] < +\infty$.

**Goal:** Estimate $Q := \mathbb{E}\left[Y\right]$

**Definition 2.1.** *Let $N \in \mathbb{N}$. Consider a sequence $Y^{(1)}, \ldots, Y^{(N)}$ of independent, identically distributed copies of $Y$. The standard Monte Carlo estimator (a.k.a. sample mean) for $Q = \mathbb{E}\left[Y\right]$ is defined as*

$$E^{MC}[Q] := \frac{1}{N} \sum_{i=1}^{N} Y^{(i)}. \tag{2.1}$$

# Standard Monte Carlo II

Let $\widehat{Q}$ denote a statistical estimator for the deterministic quantity $Q$.

**Definition 2.2.** *The mean-square error of $\widehat{Q}$ is defined as*

$$\mathrm{mse}(\widehat{Q}) := \|\widehat{Q} - Q\|_{L^2}^2 = \mathbb{E}\left[(\widehat{Q} - Q)^2\right]. \tag{2.2}$$

*The root-mean-square error is defined as*

$$\mathrm{rmse}(\widehat{Q}) := \|\widehat{Q} - Q\|_{L^2} = \sqrt{\mathbb{E}\left[(\widehat{Q} - Q)^2\right]}. \tag{2.3}$$

*The coefficient of variation (c.o.v.) (a.k.a. relative error) is defined as*

$$\mathrm{c.o.v.}(\widehat{Q}) := \frac{\sqrt{\mathrm{var}(\widehat{Q})}}{\mathbb{E}\left[\widehat{Q}\right]}. \tag{2.4}$$

# Standard Monte Carlo III

**Corollary 2.3.** *The standard Monte Carlo estimator has the following properties:*

(i) $E^{MC}[Q] \to Q$ *almost surely as* $N \to +\infty$.

(ii) *If* $\mathbb{E}\left[Y^2\right] < +\infty$, *then the distribution of the random variable*

$$S_N := \sqrt{N}(E^{MC}[Q] - Q)$$

*converges to a standard normal distribution with mean zero and variance* $\operatorname{var} Y$ *as* $N \to +\infty$.

(iii) $E^{MC}[Q]$ *is an unbiased estimator for* $Q$.

(iv) *If* $\operatorname{var} Y < +\infty$, *then* $E^{MC}[Q] \to Q$ *in mean-square as* $N \to +\infty$.

# Standard Monte Carlo IV

*Proof.* $(i)$ by the Strong Law of Large Numbers

$(ii)$ by the Central Limit Theorem

$(iii)$ $\mathbb{E}\left[E^{MC}[Q]\right] = \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N} Y^{(i)}\right] = \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left[Y^{(i)}\right] = \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left[Y\right] = \mathbb{E}\left[Y\right] = Q.$

$(iv)$

$$\begin{aligned}
\operatorname{rmse}(E^{MC}[Q]) &= \mathbb{E}\left[(E^{MC}[Q]-Q)^2\right] \\
&= \mathbb{E}\left[(E^{MC}[Q]-\mathbb{E}[E^{MC}[Q]])^2\right] \quad \text{by } (iii) \\
&= \operatorname{var} E^{MC}[Q] = \frac{1}{N^2}\operatorname{var}\left(\sum_{i=1}^{N} Y^{(i)}\right) = \frac{N\operatorname{var} Y}{N^2} = \frac{\operatorname{var} Y}{N} \to 0 \quad \text{as } N \to +\infty.
\end{aligned}$$

□

# Standard Monte Carlo for failure probabilities I

Let $\mathbb{1}_{\{G(U)\leq 0\}}$ denote the random variable indicating the failure event. Note that $P_f = \mathbb{E}\left[\mathbb{1}_{\{G(U)\leq 0\}}\right]$. The Monte Carlo estimator for $P_f$ is

$$E^{MC}[P_f] := \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{\{G(U^{(i)})\leq 0\}}. \tag{2.5}$$

**Properties of $E^{MC}[P_f]$.**

$$\operatorname{var} \mathbb{1}_{\{G(U)\leq 0\}} = \mathbb{E}\left[\mathbb{1}_{\{G(U)\leq 0\}}\right] - \mathbb{E}\left[\mathbb{1}_{\{G(U)\leq 0\}}\right]^2 = P_f - P_f^2 = P_f(1-P_f)$$

$$\text{c. o. v.}(E^{MC}[P_f]) = \frac{\sqrt{\operatorname{var}(E^{MC}[P_f])}}{\mathbb{E}\left[E^{MC}[P_f]\right]} = \frac{\sqrt{\operatorname{var}(\mathbb{1}_{\{G\leq 0\}})N^{-1}}}{\mathbb{E}\left[\mathbb{1}_{\{G\leq 0\}}\right]} = \sqrt{\frac{1-P_f}{P_f N}}.$$

# Standard Monte Carlo for failure probabilities II

## Conclusion

We require

$$N = \frac{1 - P_f}{P_f \times \tau^2}$$

samples to estimate $P_f$ by standard Monte Carlo with relative error (c.o.v.) $\tau > 0$.

For $P_f \ll 1$ this means $N \sim P_f^{-1} \times \tau^{-2}$.

# Part II: Variance reduction

3 Importance Sampling

4 Sequential Importance Sampling

5 Cross-entropy method

6 Subset simulation

# Outline

# Basic Idea I

- $f\colon \mathbb{R}^d \to \mathbb{R}$ probability density function (pdf) of the random input vector $\boldsymbol{U}$ with support $\mathcal{D}_f \subset \mathbb{R}^n$
- importance sampling density $g\colon \mathbb{R}^d \to \mathbb{R}$ of the random vector $\boldsymbol{V}$ with support $\mathcal{D}_g \subset \mathbb{R}^n$, such that $Hf$ is dominated by $g$: If $g(\boldsymbol{u}) = 0$ then $H(\boldsymbol{u})f(\boldsymbol{u}) = 0$.
- Likelihood ratio $W(\boldsymbol{u}) := f(\boldsymbol{u})/g(\boldsymbol{u})$, $\boldsymbol{u} \in \mathcal{D}_g$
- **Goal:** Estimate

$$Q = \mathbb{E}[H(\boldsymbol{U})] = \int_{\mathcal{D}_f} H(\boldsymbol{u}) f(\boldsymbol{u}) \mathrm{d}\boldsymbol{u}$$

- Change of measure:

$$Q = \int_{\mathcal{D}_g} H(\boldsymbol{u}) \frac{f(\boldsymbol{u})}{g(\boldsymbol{u})} g(\boldsymbol{u}) \mathrm{d}\boldsymbol{u} = \mathbb{E}\left[ H(\boldsymbol{V}) \frac{f(\boldsymbol{V})}{g(\boldsymbol{V})} \right] \tag{3.1}$$

# Basic Idea II

**Definition 3.1.** *Let $\boldsymbol{U}$ be a random vector with pdf $f$. Let $\boldsymbol{V}$ denote a random vector with pdf $g$. Let $N \in \mathbb{N}$. Consider a sequence $\boldsymbol{V}^{(1)}, \dots, \boldsymbol{V}^{(N)}$ of independent, identically distributed copies of $\boldsymbol{V}$. The importance sampling (IS) estimator for $Q = \mathbb{E}[H(\boldsymbol{U})]$ is defined as*

$$E_g^{IS}[Q] := \frac{1}{N} \sum_{i=1}^{N} H(\boldsymbol{V}^{(i)}) \frac{f(\boldsymbol{V}^{(i)})}{g(\boldsymbol{V}^{(i)})}. \tag{3.2}$$

**Corollary 3.2.** *Assume that $g$ dominates $Hf$, that is, for all $\boldsymbol{u} \notin \mathcal{D}_g$ we have $H(\boldsymbol{u})f(\boldsymbol{u}) = 0$. Then the importance sampling estimator $E_g^{IS}[Q]$ is an unbiased estimator for $Q = \mathbb{E}[H(\boldsymbol{U})]$.*

# Basic Idea III

*Proof.*

$$\begin{aligned}\mathbb{E}\left[H(\boldsymbol{V}^{(i)})\frac{f(\boldsymbol{V}^{(i)})}{g(\boldsymbol{V}^{(i)})}\right] &= \int_{\mathcal{D}_g}\frac{H(\boldsymbol{u})f(\boldsymbol{u})}{g(\boldsymbol{u})}g(\boldsymbol{u})\mathrm{d}\boldsymbol{u} = \int_{\mathcal{D}_g}H(\boldsymbol{u})f(\boldsymbol{u})\mathrm{d}\boldsymbol{u}\\ &= \int_{\mathcal{D}_f}H(\boldsymbol{u})f(\boldsymbol{u})\mathrm{d}\boldsymbol{u} + \int_{\mathcal{D}_f^c\cap\mathcal{D}_g}H(\boldsymbol{u})\underbrace{f(\boldsymbol{u})}_{=0}\mathrm{d}\boldsymbol{u} - \int_{\mathcal{D}_f\cap\mathcal{D}_g^c}\underbrace{H(\boldsymbol{u})f(\boldsymbol{u})}_{=0}\mathrm{d}\boldsymbol{u}\\ &= \int_{\mathcal{D}_f}H(\boldsymbol{u})f(\boldsymbol{u})\mathrm{d}\boldsymbol{u} = \mathbb{E}[H(\boldsymbol{U})]\end{aligned}$$

$$\mathbb{E}[E_g^{IS}[Q]] = \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left[H(\boldsymbol{V}^{(i)})\frac{f(\boldsymbol{V}^{(i)})}{g(\boldsymbol{V}^{(i)})}\right] = \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[H(\boldsymbol{U})] = \mathbb{E}[H(\boldsymbol{U})] = Q$$

□

# Choice of importance sampling density I

Required:

- Finite variance $\mathrm{var}(E_g^{IS}[Q])$, equivalently

$$\mathbb{E}\left[H^2(\boldsymbol{V})\frac{f^2(\boldsymbol{V})}{g^2(\boldsymbol{V})}\right] = \mathbb{E}\left[H^2(\boldsymbol{U})\frac{f(\boldsymbol{U})}{g(\boldsymbol{U})}\right] < +\infty.$$

- Variance reduction, $\mathrm{var}(E_g^{IS}[Q]) \leq \mathrm{var}(E^{MC}[Q])$, equivalently

$$\mathrm{var}(H(\boldsymbol{V})f(\boldsymbol{V})/g(\boldsymbol{V})) \leq \mathrm{var}(H(\boldsymbol{U})).$$

**Theorem 3.3.** *The minimum of* $\mathrm{var}(H(\boldsymbol{V})f(\boldsymbol{V})/g(\boldsymbol{V}))$ *is equal to*

$$V_0 := \left(\int_{\mathcal{D}_{|H|f}} |H(\boldsymbol{u})|f(\boldsymbol{u})\mathrm{d}\boldsymbol{u}\right)^2 - \left(\int_{\mathcal{D}_{|H|f}} H(\boldsymbol{u})f(\boldsymbol{u})\mathrm{d}\boldsymbol{u}\right)^2 \tag{3.3}$$

# Choice of importance sampling density II

*and occurs for the random vector $\boldsymbol{V}_{opt}$ with the pdf*

$$g_{opt}(\boldsymbol{u}) := \frac{|H(\boldsymbol{u})|f(\boldsymbol{u})}{\int_{\mathcal{D}_{|H|f}} |H(\boldsymbol{u})|f(\boldsymbol{u})\mathrm{d}\boldsymbol{u}}. \tag{3.4}$$

*Proof.* (3.3) follows by using $g_{opt}$ in (3.4) to calculate $\mathrm{var}(H(\boldsymbol{V}_{opt})f(\boldsymbol{V}_{opt})/g(\boldsymbol{V}_{opt}))$. Indeed, with $Z := \int_{\mathcal{D}_{|H|f}} |H(\boldsymbol{u})|f(\boldsymbol{u})\mathrm{d}\boldsymbol{u}$ we obtain

$$\begin{aligned}
\mathrm{var}(H(\boldsymbol{V}_{opt})f(\boldsymbol{V}_{opt})/g_{opt}(\boldsymbol{V}_{opt})) &= \int_{\mathcal{D}_{|H|f}} \frac{H^2(\boldsymbol{u})f^2(\boldsymbol{u})}{g_{opt}(\boldsymbol{u})}\mathrm{d}\boldsymbol{u} - \left(\int_{\mathcal{D}_{|H|f}} H(\boldsymbol{u})f(\boldsymbol{u})\mathrm{d}\boldsymbol{u}\right)^2 \\
&= \int_{\mathcal{D}_{|H|f}} Z|H(\boldsymbol{u})|f(\boldsymbol{u})\mathrm{d}\boldsymbol{u} - \left(\int_{\mathcal{D}_{|H|f}} H(\boldsymbol{u})f(\boldsymbol{u})\mathrm{d}\boldsymbol{u}\right)^2 \\
&= \left(\int_{\mathcal{D}_{|H|f}} |H(\boldsymbol{u})|f(\boldsymbol{u})\mathrm{d}\boldsymbol{u}\right)^2 - \left(\int_{\mathcal{D}_{|H|f}} H(\boldsymbol{u})f(\boldsymbol{u})\mathrm{d}\boldsymbol{u}\right)^2.
\end{aligned}$$

# Choice of importance sampling density III

To show $V_0 \leq \mathrm{var}(H(\boldsymbol{V})f(\boldsymbol{V})/g(\boldsymbol{V}))$ for any other random vector $\boldsymbol{V}$ with pdf $g$ on $\mathbb{R}^n$, it is sufficient to prove that

$$\left(\int_{\mathcal{D}_g} |H(\boldsymbol{u})|f(\boldsymbol{u})\mathrm{d}\boldsymbol{u}\right)^2 \leq \int_{\mathcal{D}_g} \frac{H^2(\boldsymbol{u})f^2(\boldsymbol{u})}{g(\boldsymbol{u})}\mathrm{d}\boldsymbol{u},$$

which follows from the Cauchy–Schwarz inequality. Indeed,

$$\begin{aligned}\left(\int_{\mathcal{D}_g} |H(\boldsymbol{u})|f(\boldsymbol{u})\mathrm{d}\boldsymbol{u}\right)^2 &= \left(\int_{\mathcal{D}_g} \frac{|H(\boldsymbol{u})|f(\boldsymbol{u})}{g(\boldsymbol{u})^{1/2}} g(\boldsymbol{u})^{1/2}\mathrm{d}\boldsymbol{u}\right)^2 \\ &\leq \int_{\mathcal{D}_g} \frac{H^2(\boldsymbol{u})f^2(\boldsymbol{u})}{g(\boldsymbol{u})}\mathrm{d}\boldsymbol{u} \int_{\mathcal{D}_g} g(u)\mathrm{d}\boldsymbol{u} = \int_{\mathcal{D}_g} \frac{H^2(\boldsymbol{u})f^2(\boldsymbol{u})}{g(\boldsymbol{u})}\mathrm{d}\boldsymbol{u}.\end{aligned}$$

□

# Choice of importance sampling density IV

**Corollary 3.4.** *If $H(\boldsymbol{u}) \geq 0$, then the optimal importance sampling density is given by*

$$g_{opt}(\boldsymbol{u}) = \frac{H(\boldsymbol{u}) f(\boldsymbol{u})}{Q}. \tag{3.5}$$

*Moreover,* $\mathrm{var}(E^{IS}_{g_{opt}}[Q]) = \mathrm{var}(H(\boldsymbol{V}_{opt}) f(\boldsymbol{V}_{opt}) / g_{opt}(\boldsymbol{V}_{opt})) = 0.$

**Corollary 3.5.** *For $H(\boldsymbol{u}) = \mathcal{I}(G(\boldsymbol{u}) \leq 0)$ the optimal importance sampling density is given by*

$$g_{opt}(\boldsymbol{u}) = \frac{\mathcal{I}(G(\boldsymbol{u}) \leq 0)}{P_f} f(\boldsymbol{u}). \tag{3.6}$$

*The corresponding IS estimator has variance equal to zero.*

# Self-normalized importance sampling I

Suppose $f(\boldsymbol{u}) = c_f f_u(\boldsymbol{u})$ and $g(\boldsymbol{u}) = c_g g_u(\boldsymbol{u})$, where $c_f, c_g > 0$ are unknown constants. Suppose that we can evalute $f_u$ and $g_u$. Hence we can also evaluate the ratio

$$W_u(\boldsymbol{u}) := \frac{f_u(\boldsymbol{u})}{g_u(\boldsymbol{u})} = \frac{c_g}{c_f} \frac{f(\boldsymbol{u})}{g(\boldsymbol{u})}.$$

**Definition 3.6.** *Let $\boldsymbol{U}$ denote a random vector with pdf $f$. Let $\boldsymbol{V}$ denote a random vector with pdf $g$. Let $N \in \mathbb{N}$. Consider a sequence $\boldsymbol{V}^{(1)}, \ldots, \boldsymbol{V}^{(N)}$ of independent, identically distributed copies of $\boldsymbol{V}$. The self-normalized importance sampling estimator for $Q = \mathbb{E}[H(\boldsymbol{U})]$ is defined as*

$$E^{IS}_{sn,g}[Q] := \frac{\frac{1}{N}\sum_{i=1}^{N} H(\boldsymbol{V}^{(i)}) W_u(\boldsymbol{V}^{(i)})}{\frac{1}{N}\sum_{i=1}^{N} W_u(\boldsymbol{V}^{(i)})}. \tag{3.7}$$

# Self-normalized importance sampling II

**Theorem 3.7.** *Assume that $g$ dominates $f$, that is, $\mathcal{D}_f \subseteq \mathcal{D}_g$. Then $E^{IS}_{sn,g}[Q] \to Q$ almost surely as $N \to +\infty$.*

*Proof.* Write the self-normalized IS estimator as

$$E^{IS}_{sn,g}[Q] = \frac{\frac{1}{N}\sum_{i=1}^{N} H(\boldsymbol{V}^{(i)})W(\boldsymbol{V}^{(i)})}{\frac{1}{N}\sum_{i=1}^{N} W(\boldsymbol{V}^{(i)})}. \tag{3.8}$$

The numerator in (3.8) is the sample average of the i.i.d. random variables $H(\boldsymbol{V}^{(i)})f(\boldsymbol{V}^{(i)})/g(\boldsymbol{V}^{(i)})$ with expectation

$$\int_{\mathcal{D}_g} H(\boldsymbol{u})\frac{f(\boldsymbol{u})}{g(\boldsymbol{u})}g(\boldsymbol{u})\mathrm{d}\boldsymbol{u} = \int_{\mathcal{D}_g} H(\boldsymbol{u})f(\boldsymbol{u})\mathrm{d}\boldsymbol{u} = \int_{\mathcal{D}_f} H(\boldsymbol{u})f(\boldsymbol{u})\mathrm{d}\boldsymbol{u} = Q.$$

# Self-normalized importance sampling III

Hence it converges to $Q$ almost surely as $N \to +\infty$ according to the Strong Law of Large Numbers (SLLN). The denominator in (3.8) converges to $1$ almost surely as $N \to +\infty$, again by the SLLN. □

**Properties.**

- $E^{IS}_{sn,g}[Q]$ is a ratio estimator.
- $E^{IS}_{sn,g}[Q]$ has bias is of order $\mathcal{O}(1/N)$ as $N \to +\infty$ (Kong, 1992).
- The variance of $E^{IS}_{sn,g}[Q]$ can be approximated as

$$\mathrm{var}(E^{IS}_{sn,g}[Q]) \approx \mathrm{var}(E^{MC}[Q]) \times (1 + \mathrm{var}(W(\boldsymbol{V}))), \tag{3.9}$$

where $\boldsymbol{V} \sim g$ and $W(\boldsymbol{V}) = f(\boldsymbol{V})/g(\boldsymbol{V})$ is the likelihood ratio (Kong, 1992).
Hence $\mathrm{var}(E^{IS}_{sn,g}[Q]) = \mathcal{O}(1/N)$ as $N \to +\infty$.

# Effective sample size

- Standard Monte Carlo uses $N$ equally weighted samples ($w_i = 1/N$, $i = 1, \dots, N$).
- Importance sampling uses $N$ unequally weighted samples $w_i$, which are realizations of the random variables

$$W_i := W(\boldsymbol{V}^{(i)}) = f(\boldsymbol{V}^{(i)})/g(\boldsymbol{V}^{(i)}) \geq 0, \quad i = 1, \dots, N. \tag{3.10}$$

- If the variance of $W_i$ is large, then some samples carry (a lot) more weight than others.
  $\longrightarrow$ Detect this situation using the ESS!

**Definition 3.8.** *The effective sample size (ESS) is defined as*

$$ESS := N \times \frac{\mathrm{var}(E^{MC}[Q])}{\mathrm{var}(E^{IS}_{sn,g}[Q])}. \tag{3.11}$$

# Heuristic approximations of ESS I

- **Option 1:** Assume that $H(\boldsymbol{U}^{(i)})$ are independent random variables with common mean and common variance $\sigma^2 > 0$, and that $\boldsymbol{w} = (w_1, \dots, w_N)^\top \in [0, +\infty)^N$ is a given realization of the random vector $\boldsymbol{W} := (W_1, \dots, W_N)^\top$. Let $Q := \mathbb{E}[H(\boldsymbol{U}^{(1)})]$. Then it holds $\mathrm{var}(E^{MC}[Q]) = \sigma^2/N$ and

$$\mathrm{var}\left(\sum_{i=1}^{N} w_i H(\boldsymbol{U}^{(i)})/\overline{w}\right) = \sigma^2 \sum_{i=1}^{N} w_i^2/\overline{w}^2,$$

giving

$$N_{\text{eff}} := \frac{(\sum_{i=1}^{N} w_i)^2}{\sum_{i=1}^{N} w_i^2}, \quad \text{where} \quad \frac{\sigma^2}{N_{\text{eff}}} = \mathrm{var}\left(\sum_{i=1}^{N} w_i H(\boldsymbol{U}^{(i)})/\overline{w}\right). \tag{3.12}$$

# Heuristic approximations of ESS II

- **Option 2:** Using $\mathrm{var}(E^{IS}_{sn,g}[Q]) \approx \mathrm{var}(E^{MC}[Q])(1+\mathrm{var}(W(\boldsymbol{V})))$ in (3.9) we obtain

$$ESS = N \times \frac{\mathrm{var}(E^{MC}[Q])}{\mathrm{var}(E^{IS}_{sn,g}[Q])} \approx \frac{N}{1+\mathrm{var}(W(\boldsymbol{V}))},$$

where $\boldsymbol{V} \sim g$ is a random vector with pdf $g$. In practise the densities $f, g$ are often not normalized. Thus we replace $\mathrm{var}(W(\boldsymbol{V}))$ by the coefficient of variation squared, $\mathrm{var}(W(\boldsymbol{V}))/\mathbb{E}(W(\boldsymbol{V}))^2$. This gives

$$N_{\text{eff,cv}} := \frac{N}{1+\widehat{\text{c. o. v.}}(\boldsymbol{w})^2}, \tag{3.13}$$

$$\widehat{\text{c. o. v.}}(\boldsymbol{w}) = \frac{\widehat{\mathrm{var}}(\boldsymbol{w})^{1/2}}{\widehat{\mathbb{E}}[\boldsymbol{w}]}, \quad \widehat{\mathbb{E}}[\boldsymbol{w}] = \frac{1}{N}\sum_{i=1}^{N} w_i, \quad \widehat{\mathrm{var}}(\boldsymbol{w}) = \frac{1}{N-1}\sum_{i=1}^{N}(w_i - \widehat{\mathbb{E}}[\boldsymbol{w}])^2. \tag{3.14}$$

# Adaptive Importance Sampling

- In rare event estimation

$$g_{opt}(\boldsymbol{u}) = \mathcal{I}(G(\boldsymbol{u}) \leq 0) f(\boldsymbol{u}) / P_f \tag{3.15}$$

  is not accessible in practice since we don't know $P_f$ and the failure domain $\{\boldsymbol{u} \in \mathbb{R}^n \colon G(\boldsymbol{u}) \leq 0\}$.

- **Key idea:** Approximate $g_{opt}$ through a sequence of auxiliary IS densities in a parametric family of densities

$$\mathcal{G} = \{g(\boldsymbol{u}; \boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}.$$

- Examples:
  - Sequential Importance Sampling (Papaioannou, Papadimitriou, et al., 2016)
  - Cross-entropy method (Boer et al., 2005)

# Outline

# Sequential Importance Sampling
## Basic idea

- Assume that all distributions in this section have a pdf with respect to the Lebesgue measure on $\mathbb{R}^n$. Write $\boldsymbol{U}_j \sim g_j$ meaning that $\boldsymbol{U}_j$ has the pdf $g_j$.
- Assume

$$g_j(\boldsymbol{u}) = \frac{\eta_j(\boldsymbol{u})}{P_j},$$

  where $\eta_j(\boldsymbol{u})$ is known analytically for all $\boldsymbol{u} \in \mathbb{R}^n$, and $P_j$ is not known. Assume $P_0 = 1$ and thus $g_0 = \eta_0$.
- **Goal:** Obtain samples of $\boldsymbol{U}_M \sim g_M$ and estimate the normalizing constant $P_M$.
- **Idea:** Sample sequentially from the densities; estimate the ratio $P_j/P_{j-1}$ by Importance Sampling using $g_{j-1}$ as IS density. Estimate $P_M$ using the telescoping product

$$P_M = \frac{P_M}{P_{M-1}} \times \frac{P_{M-1}}{P_{M-2}} \times \cdots \times \frac{P_2}{P_1} \times \frac{P_1}{P_0}. \tag{4.1}$$

# Sequential Importance Sampling I

## A single step of SIS

Assumption: From step $j-1$ we have $N$ equally weighted samples of $\boldsymbol{U}_{j-1}$, notation $\boldsymbol{u}_{j-1}^{(i)}$, $i = 1, \ldots, N$. Define the weight function

$$W_j(\boldsymbol{u}) := \frac{\eta_j(\boldsymbol{u})}{\eta_{j-1}(\boldsymbol{u})}. \tag{4.2}$$

Write the normalizing constant in step $j$ as

$$P_j = \int_{\mathbb{R}^n} \eta_j(\boldsymbol{u}) \mathrm{d}\boldsymbol{u} = P_{j-1} \int_{\mathbb{R}^n} W_j(\boldsymbol{u}) \frac{\eta_{j-1}(\boldsymbol{u})}{P_{j-1}} \mathrm{d}\boldsymbol{u} = P_{j-1} \mathbb{E}\left[W_j(\boldsymbol{U}_{j-1})\right].$$

Hence

$$S_j := \frac{P_j}{P_{j-1}} = \mathbb{E}\left[W_j(\boldsymbol{U}_{j-1})\right], \quad \widehat{S}_j = E^{IS}_{g_{j-1}}[S_j] = \frac{1}{N} \sum_{i=1}^{N} W_j(\boldsymbol{u}_{j-1}^{(i)}).$$

# Sequential Importance Sampling II

## A single step of SIS

By (4.1) we obtain an estimate for the normalizing constant $P_j$ as

$$\widehat{P}_j = \prod_{i=1}^{j} \widehat{S}_j. \tag{4.3}$$

Samples of $\boldsymbol{U}_j$ are obtained by a resample-move scheme: We select randomly with replacement $N_c < N$ samples from $\{\boldsymbol{U}_{j-1}^{(i)}, i = 1, \ldots, N\}$. The probability of selecting $\boldsymbol{u}_{j-1}^{(i)}$ is $W_j(\boldsymbol{u}_{j-1}^{(i)})$. Each of the selected $N_c$ samples is used as seed for a Markov chain of length $N/N_c$ with stationary density $g_j(\boldsymbol{u})$. See Algorithm 2.

# Sequential Importance Sampling I
## Choice of auxiliary densities

(Papaioannou, Papadimitriou, et al., 2016)

- Assume that $g_0$ is the $n$-variate standard normal pdf.
- **Idea:** Approximate the indicator function as

$$\mathcal{I}(G(\boldsymbol{u}) \leq 0) = \lim_{\sigma \to 0+} \Phi\left(-\frac{G(\boldsymbol{u})}{\sigma}\right), \quad G(\boldsymbol{u}) \neq 0,$$

  where $\Phi \colon \mathbb{R} \to \mathbb{R}$ is the univariate standard normal cdf.
- By choosing $\sigma_M > 0$ sufficiently small we obtain the approximation

$$g_M(\boldsymbol{u}) = \frac{1}{P_M}\Phi\left(-\frac{G(\boldsymbol{u})}{\sigma_M}\right) f(\boldsymbol{u}) \approx \frac{1}{P_f}\mathcal{I}(G(\boldsymbol{u}) \leq 0) f(\boldsymbol{u}) = g_{opt}(\boldsymbol{u}).$$

# Sequential Importance Sampling II

## Choice of auxiliary densities

- Summary: Define

$$\eta_j(\boldsymbol{u}) = \Phi(-G(\boldsymbol{u})/\sigma_j) f(\boldsymbol{u}), \quad g_j(\boldsymbol{u}) = \eta_j(\boldsymbol{u})/P_j, \quad j = 0, \dots, M,$$

  where formally $\sigma_0 = +\infty$.

- How to determine the temperature $\sigma_j$?
  Let $\boldsymbol{w}_j := (W_j(\boldsymbol{u}_{j-1}^{(1)}), \dots, W_j(\boldsymbol{u}_{j-1}^{(N)}))^\top$. Note that $\boldsymbol{w}_j = \boldsymbol{w}_j(\sigma)$, where $\sigma$ is the yet unknown scaling parameter in the density $g_j$. Now, given $\boldsymbol{w}_j$ and $\delta_{\text{target}} > 0$, solve the optimization problem

$$\sigma_j = \operatorname*{argmin}_{\sigma \in (0, \sigma_{j-1})} |\widehat{\text{c. o. v.}}(\boldsymbol{w}_j(\sigma)) - \delta_{\text{target}}|.$$

# Sequential Importance Sampling
## Coefficient of variation of weights

Note: The sample coefficient of variation of the weights squared $\widehat{\text{c. o. v.}}(\boldsymbol{w}_j)^2$ is an approximation of the Pearson $\chi^2$-divergence of the densities $g_j$ and $g_{j-1}$.

$$\begin{aligned}
\frac{\text{var}(W_j(\boldsymbol{U}_{j-1}))}{\mathbb{E}[W_j(\boldsymbol{U}_{j-1})]^2} &= \frac{\mathbb{E}[W_j(\boldsymbol{U}_{j-1})^2] - \mathbb{E}[W_j(\boldsymbol{U}_{j-1})]^2}{\mathbb{E}[W_j(\boldsymbol{U}_{j-1})]^2} \\
&= \frac{P_j^2/P_{j-1}^2 \mathbb{E}[g_j(\boldsymbol{U}_{j-1})/g_{j-1}(\boldsymbol{U}_{j-1})] - (P_j/P_{j-1})^2}{(P_j/P_{j-1})^2} \\
&= \mathbb{E}[g_j(\boldsymbol{U}_j)/g_{j-1}(\boldsymbol{U}_j)] - 1 \\
D_{\chi^2}(g_j, g_{j-1}) &= \int_{\mathbb{R}^n} \frac{(g_j(\boldsymbol{u}) - g_{j-1}(\boldsymbol{u}))^2}{g_{j-1}(\boldsymbol{u})} \mathrm{d}\boldsymbol{u} \\
&= \mathbb{E}[g_j(\boldsymbol{U}_j)/g_{j-1}(\boldsymbol{U}_j)] - 2\mathbb{E}[1] + \mathbb{E}[1] = \mathbb{E}[g_j(\boldsymbol{U}_j)/g_{j-1}(\boldsymbol{U}_j)] - 1.
\end{aligned}$$

# Sequential Importance Sampling
## Stopping criterion

- Stop the SIS iteration if the (unnormalized) density $\eta_j(\boldsymbol{u})$ is "close" to the (unnormalized) optimal IS density $\mathcal{I}(G(\boldsymbol{u}) \leq 0) f(\boldsymbol{u})$.
- Define $W_j^{(opt)}(\boldsymbol{u}) := \mathcal{I}(G(\boldsymbol{u}) \leq 0) f(\boldsymbol{u}) / \eta_j(\boldsymbol{u})$. Given (approximate) samples $\boldsymbol{u}_j^{(i)}, i = 1, \ldots, N$ of $\boldsymbol{U}_j \sim g_j$ estimate the coefficient of variation of $W_j^{(opt)}(\boldsymbol{U}_j)$ as

$$\widehat{\text{c. o. v.}}(\boldsymbol{w}_{opt}) = \frac{\widehat{\text{var}}(\boldsymbol{w}_{opt})^{1/2}}{\widehat{\mathbb{E}}[\boldsymbol{w}_{opt}]},$$

  where $\boldsymbol{w}_{opt} = (W_j^{(opt)}(\boldsymbol{u}_j^{(1)}), \ldots, W_j^{(opt)}(\boldsymbol{u}_j^{(N)}))^\top$.
- If $\widehat{\text{c. o. v.}}(\boldsymbol{w}_{opt}) < \delta_{\text{target}}$, stop the SIS iteration.

# Sequential Importance Sampling

## Estimate of $P_f$

Assume that we have stopped the SIS iteration at $j = M$. The final estimate for $P_f$ is based on the identity

$$P_f = P_M \times \frac{P_f}{P_M} = P_M \times \mathbb{E}[W_M^{(opt)}(\boldsymbol{U}_M)].$$

The ratio $P_f/P_M$ is estimated using the weights $\boldsymbol{w}_{opt}$ from evaluating the stopping criterion. We arrive at

$$\widehat{P}_f = \widehat{P}_M \times \frac{1}{N} \sum_{i=1}^{N} W_M^{(opt)}(\boldsymbol{u}_M^{(i)}).$$

Algorithm 1 summarizes the steps of SIS for rare event estimation.

# Sequential Importance Sampling to estimate $P_f$ I

**Algorithm 1** (Papaioannou, Papadimitriou, et al., 2016)

1: Input: number of samples $N$, target c.o.v. $\delta_{\text{target}}$
2: **for** $k = 1 \colon N$ **do**
3: Sample $\boldsymbol{U}_0 \sim g_0$, notation $\boldsymbol{u}_0^{(k)}$
4: **end for**
5: Set $\sigma_0 = +\infty$, $\widehat{P}_0 = 1$, $j = 1$
6: Determine the density parameter: $\sigma_j = \underset{\sigma \in (0, \sigma_{j-1})}{\operatorname{argmin}} \left| \widehat{\text{c. o. v.}}(\boldsymbol{w}_j(\sigma)) - \delta_{\text{target}} \right|$
7: Evaluate the weights: $w_j^{(k)} = W_j(\boldsymbol{u}_{j-1}^{(k)})$, $k = 1, \ldots, N$
8: Estimate the ratio of normalizing constants: $\widehat{S}_j = \frac{1}{N} \sum_{k=1}^{N} w_j^{(k)}$

# Sequential Importance Sampling to estimate $P_f$ II

9: Estimate the normalizing constant $\widehat{P}_j = \widehat{P}_{j-1}\widehat{S}_j$

10: Move the samples $\boldsymbol{u}_{j-1}^{(k)}$ with MH-MCMC to obtain samples from $\boldsymbol{U}_j$, notation $\boldsymbol{u}_j^{(k)}$, $k = 1, \ldots, N$: Select randomly with replacement $N_c < N$ samples from $\{\boldsymbol{u}_{j-1}^{(k)}, k = 1, \ldots, N\}$, where $w_j^{(k)}$ is the probability of selecting $\boldsymbol{u}_{j-1}^{(k)}$. Each of the selected $N_c$ samples is a seed for a Markov chain of length $N/N_c$ with stationary density $g_j(\boldsymbol{u})$, see Algorithm 2.

11: Evaluate the weights for the stopping criterion: $w_{opt}^{(k)} = W_j^{(opt)}(\boldsymbol{u}_j^{(k)})$, $k = 1, \ldots, N$

12: **if** $\widehat{\text{c. o. v.}}\boldsymbol{w}_{opt} \geq \delta_{\text{target}}$ **then**

13: $j = j + 1$ and go to line 6.

14: **end if**

15: Estimate $\widehat{P}_f = \widehat{P}_j \times \frac{1}{N}\sum_{k=1}^{N} w_{opt}^{(k)}$

16: Output: $\widehat{P}_f$

# Metropolis–Hastings MCMC

(Metropolis et al., 1953; Hastings, 1970)

**Algorithm 2** Metropolis–Hastings MCMC to generate samples from $g_j$

1: Input: length of Markov chain $N_L$, seed $\boldsymbol{u}^{(0)}$, proposal density $q(\cdot|\cdot)$
2: **for** $k = 1 : N_L$ **do**
3: Generate a candidate state $\boldsymbol{v}$ from the proposal density $q(\cdot|\boldsymbol{u}^{(k-1)})$
4: Calculate the ratio

$$r(\boldsymbol{u}^{(k-1)}, \boldsymbol{v}) = \frac{\eta_j(\boldsymbol{v})}{\eta_j(\boldsymbol{u}^{(k-1)})} \frac{q(\boldsymbol{u}^{(k-1)}|\boldsymbol{v})}{q(\boldsymbol{v}|\boldsymbol{u}^{(k-1)})}$$

5: Calculate $\alpha(\boldsymbol{u}^{(k-1)}, \boldsymbol{v}) = \min\{1, r(\boldsymbol{u}^{(k-1)}, \boldsymbol{v})\}$
6: Set $\boldsymbol{u}^{(k)} = \boldsymbol{v}$ with probability $\alpha(\boldsymbol{u}^{(k-1)}, \boldsymbol{v})$, otherwise $\boldsymbol{u}^{(k)} = \boldsymbol{u}^{(k-1)}$
7: **end for**
8: Output: Samples $\boldsymbol{u}^{(1)}, \ldots, \boldsymbol{u}^{(N_L)}$

# Outline

# Cross-entropy (CE) method I

(Rubinstein, 1997), (Boer et al., 2005; Geyer et al., 2019)

- We assume that all distributions in this section have a density with respect to the Lebesgue measure on $\mathbb{R}^d$.
- Consider a parametric family of densities

$$\mathcal{G} = \{g(\boldsymbol{u}; \boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}, \tag{5.1}$$

  where $\boldsymbol{\theta} \in \Theta$.
- **Idea:** Determine a parameter $\boldsymbol{\theta}^* \in \Theta$, such that

$$\boldsymbol{\theta}^* = \underset{\boldsymbol{\theta} \in \Theta}{\operatorname{argmin}} \, D_{\mathsf{KL}}(g_{opt}, g(\cdot; \boldsymbol{\theta})), \tag{5.2}$$

# Cross-entropy (CE) method II

where $g_{opt} = \mathbb{1}_{\{G \leq 0\}}(\boldsymbol{u}) f(\boldsymbol{u}) / P_f$ is the optimal importance sampling density for estimating $P_f$, and $D_{\mathsf{KL}}(\cdot, \cdot)$ is the Kullback–Leibler (KL) divergence of $g_{opt}$ and the parametric density $g(\cdot, \boldsymbol{\theta})$,

$$D_{\mathsf{KL}}(g_{opt}, g(\cdot; \boldsymbol{\theta})) = \int_{\mathbb{R}^n} g_{opt}(\boldsymbol{u}) \ln g_{opt}(\boldsymbol{u}) \mathrm{d}\boldsymbol{u} - \int_{\mathbb{R}^n} g_{opt}(\boldsymbol{u}) \ln g(\boldsymbol{u}; \boldsymbol{\theta}) \mathrm{d}\boldsymbol{u}.$$

- Since the first term in the sum on the right-hand side does not depend on $\boldsymbol{\theta}$ the solution of (5.2) can equivalently be found by minimizing the so-called cross-entropy of $g_{opt}$ and $g(\cdot; \boldsymbol{\theta})$,

$$- \int_{\mathbb{R}^n} g_{opt}(\boldsymbol{u}) \ln g(\boldsymbol{u}; \boldsymbol{\theta}) \mathrm{d}\boldsymbol{u}.$$

- Equivalent maximization problem:

$$\max_{\boldsymbol{\theta} \in \Theta} \int_{\mathbb{R}^n} g_{opt}(\boldsymbol{u}) \ln g(\boldsymbol{u}; \boldsymbol{\theta}) \mathrm{d}\boldsymbol{u} \tag{5.3}$$

# Cross-entropy (CE) method III

- Substitute $g_{opt}$ into (5.3) and obtain

$$\max_{\boldsymbol{\theta}\in\Theta} \frac{1}{P_f} \int_{\mathbb{R}^n} \mathbb{1}_{\{G\leq 0\}}(\boldsymbol{u}) f(\boldsymbol{u}) \ln g(\boldsymbol{u};\boldsymbol{\theta}) \mathrm{d}\boldsymbol{u}$$

  which is equivalent to

$$\max_{\boldsymbol{\theta}\in\Theta} \mathbb{E}\left[\mathbb{1}_{\{G(\boldsymbol{U})\leq 0\}} \ln g(\boldsymbol{U};\boldsymbol{\theta})\right]. \tag{5.4}$$

- Estimate the expectation in (5.4) by importance sampling.

# A single step of the CE method I

- To simplify the presentation we assume that the nominal density $f(\boldsymbol{u}) = g(\boldsymbol{u}; \boldsymbol{\theta}_0) \in \mathcal{G}$.
- Let $g(\cdot; \boldsymbol{\theta}_{\text{ref}}) \in \mathcal{G}$ denote a density in the parametric family, where $\boldsymbol{\theta}_{\text{ref}} \in \Theta$ is fixed.
- Denoting the likelihood ratio by

$$W(\boldsymbol{u}; \boldsymbol{\theta}_0, \boldsymbol{\theta}_{\text{ref}}) = \frac{g(\boldsymbol{u}; \boldsymbol{\theta}_0)}{g(\boldsymbol{u}; \boldsymbol{\theta}_{\text{ref}})},$$

we obtain the optimization problem

$$\max_{\boldsymbol{\theta} \in \Theta} \mathbb{E}\left[ \mathbb{1}_{\{G(\boldsymbol{U}_{\text{ref}}) \leq 0\}} W(\boldsymbol{U}_{\text{ref}}; \boldsymbol{\theta}_0, \boldsymbol{\theta}_{\text{ref}}) \ln g(\boldsymbol{U}_{\text{ref}}; \boldsymbol{\theta}) \right], \tag{5.5}$$

where $\boldsymbol{U}_{\text{ref}} \sim g(\cdot; \boldsymbol{\theta}_{\text{ref}})$. Note that in (5.5) we have three densities: the nominal density $g(\cdot; \boldsymbol{\theta}_0)$, the importance sampling density $g(\cdot; \boldsymbol{\theta}_{\text{ref}})$ and the candidate density $g(\cdot; \boldsymbol{\theta})$, where we optimize with respect to $\boldsymbol{\theta}$.

# A single step of the CE method II

- Given $N$ samples of $\boldsymbol{U}_{\text{ref}}$, notation $\boldsymbol{u}_{\text{ref}}^{(i)}$, $i = 1, \ldots, N$ we approximate the expectation in (5.5) by the stochastic counterpart of (5.5),

$$\max_{\boldsymbol{\theta} \in \Theta} \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{\{G \leq 0\}}(\boldsymbol{u}_{\text{ref}}^{(i)}) W(\boldsymbol{u}_{\text{ref}}^{(i)}; \boldsymbol{\theta}_0, \boldsymbol{\theta}_{\text{ref}}) \ln g(\boldsymbol{u}_{\text{ref}}^{(i)}; \boldsymbol{\theta}). \tag{5.6}$$

- If the objective function in (5.6) is differentiable with respect to $\boldsymbol{\theta}$, then we obtain the necessary optimality condition

$$\frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{\{G \leq 0\}}(\boldsymbol{u}_{\text{ref}}^{(i)}) W(\boldsymbol{u}_{\text{ref}}^{(i)}; \boldsymbol{\theta}_0, \boldsymbol{\theta}_{\text{ref}}) \nabla_{\boldsymbol{\theta}} \ln g(\boldsymbol{u}_{\text{ref}}^{(i)}; \boldsymbol{\theta}) = \mathbf{0}, \tag{5.7}$$

  where $\nabla_{\boldsymbol{\theta}}$ denotes the gradient with respect to $\boldsymbol{\theta}$.

# Natural exponential family

Assume that the parametric family $\mathcal{G}$ in (5.1) is a so-called natural exponential family (NEF),

$$\mathcal{G} = \{\exp(\boldsymbol{\theta}^\top \boldsymbol{u} - A(\boldsymbol{u}) - C(\boldsymbol{\theta})), \boldsymbol{\theta} \in \Theta\},$$

where $C(\cdot)$ is a smooth, convex function of $\boldsymbol{\theta}$. Then the optimal parameter $\boldsymbol{\theta}$ solving (5.7) can be computed *analytically*. To see this, observe that in this case

$$\nabla_{\boldsymbol{\theta}} \ln g(\boldsymbol{u}; \boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}}(\boldsymbol{\theta}^\top \boldsymbol{u} - A(\boldsymbol{u}) - C(\boldsymbol{\theta})) = \boldsymbol{u} - \nabla_{\boldsymbol{\theta}} C(\boldsymbol{\theta}).$$

Hence the optimal $\boldsymbol{\theta}_{opt}$ solves the equation

$$\frac{\sum_{i=1}^N H^{(i)}(\boldsymbol{u}_{\text{ref}}^{(i)}) \boldsymbol{u}_{\text{ref}}^{(i)}}{\sum_{i=1}^N H^{(i)}(\boldsymbol{u}_{\text{ref}}^{(i)})} = \nabla_{\boldsymbol{\theta}} C(\boldsymbol{\theta}_{opt}), \tag{5.8}$$

where

$$H^{(i)}(\boldsymbol{u}_{\text{ref}}^{(i)}) := \mathbb{1}_{\{G \leq 0\}}(\boldsymbol{u}_{\text{ref}}^{(i)}) W(\boldsymbol{u}_{\text{ref}}^{(i)}; \boldsymbol{\theta}_0, \boldsymbol{\theta}_{\text{ref}}), \quad i = 1, \ldots, N.$$

# Iterative choice of reference density

**Question:** How to choose the parameter $\boldsymbol{\theta}_{\text{ref}}$ in (5.6)?

- For an arbitrary guess $\boldsymbol{\theta} \in \Theta$ the density $g(\cdot;\boldsymbol{\theta})$ might have a large distance to $g_{opt}$. Then, most samples of $\boldsymbol{U}_{\text{ref}}$ will be outside of the failure domain $\{\boldsymbol{u} \in \mathbb{R}^n, G(\boldsymbol{u}) \leq 0\}$.
  $\rightarrow$ Inaccurate approximation of the expectation in the objective function
- **Idea:** Replace the failure event in the objective function in (5.6) by a less rare event with associated failure domain $\{\boldsymbol{u} \in \mathbb{R}^n, G(\boldsymbol{u}) \leq \gamma\}$ with failure level $\gamma > 0$.
- Start with an initial choice $\gamma_0$ and iterate in $\gamma_j$ and $\boldsymbol{\theta}_j$.

# Stopping criterion, Estimate of $P_f$

- We stop the CE method at iteration $j$, where the failure level $\gamma_j \leq 0$ for the first time.
- Suppose we stop the CE method at iteration $j = M$ with associated importance sampling density $g(\cdot; \boldsymbol{\theta}_M)$ and samples $\boldsymbol{u}_M^{(i)}$, $i = 1, \ldots, N$. We can then estimate the probability of failure as

$$\widehat{P}_f = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{\{G \leq 0\}}(\boldsymbol{u}_M^{(i)}) W(\boldsymbol{u}_M^{(i)}; \boldsymbol{\theta}_0, \boldsymbol{\theta}_M).$$

# Complete algorithm I

- We define the objective function

$$J(\boldsymbol{\theta}; \gamma, \boldsymbol{\theta}_{\text{ref}}) := \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{\{G \leq \gamma\}}(\boldsymbol{u}_{\text{ref}}^{(i)}) W(\boldsymbol{u}_{\text{ref}}^{(i)}; \boldsymbol{\theta}_0, \boldsymbol{\theta}_{\text{ref}}) \ln g(\boldsymbol{u}_{\text{ref}}^{(i)}; \boldsymbol{\theta}), \tag{5.9}$$

  where $\boldsymbol{u}_{\text{ref}}^{(i)}$, $i = 1, \ldots, N$ are samples of $\boldsymbol{U}_{\text{ref}}^{(i)} \sim g(\cdot; \boldsymbol{\theta}_{\text{ref}})$ i.i.d.
- Let $\rho \in (0, 1)$ denote a fixed parameter.
- We generate $N$ samples of $\boldsymbol{U}_0 \sim g(\boldsymbol{u}; \boldsymbol{\theta}_0)$ and determine $\gamma_0$ such that $\lceil \rho N \rceil$ samples are contained in the intermediate failure domain $\{\boldsymbol{u} \in \mathbb{R}^n, G(\boldsymbol{u}) \leq \gamma_0\}$. In other words, we estimate the $\rho$-quantile of $Y = G(\boldsymbol{U}_0)$.

# Complete algorithm II

**Definition 5.1.** *Let $Y$ denote a real-valued random variable with probability distribution $\mathbb{P}_Y$. Let $\rho \in [0,1]$. The $\rho$-quantile of $Y$ is a real number $\gamma$ such that*

$$\mathbb{P}_Y(Y \leq \gamma) \geq \rho \text{ and } \mathbb{P}_Y(Y \geq \gamma) \geq 1 - \rho.$$

A (biased) estimator for the $\rho$-quantile of $Y$ can be obtained by drawing $N$ i.i.d. samples of $Y$, ordering the samples from smallest to largest, $y^{(1)} \leq y^{(2)} \leq \cdots \leq y^{(N)}$, and estimating the $\rho$-quantile as

$$\widehat{\gamma} = y^{(\lceil \rho N \rceil)}. \tag{5.10}$$

The estimate in (5.10) is called sample $\rho$-quantile.

# Complete algorithm III

- Next we calculate

$$\boldsymbol{\theta}_1 = \underset{\boldsymbol{\theta}\in\Theta}{\operatorname{argmax}} \; J(\boldsymbol{\theta}; \gamma_0, \boldsymbol{\theta}_0).$$

  Then we generate $N$ samples of $\boldsymbol{U}_1 \sim g(\boldsymbol{x}; \boldsymbol{\theta}_1)$ and determine $\gamma_1$ such that $\lceil \rho N \rceil$ samples are contained in the intermediate failure domain $\{\boldsymbol{u} \in \mathbb{R}^n, G(\boldsymbol{u}) \leq \gamma_1\}$.

- Then we calculate

$$\boldsymbol{\theta}_2 = \underset{\boldsymbol{\theta}\in\Theta}{\operatorname{argmax}} \; J(\boldsymbol{\theta}; \gamma_1, \boldsymbol{\theta}_1),$$

  and so on, until $\gamma_j \leq 0$ for the first time. Algorithm 3 summarizes the steps of the CE method for rare event estimation.

# Cross-entropy method to estimate $P_f$ I

**Algorithm 3** (Boer et al., 2005)

1: Input: number of samples $N$, fraction $\rho \in (0,1)$
2: Set $j = 0$ and $\boldsymbol{\theta}_0$ such that $g(\cdot;\boldsymbol{\theta}_0) = f(\boldsymbol{u})$.
3: Generate $N$ samples of $\boldsymbol{U}_j \sim g(\cdot;\boldsymbol{\theta}_j)$, notation $\boldsymbol{u}_j^{(i)}$, $i = 1,\dots,N$
4: Evaluate $y_j^{(i)} = G(\boldsymbol{u}_j^{(i)})$, $i = 1,\dots,N$
5: Determine the sample $\rho$-quantile of the $y_j^{(i)}$, $i = 1,\dots,N$, notation $\widehat{\gamma}_j$.
6: **if** $\widehat{\gamma}_j \leq 0$ **then** go to line 10
7: **end if**
8: Calculate $\boldsymbol{\theta}_{j+1} = \underset{\boldsymbol{\theta}\in\Theta}{\operatorname{argmax}} \; J(\boldsymbol{\theta};\widehat{\gamma}_j,\boldsymbol{\theta}_j)$

# Cross-entropy method to estimate $P_f$ II

9: Set $j = j + 1$ and go to line 3.

10: Estimate $\widehat{P}_f = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{\{G \leq 0\}}(\boldsymbol{u}_j^{(i)}) W(\boldsymbol{u}_j^{(i)}; \boldsymbol{\theta}_0, \boldsymbol{\theta}_j)$

11: Output: $\widehat{P}_f$

---

# Outline

# Basic idea I

- Stationary process models: Subset simulation (SuS) (Au & Beck, 2001)
- Time-dependent process models: Adaptive Multilevel Splitting (ADAM) (Cérou & Guyader, 2007; Botev & Kroese, 2010)

Consider estimating the failure probability $P_f$ with failure domain

$$F := \{\boldsymbol{u} \in \mathbb{R}^n \colon G(\boldsymbol{u}) \leq 0\}.$$

**Lemma 6.1.** *Let $F_\ell \subset \mathbb{R}^n$, $\ell = 1, \dots, L$, denote auxiliary failure domains, where*

$$F_1 \supset F_2 \supset \cdots \supset F_{L-1} \supset F_L = F. \tag{6.1}$$

*Then it holds*

$$P_f = \mathbb{P}(F_L) = \mathbb{P}(F_1) \prod_{\ell=2}^{L} \mathbb{P}(F_\ell | F_{\ell-1}). \tag{6.2}$$

# Basic idea II

*Proof.* Recall that for a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and elementary events $A, B \in \mathcal{A}$ it holds

$$\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B).$$

Because of the nestedness $F_{L-1} \supset F_L$ we have $F_L = F_L \cap F_{L-1}$. Hence we obtain

$$P_f = \mathbb{P}(F_L) = \mathbb{P}\left(F_L \cap F_{L-1}\right) = \mathbb{P}\left(F_L \mid F_{L-1}\right)\mathbb{P}\left(F_{L-1}\right).$$

Repeating the argument for $\mathbb{P}(F_{L-1}), \ldots, \mathbb{P}(F_2)$ shows the claim. □

# Nested auxiliary failure domains

- Let $F_\ell := \{\boldsymbol{u} \in \mathbb{R}^n \colon G(\boldsymbol{u}) \leq \gamma_\ell\}$ denote a failure domain with associated failure level $\gamma_\ell > 0$ and probability of failure $\mathbb{P}(F_\ell)$, $\ell = 1, \dots, L-1$.
- The final failure domain $F_L = F$, thus $\gamma_L = 0$.
- We also define the failure domain $F_0 = \mathbb{R}^n$, where formally $\gamma_0 = +\infty$.
- Fix $p_0 \in (0, 1)$ and select the failure levels $\gamma_\ell$ recursively, such that $\gamma_\ell$ is the $p_0$-quantile of $Y = G(\boldsymbol{U})$ under the input distribution $\mathbb{P}_{\boldsymbol{U}|F_{\ell-1}}$, $\ell = 1, \dots, L-1$.
- $\gamma_1$ is the $p_0$-quantile of the unconditional distribution $\mathbb{P}_{\boldsymbol{U}}$.
- In practice: Use a quantile estimate $\widehat{\gamma}_\ell$ instead of $\gamma_\ell$, $\ell = 1, \dots, L-1$.

# Conditional sampling

(Papaioannou, Betz, et al., 2015)

**Algorithm 4** Metropolis–Hastings MCMC to generate samples from $\varphi_n(\cdot|F_{\ell-1})$ with Gaussian proposal density $N(\rho \boldsymbol{u}^{(k-1)}, (1-\rho^2) I_n)$

1: Input: length of Markov chain $N_L$, seed $\boldsymbol{u}^{(0)}$, correlation parameter $\rho \in [0,1]$
2: **for** $k = 1 : N_L$ **do**
3: Sample $\boldsymbol{Z} \sim N(0, I_n)$, notation $\boldsymbol{z}$
4: $\boldsymbol{v} = \rho \boldsymbol{u}^{(k-1)} + \sqrt{1-\rho^2}\boldsymbol{z}$ ▷ Generate a candidate state $\boldsymbol{v}$
5: $\boldsymbol{u}^{(k)} = \begin{cases} \boldsymbol{v}, & \boldsymbol{v} \in F_{\ell-1}, \\ \boldsymbol{u}^{(k-1)}, & \boldsymbol{v} \notin F_{\ell-1}. \end{cases}$ ▷ Accept or reject $\boldsymbol{v}$
6: **end for**
7: Output: Samples $\boldsymbol{u}^{(1)}, \ldots, \boldsymbol{u}^{(N_L)}$

# Subset simulation to estimate $P_f$ I

**Algorithm 5** (Au & Beck, 2001)

1: Input: number of samples $N$, probability of auxiliary domains $p_0 \in (0, 1)$
2: Generate $N$ samples from the distribution $\mathbb{P}_U$, notation $\boldsymbol{u}_0^{(1)}, \ldots, \boldsymbol{u}_0^{(N)}$
3: Evaluate $y_0^{(i)} = G(\boldsymbol{u}_0^{(i)})$, $i = 1, \ldots, N$
4: Order the LSF values increasingly, notation $y_0^{(1)} \leq y_0^{(2)} \leq \cdots \leq y_0^{(N)}$
5: Set $\widehat{\gamma}_1 = y_0^{(\lceil p_0 N \rceil)}$ and $\ell = 1$
6: **while** $\widehat{\gamma}_\ell > 0$ **do**
7: Determine the number of samples $N_\ell$ for which $\boldsymbol{u}_{\ell-1}^{(i)} \in F_\ell$
8: Calculate $\widehat{P}_\ell = N_\ell / N$

# Subset simulation to estimate $P_f$ II

9: Use the $N_\ell$ failure points $\boldsymbol{u}_{\ell-1}^{(i)} \in F_\ell$ as seeds and generate $N - N_\ell$ samples $\boldsymbol{u}_\ell^{(i)}$ with Algorithm 4.

10: Evaluate $y_\ell^{(i)} = G(\boldsymbol{u}_\ell^{(i)})$, $i = 1, \dots, N$

11: Order the LSF values increasingly, $y_\ell^{(1)} \leq y_\ell^{(2)} \leq \cdots \leq y_\ell^{(N)}$

12: Set $\widehat{\gamma}_{\ell+1} = y_\ell^{(\lceil p_0 N \rceil)}$ and $\ell = \ell + 1$

13: **end while**

14: Determine the number of samples $N_F$ for which $\boldsymbol{u}_{\ell-1}^{(i)} \in F$

15: Calculate $\widehat{P}_\ell = N_F / N$

16: Calculate $\widehat{P}_f = \prod_{k=1}^{\ell} \widehat{P}_k$

17: Output: $\widehat{P}_f$

# Non-nested auxiliary failure domains

**Lemma 6.2** ((Ullmann & Papaioannou, 2015)). *Let $F_\ell \subset \mathbb{R}^n$, $\ell = 1, \dots, L$, where the failure domains are not necessarily nested. Then it holds*

$$P_f = \mathbb{P}(F_L) = \mathbb{P}(F_1) \prod_{\ell=2}^{L} \frac{\mathbb{P}(F_\ell | F_{\ell-1})}{\mathbb{P}(F_{\ell-1} | F_\ell)}. \tag{6.3}$$

*Proof.* Consider the failure domains $F_L$ and $F_{L-1}$. The application of Bayes' rule in two ways gives

$$\mathbb{P}(F_{L-1}|F_L)\mathbb{P}(F_L) = \mathbb{P}(F_L \cap F_{L-1}) = \mathbb{P}(F_L|F_{L-1})\mathbb{P}(F_{L-1}).$$

Hence

$$P_f = \mathbb{P}(F_L) = \frac{\mathbb{P}(F_L|F_{L-1})}{\mathbb{P}(F_{L-1}|F_L)}\mathbb{P}(F_{L-1}).$$

Repeating the argument for $\mathbb{P}(F_{L-1}), \dots, \mathbb{P}(F_2)$ shows the claim. □

# Part III: Semi-analytical methods

7 First order reliability method

8 Line sampling

# Outline

# First order reliability method I

(Hasofer & Lind, 1974)

The FORM estimate for $P_f$ is defined as

$$P_f^{\mathrm{FORM}} = \Phi(-\|\boldsymbol{u}^{\mathrm{MLFP}}\|_2),$$

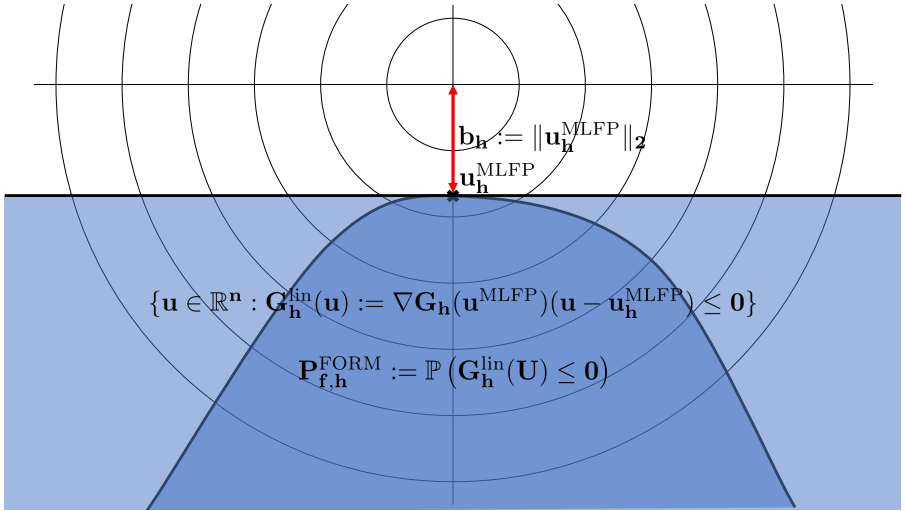
where

$\Phi$ is the cumulative distribution function of a univariate standard normal random variable,

$u^{\mathrm{MLFP}}$ is a most likely failure point (MLFP)

$$\boldsymbol{u}^{\mathrm{MLFP}} := \operatorname*{argmin}_{\boldsymbol{u}\in\mathbb{R}^n} \quad \frac{1}{2}\|\boldsymbol{u}\|_2^2, \quad \text{such that} \quad G(\boldsymbol{u}) = 0.$$

# First order reliability method II



# First order reliability method - cont.

**Step 1.** Define direction $\boldsymbol{\alpha} \in \mathbb{R}^n$, $\boldsymbol{\alpha}^\top \boldsymbol{\alpha} = 1$, pointing towards the most likely failure point (assuming there is only one).

**Step 2.** Collect $n-1$ orthonormal basis vectors of $\boldsymbol{\alpha}^\perp$ in the matrix $A \in \mathbb{R}^{n \times (n-1)}$.

**Step 3.** Partition the standard normal space

$$\boldsymbol{u} = \boldsymbol{\alpha}(\boldsymbol{\alpha}^\top \boldsymbol{u}) + A(A^\top \boldsymbol{u})$$

and rotate: $\boldsymbol{z} = R\boldsymbol{u}$, $R = \begin{bmatrix} \boldsymbol{\alpha}^\top \\ A^\top \end{bmatrix}$.

**Key observation**

The distribution of the rotated random vector $\boldsymbol{Z} = R\boldsymbol{U}$ is still standard normal.

# First order reliability method - cont.

Calculate the probability of failure in the rotated space ($\boldsymbol{u} = R^\top \boldsymbol{z}$)

$$
\begin{aligned}
P_f &= \int_{\mathbb{R}^n} \mathcal{I}\left(G(\boldsymbol{u}) \leq 0\right) \varphi_n(\boldsymbol{u}) \mathrm{d}\boldsymbol{u} \\
&= \int_{\mathbb{R}^n} \mathcal{I}(G(R^\top \boldsymbol{z}) \leq 0) \underbrace{\varphi_n(R^\top \boldsymbol{z})}_{=\varphi_n(\boldsymbol{z})} \mathrm{d}\boldsymbol{z} \\
&= \int_{\mathbb{R}^{n-1}} \int_{\mathbb{R}} \mathcal{I}(G(R^\top \boldsymbol{z}) \leq 0) \varphi_1(z_1) \mathrm{d}z_1 \varphi_{n-1}(\boldsymbol{z}_{2:n}) \mathrm{d}\boldsymbol{z}_{2:n}
\end{aligned}
$$

## Assumption

(*i*) $G(\mathbf{0}) > 0$,

(*ii*) for every $\boldsymbol{z}_{2:n} \in \mathbb{R}^{n-1}$ there is a unique solution of $G(R^\top [s; \boldsymbol{z}_{2:n}]) = 0$, denoted by $s(\boldsymbol{z}_{2:n})$.

# First order reliability method - cont.

$$P_f = \int_{\mathbb{R}^{n-1}} \Phi(-s(\boldsymbol{z}_{2:n}))\varphi_{n-1}(\boldsymbol{z}_{2:n})\mathrm{d}\boldsymbol{z}_{2:n}.$$

Under the assumptions of FORM, $\Phi(-s(\boldsymbol{z}_{2:n}))$ does not depend on $\boldsymbol{z}_{2:n}$.

$$\begin{aligned} P_f^{\mathrm{FORM}} &= \int_{\mathbb{R}^{n-1}} \Phi(-s(\boldsymbol{z}_{2:n}))\varphi_{n-1}(\boldsymbol{z}_{2:n})\mathrm{d}\boldsymbol{z}_{2:n} \\ &= \Phi(-\|\boldsymbol{u}^{\mathrm{MLFP}}\|_2) \underbrace{\int_{\mathbb{R}^{n-1}} \varphi_{n-1}(\boldsymbol{z}_{2:n})\mathrm{d}\boldsymbol{z}_{2:n}}_{=1} \\ &= \Phi(-\|\boldsymbol{u}^{\mathrm{MLFP}}\|_2). \end{aligned}$$

# Outline

# Line sampling

(Koutsourelakis et al., 2004)

$$P_f = \int_{\mathbb{R}^{n-1}} \Phi(-s(\boldsymbol{z}_{2:n}))\varphi_{n-1}(\boldsymbol{z}_{2:n})\mathrm{d}\boldsymbol{z}_{2:n} = \mathbb{E}_{\varphi_{n-1}}[\Phi(-s(\boldsymbol{z}_{2:n}))]$$

We approximate the expectation by Crude Monte Carlo:

$$\widehat{P}_f^{\mathrm{LS}} = \frac{1}{N}\sum_{i=1}^{N} \Phi(-s(\boldsymbol{z}_{2:n}^{(i)})).$$

# Line Sampling - cont.

Line Sampling combines two mathematical properties:

1. Smoothing by Pre-Integration: Replace indicator function by standard normal cdf

$$P_f = \int_{\mathbb{R}^n} \mathcal{I}(G(\boldsymbol{u}) \leq 0)\varphi_n(\boldsymbol{u})\mathrm{d}\boldsymbol{u} = \int_{\mathbb{R}^{n-1}} \Phi(-s(\boldsymbol{z}_{2:n}))\varphi_{n-1}(\boldsymbol{z}_{2:n})\mathrm{d}\boldsymbol{z}_{2:n}$$

   See the Quasi-Monte Carlo literature, e.g., (Griewank et al., 2018; Gilbert et al., 2023; S. Liu & Owen, 2023)

2. Variance reduction by Conditioning: Depends heavily on the direction $\boldsymbol{\alpha}$

$$P_f = \mathbb{E}[\mathbb{1}_{\{G(\boldsymbol{U})\leq 0\}}] = \mathbb{E}[\mathbb{E}[\mathbb{1}_{\{G(\boldsymbol{Z})\leq 0\}}|\boldsymbol{Z}_{2:n}]]$$

# Part IV: Advanced topics

9 Error analysis for failure probabilities with approximate models

10 Ensemble Kalman filter for rare event estimation

11 Consensus-based rare event estimation

# Acknowledgments

Konstantin Althaus
Jonas Latz *(The University of Manchester)*
Fabian Wagner
Iason Papaioannou *(TUM)*

# Outline

# Problem formulation

Estimate the probability of failure

$$P_f := \mathbb{P}\left(G(\boldsymbol{U}) \leq 0\right) = \int_{\mathbb{R}^n} \mathcal{I}\left(G(\boldsymbol{u}) \leq 0\right) \varphi_n(\boldsymbol{u}) \mathrm{d}\boldsymbol{u}$$

**In practice:** Approximate limit state function $G_h : \mathbb{R}^n \to \mathbb{R}$

Approximate probability of failure

$$P_{f,h} := \mathbb{P}\left(G_h(\boldsymbol{U}) \leq 0\right) = \int_{\mathbb{R}^n} \mathcal{I}\left(G_h(\boldsymbol{u}) \leq 0\right) \varphi_n(\boldsymbol{u}) \mathrm{d}\boldsymbol{u}$$

with discretization parameter $h > 0$

# Challenge

**In practice:** Approximate limit state function $G_h : \mathbb{R}^n \to \mathbb{R}$
Approximate probability of failure

$$P_{f,h} := \mathbb{P}\left(G_h(\boldsymbol{U}) \leq 0\right) = \int_{\mathbb{R}^n} \mathcal{I}\left(G_h(\boldsymbol{u}) \leq 0\right) \varphi_n(\boldsymbol{u}) \mathrm{d}\boldsymbol{u}$$

with discretization parameter $h > 0$

# Approximation error

(Elfverson, Hellman, et al., 2016) bound the absolute error

$$|P_f - P_{f,h}| \leq Ch^s.$$

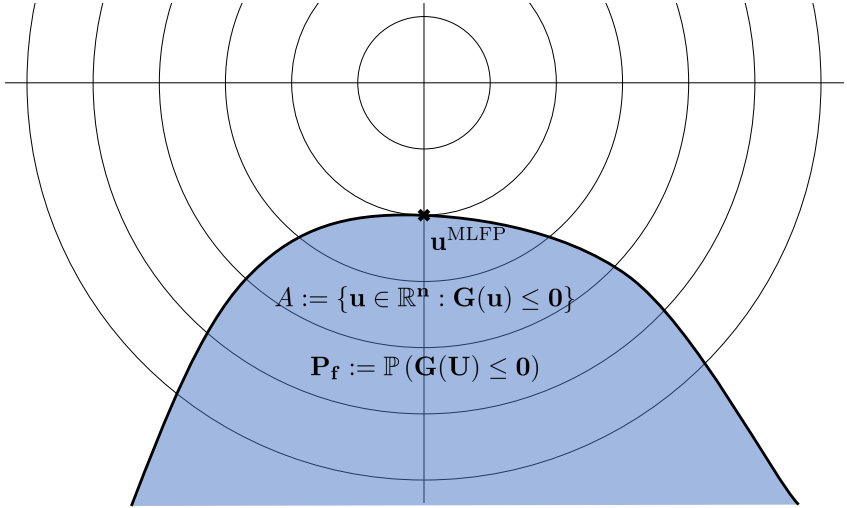
**Goal:** Bound the relative error

$$\frac{|P_f - P_{f,h}|}{P_f} \leq Ch^s.$$

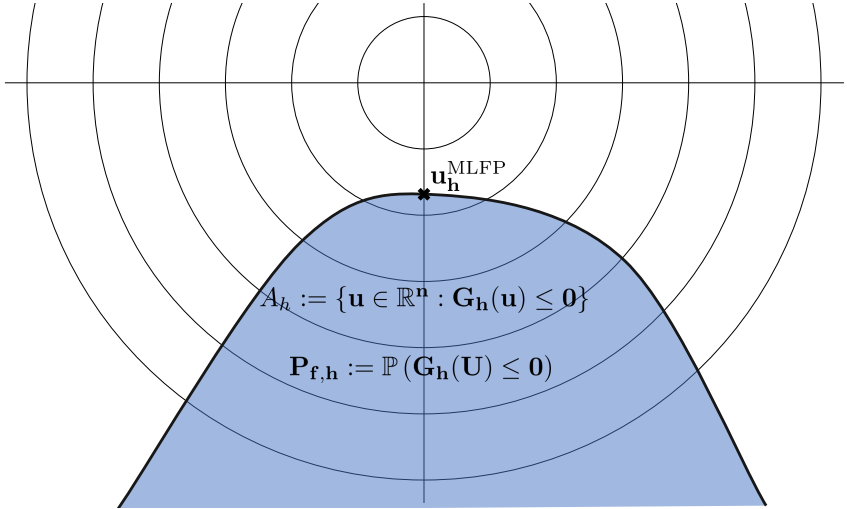
Starting point: Gaussian measure of the symmetric difference

$$|P_f - P_{f,h}| \leq \mathbb{P}\left(\{G(\boldsymbol{U}) \leq 0\} \cap \{G_h(\boldsymbol{U}) > 0\}\right) + \mathbb{P}\left(\{G(\boldsymbol{U}) > 0\} \cap \{G_h(\boldsymbol{U}) \leq 0\}\right)$$

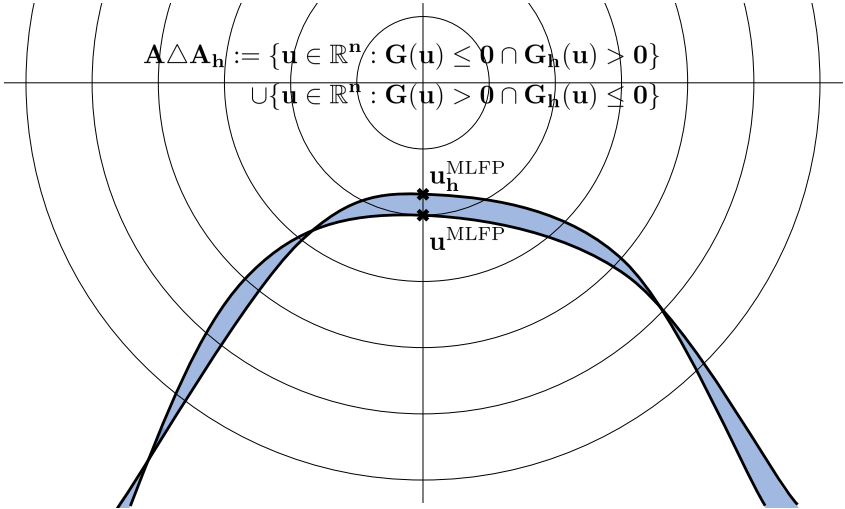
# Exact failure domain



# Approximate failure domain



# Upper bound for approximation error $|P_f - P_{f,h}|$



# Distance of limit state surfaces

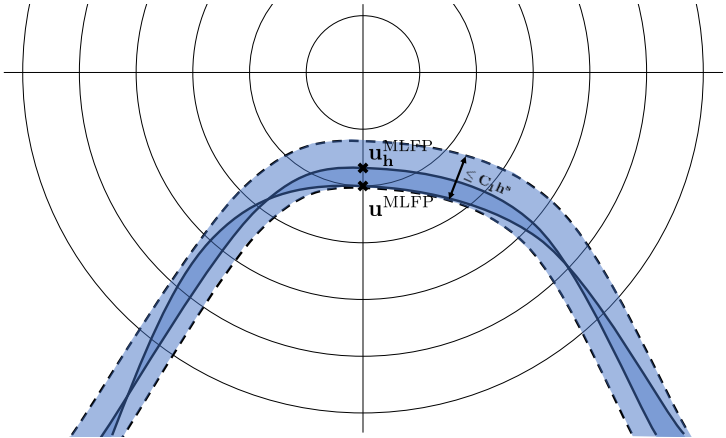



**Figure 1** The distance between the exact and approximate failure surface behaves as $C_1 h^s$; proof uses results from PDE-constrained parameter identification problems (Vexler, 2004).

# Bound of the Gaussian measure

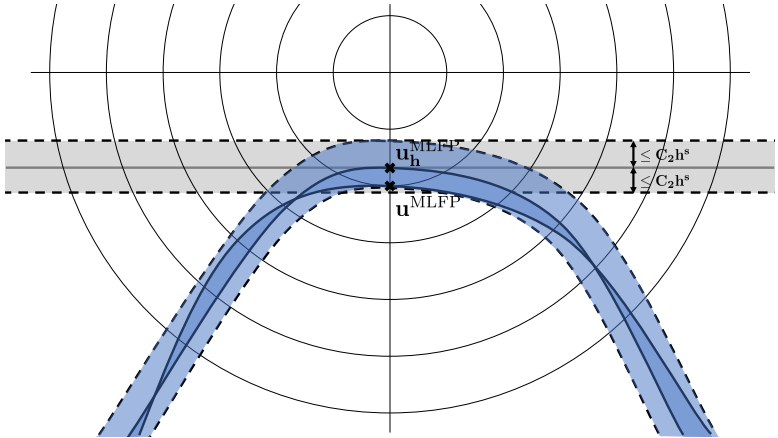



**Figure 2** The Gaussian measure of the **blue set** is upper bounded by the Gaussian measure of the **gray set**; proof uses convexity of $A$ and $A_h$.

# PDE and assumptions

The evaluation of $G$ requires the weak solution $y : \overline{D} \times \Omega \to \mathbb{R}$ of the elliptic diffusion equation

$$\begin{aligned} -\nabla_x \cdot (a(x,\omega)\nabla_x y(x,\omega)) &= f(x) \quad \forall x \in D, \\ y(x,\omega) &= 0 \qquad \forall x \in \partial D, \quad \mathbb{P}\text{-a.s. in } \Omega, \end{aligned}$$

- $(\Omega, \mathcal{A}, \mathbb{P})$ is a probability space,
- $D \subset \mathbb{R}^d$, $d = 1, 2, 3$, is an open, bounded, convex, polygonal domain,
- $a : D \times \Omega \to \mathbb{R}$ is a uniformly bounded and elliptic random diffusion coefficient,
- depends on $n$ i.i.d. standard normal random variables, s.t. $y(\cdot, \omega) \in H^2(D)$, $f \in L^2(D)$.

**Assumption:** uniformly bounded approximation error

$$|G(U(\omega)) - G_h(U(\omega))| = |\mathcal{F} y(\cdot,\omega) - \mathcal{F}_h y_h(\cdot,\omega)| \leq C_{\mathrm{FE}} h^s, \quad \mathbb{P}\text{-a.s.},$$

with $C_{\mathrm{FE}} > 0$, $s > 0$ indep. of $h$.

# Main results

**Theorem 9.1** (informal, (Wagner, Latz, et al., 2021)). *For convex, unbounded failure domains $A$, $A_h$ and for $h > 0$ sufficiently small [2], the error of the exact and approximate probability of failure is upper bounded by*

$$|P_f - P_{f,h}| \leq \widehat{C} \cdot h^s \cdot P_{f,h}^{\mathrm{FORM}},$$

*where $\widehat{C}$ is independent of $P_{f,h}^{\mathrm{FORM}}$.*

**Theorem 9.2** (informal, (Wagner, Latz, et al., 2021)). *For $h > 0$ sufficiently small [1], the relative error of the FORM estimates is upper bounded by*

$$\frac{|P_f^{\mathrm{FORM}} - P_{f,h}^{\mathrm{FORM}}|}{P_f^{\mathrm{FORM}}} \leq \widehat{C}^{\mathrm{FORM}} \cdot h^s,$$

*where $\widehat{C}^{\mathrm{FORM}}$ is independent of $P_f^{\mathrm{FORM}}$.*

[2] and assumptions on parametric regularity of $a$, local Lipschitz continuity of CDFs of $G$, $G_h$, gradients $\nabla_u G$, $\nabla_u G_h$; details in SINUM paper

# Proof outline

1) The distance between the exact and approximate failure surface behaves as $C_1 h^s$.

2) The symmetric difference of the failure domains is bounded by an interval in $1$D with length $C_2 h^s$.

3) The Gaussian measure of the interval in $1$D is upper bounded by $C_3 h^s P_{f,h}^{\mathrm{FORM}}$.

# Final step

So far: The approximation error is bounded by the Gaussian measure of the gray set,

$$|P_f - P_{f,h}| \leq \mathbb{P}\left(U_1 \in ]-b_h - C_2 h^s, -b_h + C_2 h^s]\right),$$

where $U_1$ is a univariate Gaussian random variable and $b_h := \|u_h^{\mathrm{MLFP}}\|_2$.

In the final step, with $G_{\mathrm{lin}} := U_1 + b_h$ we show that

$$\mathbb{P}\left(U_1 \in ]-b_h - C_2 h^s, -b_h + C_2 h^s]\right) \leq C_3 \cdot h^s \cdot P_{f,h}^{\mathrm{FORM}},$$

where $P_{f,h}^{\mathrm{FORM}} = \mathbb{P}\left(U_1 \leq -b_h\right)$.

# Numerical example

$y$ is the weak solution of the elliptic BVP

$$-\nabla_x \cdot (a(x,\omega)\nabla_x y(x,\omega)) = 1, \quad x \in D = (0,1),$$
$$y(0,\omega) = 1, \quad y(1,\omega) = 0, \ \mathbb{P} - \text{a.s. in } \Omega,$$

with lognormal diffusion coefficient $a(x,\omega) = \exp(Z(x,\omega))$.

$Z$ is Gaussian random field with constant mean $\mathbb{E}[Z] = 0.1$, Whittle–Matérn covariance function, smoothness $\nu = 1/2$, variance $\sigma^2 = 0.04$, correlation length $\rho = 0.1$. The KLE of $Z$ is truncated after $n = 50$ leading terms, $Z \approx Z_n(x, U(\omega))$, $U$ is a $n$-dim. standard normal random vector, $a \approx a_n(x, U(\omega))$.

Failure events $\omega \in \Omega$ satisfy $G(U(\omega)) = 1.5 - (-a_n(1, U(\omega))y_x(1,\omega)) \leq 0$. The reference probability of failure is estimated as $P_f = 7.18 \times 10^{-5}$ for $h = 2^{-12}$.

The approximation error of the limit state function satisfies (not uniformly bounded)

$$|G(U(\omega)) - G_h(U(\omega))| \leq C_{\mathrm{FE}}(\omega)h.$$

# Error bounds

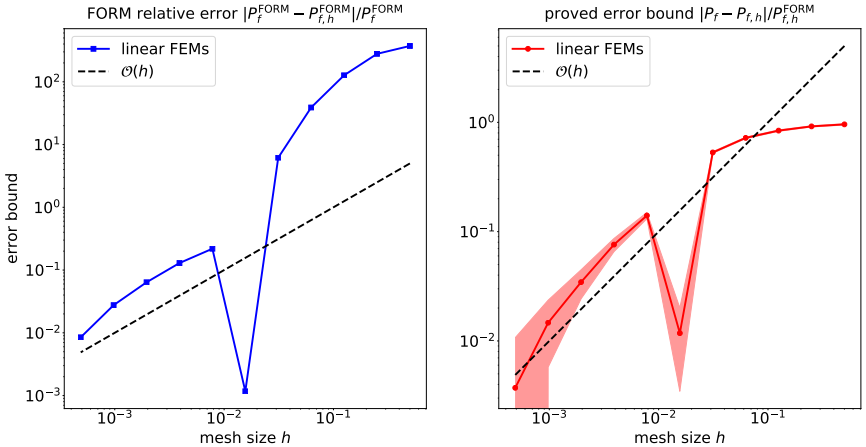



**Figure 3** Left: relative error of the FORM estimates; Right: error bound

# Summary

- Approximation error of the probability of failure $P_f$ bounded by $h^s \times P_{f,h}^{\mathrm{FORM}}$
- Relative error of the FORM estimate bounded by $h^s$

**Open problems:**

- Bounding the relative error of the probability of failure
- Practical, reliable error estimate $\longrightarrow$ Adaptivity!
- Error analysis in infinite-dimensional parameter space

# Outline

# Importance sampling

**Idea:** Change of measure

$$P_f = \int_{\mathbb{R}^n} \underbrace{\mathcal{I}\left(G(u) \leq 0\right)}_{\text{Indicates failure event}} \varphi_n(u)\mathrm{d}u = \int_{\mathbb{R}^n} \mathcal{I}\left(G(u) \leq 0\right) w(u)p(u)\mathrm{d}u$$

$p(u)$ importance sampling (IS) density, $p(u) > 0$ on failure domain $A$

$w(u) := \frac{\varphi_n(u)}{p(u)}$ importance weight, likelihood ratio

The **optimal IS density**

$$p_{\text{opt}}(u) = \frac{1}{P_f}\mathcal{I}\left(G(u) \leq 0\right)\varphi_n(u)$$

gives a zero-variance Monte Carlo estimator.

# Sequential Importance Sampling (SIS)

Problem: The optimal IS density is not accessible.

Idea:

- Construct a **sequence of auxiliary densities** which gradually approximate the optimal IS density

$$p_0 = \varphi_n \longrightarrow p_1 \longrightarrow p_2 \longrightarrow \cdots \longrightarrow p_k \approx p_{\text{opt}}$$

- Approximate each density by weighted **samples**

# In this talk

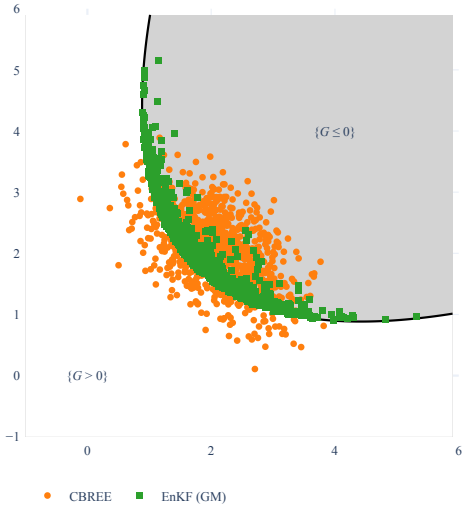
**Key idea**

**Step 1.** Generate failure samples: Move initial ensemble of particles to the failure domain along the trajectories of a stochastic differential equation (SDE)

**Step 2.** Use final ensemble with Importance Sampling

- No MCMC steps & MCMC parameter tuning required
- Use established SDE theory and discretization

# In a picture



# Sampling-based methods

In Bayesian inverse problems: Generate samples of the posterior distribution

- Ensemble Kalman filter for inverse problems (M. A. Iglesias et al., 2013)
- Consensus-based sampling (Carrillo et al., 2022)

In rare event estimation: Generate samples of the optimal IS distribution

- Ensemble Kalman filter for rare event estimation (Wagner, Papaioannou, et al., 2022)
- Consensus-based rare event estimation (Althaus et al., 2024)

# Rare event estimation as an inverse problem

Define the auxiliary limit state function $\widetilde{G}(u) := \max\{0, G(u)\}$, the prior measure $\mu_0(du) = \varphi_n(du)$, and the data $y = 0$.

The goal is to find states $u \in \mathbb{R}^n$ such that

$$y = \widetilde{G}(u) + \eta, \tag{RE-IP}$$

where $\eta \sim \mathrm{N}(0, 1)$ is observational noise.

**Two steps:**

1. Apply the Ensemble Kalman filter for inverse problems (EKI) to problem (RE-IP)

   $\longrightarrow$ generate failure samples

2. Fit a distribution model with the final ensemble and apply Importance Sampling

# Step 1: Approximation of $p_{\text{opt}}$

The posterior measure of the inverse problem (RE-IP) is (Dashti & Stuart, 2017)

$$\mu^y(du) = \frac{1}{Z} \exp\left(-\frac{1}{2}\widetilde{G}(u)^2\right) \varphi_n(du).$$

With the temperatures $\infty = \sigma_0 > \sigma_1 > \sigma_2 > \cdots > 0$, we define the EnKF densities

$$p_k^{\text{EnKF}}(u) = \frac{1}{Z_k} \exp\left(-\frac{1}{2\sigma_k}\widetilde{G}(u)^2\right) \varphi_n(u),$$

which are approximated by an ensemble of particles as in (Schillings & Stuart, 2017).

**Key observation**

$$\lim_{\sigma_k \to 0} p_k^{\text{EnKF}}(u) = \frac{1}{P_f} \mathcal{I}(G(u) \leq 0) \varphi_n(u) = p_{\text{opt}}(u)$$

# Step 2: Estimation of the probability of failure

Problem: In general, the final EnKF ensemble is not distributed according to optimal IS density (Ernst et al., 2015).

Idea: Fit a distribution model with the final ensemble and apply Importance Sampling.

We generate the ensemble $\{\widehat{u}^{(j)}\}_{j=1}^{J}$ from the fitted distribution and estimate $P_f$ by

$$\widehat{P}_f = \frac{1}{J}\sum_{j=1}^{J} \mathcal{I}\left(G\left(\widehat{u}^{(j)}\right) \leq 0\right) \frac{\varphi_n(\widehat{u}^{(j)})}{\widehat{p}(\widehat{u}^{(j)})},$$

where $\widehat{p}(\cdot)$ is the pdf of the fitted distribution.

# Further details (Wagner et al., JUQ, 2022)

- More experiments with PDE-based limit state function in $n = 150$ dimensions and 2D physical space
- Adaptive tempering as in (Papaioannou, Papadimitriou, et al., 2016) and (M. Iglesias et al., 2018) to determine the temperatures $\{\sigma_k\}_{k=0}^{N}$ resp. EnKF step size
- Treat multi-modal failure domains by localisation of covariance matrices around each particle (Reich & Weissmann, 2021)
- Analysis of the mean-field limit and the continuous time limit of the particles in the noise-free case similar to (Schillings & Stuart, 2017)

# Details: EnKF iteration

The ensemble $\mathbf{u_k} = \{u_k^{(j)}\}_{j=1}^J$ approximates the $k$th density $p_k^{\mathrm{EnKF}}$ and the EnKF update is given by (Schillings & Stuart, 2017)

$$u_{k+1}^{(j)} = u_k^{(j)} + C_{\mathrm{up}}(\mathbf{u_k}) \left( C_{\mathrm{pp}}(\mathbf{u_k}) + \frac{1}{\Delta\sigma_{k+1}} I \right)^{-1} \left( \xi_{k+1}^{(j)} - \widetilde{G}\left(u_k^{(j)}\right) \right),$$

where $\xi_{k+1}^{(j)}$ is distributed according to $\mathrm{N}(0, \Delta\sigma_{k+1}^{-1})$ and $C_{\mathrm{up}}$, $C_{\mathrm{pp}}$ are the empirical covariance matrices

$$C_{\mathrm{pp}}(\mathbf{u_k}) = \frac{1}{J} \sum_{j=1}^{J} \left( \widetilde{G}(u_k^{(j)}) - \overline{\mathbf{G}}_{\mathbf{k}} \right) \otimes \left( \widetilde{G}(u_k^{(j)}) - \overline{\mathbf{G}}_{\mathbf{k}} \right),$$

$$C_{\mathrm{up}}(\mathbf{u_k}) = \frac{1}{J} \sum_{j=1}^{J} \left( u_k^{(j)} - \overline{\mathbf{u}}_{\mathbf{k}} \right) \otimes \left( \widetilde{G}(u_k^{(j)}) - \overline{\mathbf{G}}_{\mathbf{k}} \right).$$

# Details: EnKF with adaptive step size

Problem: How can we determine the stepsize $\Delta\sigma_{k+1}$ adaptively?

Idea: Use the adaptive tempering approach of (Papaioannou et al., 2016), (Iglesias et al., 2018) to determine the sequence $\sigma_k$.

Given $\sigma_k$, we determine $\sigma_{k+1}$ by

$$\sigma_{k+1} = \underset{\sigma \in (0,\sigma_k)}{\operatorname{argmin}} \quad \frac{1}{2}\left(\delta_{w_{k+1}} - \delta_{\text{target}}\right)^2,$$

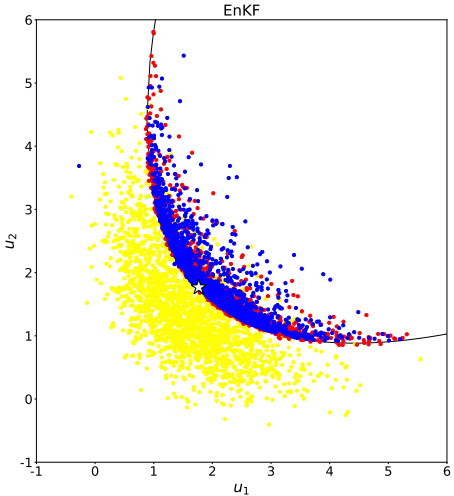
where $\delta_{w_{k+1}}$ is the coefficient of variation of the weights

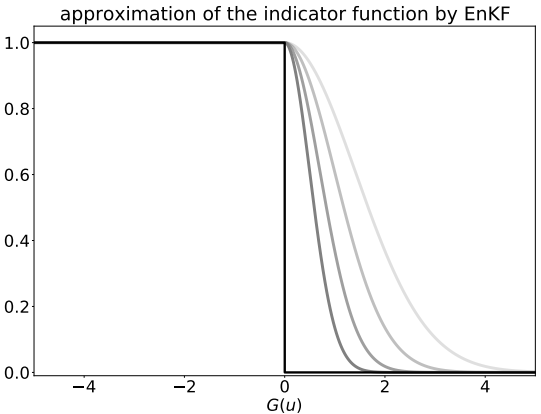
$$w_{k+1}^{(j)} = \exp\left(-\frac{1}{2}\left(\frac{1}{\sigma_{k+1}} - \frac{1}{\sigma_k}\right)\widetilde{G}(u_k^{(j)})^2\right), \qquad \text{for } j = 1, \dots, J.$$

The stepsize is defined by $\Delta\sigma_{k+1} = 1/\sigma_{k+1} - 1/\sigma_k$.

# Numerical experiment: Single-modal LSF, $n = 2$

Final ensemble of the **EnKF** for $J = 2000$ and $P_f = 4.21 \times 10^{-3}$ (left)



# Theoretical properties: Mean-field limit

**Theorem 10.1** ((Wagner, Papaioannou, et al., 2022)). *Consider* $G(u) = \boldsymbol{a}^\top \boldsymbol{u} + b$ *and assume that the EnKF is applied without noise. Then, in the large particle limit* $J \to \infty$*, the ensemble mean* $m(t)$ *converges to*

$$\lim_{t \to +\infty} m(t) = P_f u^{\mathrm{opt}} + (1 - P_f) u^{\mathrm{MLFP}},$$

*where*

- $u^{\mathrm{opt}}$ *is the mean of the optimal importance sampling density,*
- $u^{\mathrm{MLFP}}$ *is the most likely failure point (MLFP)*

$$u^{\mathrm{MLFP}} := \operatorname*{argmin}_{u \in \mathbb{R}^n} \quad \frac{1}{2}\|u\|_2^2, \quad \textit{such that} \quad G(u) = 0.$$

# Summary

- EnKF uses a sequence of auxiliary densities to approximate optimal IS density
- EnKF moves particles to the failure domain

  $\longrightarrow$ interacting particle system coupled by LSF
- Leverage analytical results for EKI

  $\longrightarrow$ nonlinear Bayesian inverse problem by construction

# Outline

# Consensus-based sampling (CBS)

(Carrillo et al., 2022)

- Terminology: $f\colon \mathbb{R}^n \to \mathbb{R}$ is a smooth energy function
- **Goal:** Generate samples of a random vector with pdf $\tau(u) \propto e^{-f(u)}$ in $\mathbb{R}^n$

**Key steps:**

1. Assume existence of Laplace approximation $N(\boldsymbol{a}, A)$ of $\tau$
2. Choose initial distribution $N(\boldsymbol{a}_0, A_0)$ close to $N(\boldsymbol{a}, A)$
3. Limit distribution $N(\boldsymbol{a}_\infty, A_\infty)$ is the equilibrium distribution of a McKean-Vlasov SDE
4. Discretize this SDE in time and probability space (“particles”)
5. Inverse temperature $\beta > 0$ controls distance of $N(\boldsymbol{a}_\infty, A_\infty)$ and $N(\boldsymbol{a}, A)$

# Consensus-based sampling - cont.

(Carrillo et al., 2022) consider the McKean–Vlasov SDE

$$
\begin{aligned}
dU_t &= (-U_t + m_\beta(\mathrm{Law}(U_t)))dt + \sqrt{2}c_\beta(\mathrm{Law}(U_t))dW_t, \\
m_\beta(\mu) &= \int_{\mathbb{R}^n} \frac{ue^{-\beta f(u)}}{Z} d\mu(u), \\
c_\beta(\mu)^2 &= (1+\beta)\left(\int_{\mathbb{R}^n} uu^T \frac{e^{-\beta f(u)}}{Z} d\mu(u) - m_\beta(\mu)m_\beta(\mu)^T\right), \\
Z &= \int_{\mathbb{R}^n} e^{-\beta f(u)} d\mu(u),
\end{aligned}
$$

where $\beta > 0$ is the inverse temperature and $\mu$ is a probability measure on $\mathbb{R}^n$.

# Consensus-based sampling - cont.

Key observation: Re-weighting of $\mathrm{Law}(U_t)$ by $\exp(-\beta f(u))$

- As $\beta$ increases the probability mass of $U_t$ concentrates around the minimizer $u^*$ of $f$ on $\mathrm{supp}(\mathrm{Law}(U_t))$. $\longrightarrow$ The drift $-U_t + m_\beta(\mathrm{Law}(U_t))$ points towards $u^*$.
- The diffusion $c_\beta(\mathrm{Law}(U_t))$ has the effect that $\mathrm{Cov}(U_t)$ approximates the curvature of $f$ at $u^*$. $\longrightarrow$ Laplace approximation of $\tau \propto e^{-f}$, e.g. (Wong, 2001; Dembo & Zeitouni, 2010)

Discretization: (Carrillo et al., 2022)

- Time $t$: Exponential Euler–Maruyama type method
- $\mathrm{Law}(U_t)$: $J$ particles (“ensemble”); empirical mean and covariance

# Consensus-based rare event estimation (CBREE)

(Althaus et al., SISC, 2024)

## Key idea

Define the smoothed optimal IS density

$$p(u,s) \propto I(G(u),s)\varphi_n(u), \quad I(u,s) = \frac{1}{2}\left(1 - \frac{su}{\sqrt{1+s^2u^2}}\right)$$

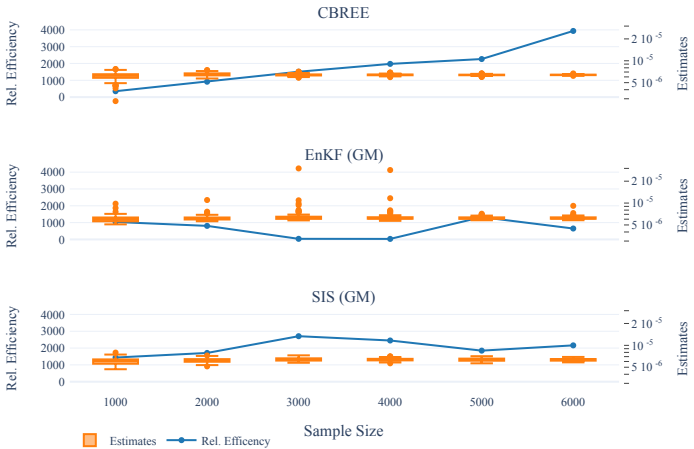
with smoothing parameter $s \geq 0$, where $p(\cdot, s) \to p_{\text{opt}}(\cdot)$ as $s \to +\infty$.

Use the energy function $f(u) = -\log p(u,s)$ in CBS.

Discretization of time and $\mathrm{Law}(U_t)$ as in (Carrillo et al., 2022); adaptive choice of inverse temperature $\beta$, smoothing parameter $s$, time step size $h$ and stopping time

# Numerical experiment, $n = 6$

Relative efficiency $P_f(1 - P_f)/(\text{MSE}(\widehat{P}_f) \times \text{cost}(\widehat{P}_f))$ of **CBREE**, SIS and EnKF for the nonlinear oscillator problem (Cheng et al., 2023) with $P_f = 6.43 \times 10^{-6}$



# Further details (Althaus et al., SISC, 2024)

- Novel adaptive choice of time step size based on exponential integrators for a proxy ODE
- Adaptive choice of inverse temperature as in (Carrillo et al., 2022)
- Adaptive choice of smoothing parameter as in (Papaioannou, Papadimitriou, et al., 2016; Wagner, Papaioannou, et al., 2022)
- More experiments with PDE-based limit state function in 1D physical space and a higher-dimensional test problem ($n = 50$)

  $\longrightarrow$ single-modal failure domains

  $\longrightarrow$ CBREE is competitive with SIS (Papaioannou, Papadimitriou, et al., 2016) and EnKF (Wagner, Papaioannou, et al., 2022) in low dimensions and for high accuracy demands

# Summary

- CBREE moves particles to failure domain using a discretized McKean–Vlasov SDE

  $\longrightarrow$ equilibrium distribution near the Laplace approx. of the smoothed optimal IS density

- CBREE implements automated parameter tuning for

  smoothing the optimal IS density

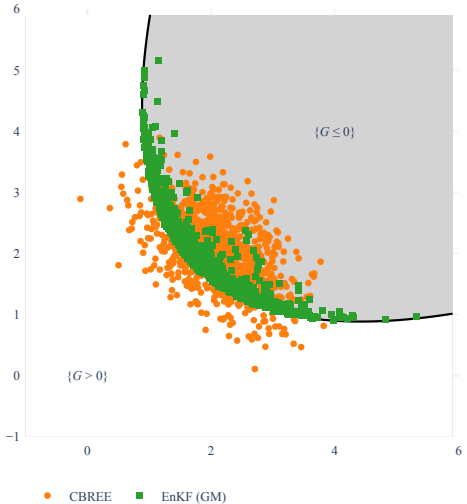
  inverse temperature in drift and diffusion coefficient

  time-step size

- Leverage analytical results for SDEs

  $\longrightarrow$ drift and diffusion coefficient not globally Lipschitz continuous

# EnKF vs CBREE



# EnKF vs CBREE: Stochastic dynamics

## EnKF

$$dU_t = -\operatorname{Cov}[U_t, \widetilde{G}(U_t)](\widetilde{G}(U)dt + dW_t)$$

$\longrightarrow$ no drift inside failure domain; scalar Wiener process $W_t$

Recall: $\widetilde{G}(u) := \max\{0, G(u)\}$, where $G$ is the limit-state function

## CBREE

$$dU_t = (-U_t + m_\beta(\mu_t))dt + \sqrt{2}c_\beta(\mu_t)dW_t, \quad \mu_t = \operatorname{Law}(U_t)$$

$\longrightarrow$ drift to minimizer of $f(u) = -\log I(G(u), s)\varphi_n(u)$; vector-valued Wiener process $W_t$

# Papers, code


- F. Wagner, I. Papaioannou, EU: *The ensemble Kalman filter for rare event estimation*. SIAM/ASA J. Uncert. Quantif., Vol. 10(1), 2022, pp. 317–349, arxiv:2106.10062

- K. Althaus, I. Papaioannou, EU: *Consensus-based rare event estimation*. SIAM J. Sci. Comput., Vol. 46(3), 2024, pp. A1487–A1513, arxiv:2304.09077.


Code and data available at

`https://github.com/AlthausKonstantin/rareeventestimation`.

# References I


- Althaus, K., I. Papaioannou, and E. Ullmann (2024). “Consensus-Based Rare Event Estimation”. In: *SIAM Journal on Scientific Computing* 46.3, A1487–A1513.
- Au, S.-K. and J. L. Beck (2001). “Estimation of small failure probabilities in high dimensions by subset simulation”. In: *Probabilistic Engineering Mechanics* 16.4, pp. 263–277.
- Boer, P.-T. de et al. (2005). “A Tutorial on the Cross-Entropy Method”. In: *Annals of Operations Research* 134.1, pp. 19–67.
- Botev, Z. I. and D. P. Kroese (2010). “Efficient Monte Carlo Simulation via the Generalized Splitting Method”. In: *Statistics and Computing* 22.1, pp. 1–16.
- Botev, Z. I. and D. P. Kroese (2012). “Efficient Monte Carlo simulation via the generalized splitting method”. In: *Stat. Comput.* 22.1, pp. 1–16.
- Breitung, K. (1984). “Asymptotic approximations for multinormal integrals”. In: *J. Eng. Mech.* 110, pp. 357–366.

# References II


- Carrillo, J. A. et al. (2022). “Consensus-based sampling”. In: *Stud. Appl. Math.* 148.3, pp. 1069–1140.
- Cérou, F. and A. Guyader (2007). “Adaptive Multilevel Splitting for Rare Event Analysis”. In: *Stochastic Analysis and Applications* 25.2, pp. 417–443.
- Cheng, K. et al. (2023). “Rare event estimation with sequential directional importance sampling”. In: *Structural Safety* 100, p. 102291.
- Dembo, A. and O. Zeitouni (2010). *Large deviations techniques and applications*. Vol. 38. Stochastic Modelling and Applied Probability. Corrected reprint of the second (1998) edition. Springer-Verlag, Berlin, pp. xvi+396.
- Ditlevsen, P. and S. Ditlevsen (2023). “Warning of a forthcoming collapse of the Atlantic meridional overturning circulation”. In: *Nature Communications* 14.1.
- Elfverson, D., F. Hellman, and A. Målqvist (2016). “A multilevel Monte Carlo method for computing failure probabilities”. In: *SIAM/ASA J. Uncertain. Quantif.* 4.1, pp. 312–330.

# References III


- Elfverson, D., R. Scheichl, et al. (2022). *Adaptive multilevel subset simulation with selective refinement*.
- Geyer, S., I. Papaioannou, and D. Straub (2019). “Cross entropy-based importance sampling using Gaussian densities revisited”. In: *Structural Safety* 76, pp. 15–27.
- Gilbert, A. D., F. Y. Kuo, and I. H. Sloan (2023). “Analysis of Preintegration Followed by Quasi–Monte Carlo Integration for Distribution Functions and Densities”. In: *SIAM Journal on Numerical Analysis* 61.1, pp. 135–166.
- Gobet, E. and G. Liu (2015). “Rare event simulation using reversible shaking transformations”. In: *SIAM J. Sci. Comput.* 37.5, A2295–A2316.
- Griewank, A. et al. (2018). “High dimensional integration of kinks and jumps—Smoothing by preintegration”. In: *Journal of Computational and Applied Mathematics* 344, pp. 259–274.

# References IV


- Guyader, A., N. Hengartner, and E. Matzner-Løber (2011). “Simulation and estimation of extreme quantiles and extreme probabilities”. In: *Appl. Math. Optim.* 64.2, pp. 171–196.
- Haji-Ali, A.-L., J. Spence, and A. L. Teckentrup (2022). “Adaptive multilevel Monte Carlo for probabilities”. In: *SIAM J. Numer. Anal.* 60.4, pp. 2125–2149.
- Hasofer, A. M. and N. Lind (1974). “An exact and invariant first order reliability format”. In: *J. Eng. Mech.* 100, pp. 111–121.
- Hastings, W. K. (1970). “Monte Carlo sampling methods using Markov chains and their applications”. In: *Biometrika* 57.1, pp. 97–109.
- Iglesias, M., M. Park, and M. V. Tretyakov (2018). “Bayesian inversion in resin transfer molding”. In: *Inverse Problems* 34.10, pp. 105002, 46.
- Iglesias, M. A., K. J. H. Law, and A. M. Stuart (2013). “Ensemble Kalman methods for inverse problems”. In: *Inverse Problems* 29.4, pp. 045001, 20.

# References V


- Kong, A. (1992). *A note on importance sampling using standardized weights*. Tech. rep. 348. Chicago, IL: Dept. Statistics, Univ. Chicago.
- Koutsourelakis, P., H. Pradlwarter, and G. Schuëller (2004). “Reliability of structures in high dimensions, part I: algorithms and applications”. In: *Probabilistic Engineering Mechanics* 19.4, pp. 409–417.
- Liu, S. and A. B. Owen (2023). “Preintegration via Active Subspace”. In: *SIAM Journal on Numerical Analysis* 61.2, pp. 495–514.
- Metropolis, N. et al. (1953). “Equation of State Calculations by Fast Computing Machines”. In: *The Journal of Chemical Physics* 21.6, pp. 1087–1092.
- Morio, J. and M. Balesdent (2015). *Estimation of rare event probabilities in complex aerospace and other systems*. Cambridge, England: Woodhead Publishing.
- Owen, A. B. (2013). *Monte Carlo theory, methods and examples*. https://artowen.su.domains/mc/.

# References VI


- Papaioannou, I., W. Betz, et al. (2015). “MCMC algorithms for Subset Simulation”. In: *Probabilistic Engineering Mechanics* 41, pp. 89–103.
- Papaioannou, I., C. Papadimitriou, and D. Straub (2016). “Sequential importance sampling for structural reliability analysis”. In: *Structural Safety* 62, pp. 66–75.
- Peherstorfer, B., B. Kramer, and K. Willcox (2018). “Multifidelity preconditioning of the cross-entropy method for rare event simulation and failure probability estimation”. In: *SIAM/ASA J. Uncertain. Quantif.* 6.2, pp. 737–761.
- Proppe, C. (2020). “A multilevel moving particles method for reliability estimation”. In: *Probabilistic Engineering Mechanics* 59, p. 103018.
- Reich, S. and S. Weissmann (2021). “Fokker-Planck particle systems for Bayesian inference: computational approaches”. In: *SIAM/ASA J. Uncertain. Quantif.* 9.2, pp. 446–482.

# References VII


- Rubinstein, R. Y. (1997). “Optimization of computer simulation models with rare events”. In: *European Journal of Operational Research* 99.1, pp. 89–112.
- Rubinstein, R. Y. and D. P. Kroese (2016). *Simulation and the Monte Carlo Method*. Wiley.
- Schillings, C. and A. M. Stuart (2017). “Analysis of the ensemble Kalman filter for inverse problems”. In: *SIAM J. Numer. Anal.* 55.3, pp. 1264–1290.
- Tong, S. and G. Stadler (2022). *Large deviation theory-based adaptive importance sampling for rare events in high dimensions*. Tech. rep. available at arXiv:2209.06278.
- Tong, S., E. Vanden-Eijnden, and G. Stadler (2021). “Extreme Event Probability Estimation Using PDE-Constrained Optimization And Large Deviation Theory, With Application To Tsunamis”. In: *Commun. Appl. Math. Comput. Sci.* 16.2, pp. 181–225.
- Ullmann, E. and I. Papaioannou (2015). “Multilevel estimation of rare events”. In: *SIAM/ASA J. Uncertain. Quantif.* 3.1, pp. 922–953.

# References VIII


- Vexler, B. (2004). “Adaptive Finite Element Methods for Parameter Identification Problems”. PhD thesis. Heidelberg University.
- Wagner, F., J. Latz, et al. (2020). “Multilevel Sequential Importance Sampling for Rare Event Estimation”. In: *SIAM Journal on Scientific Computing* 42.4, A2062–A2087.
- — (2021). “Error analysis for probabilities of rare events with approximate models”. In: *SIAM Journal on Numerical Analysis* 59.4, pp. 1948–1975.
- Wagner, F., I. Papaioannou, and E. Ullmann (2022). “The ensemble Kalman filter for rare event estimation”. In: *SIAM/ASA Journal on Uncertainty Quantification* 10.1, pp. 317–349.
- Wong, R. (2001). *Asymptotic Approximations of Integrals*. Society for Industrial and Applied Mathematics.