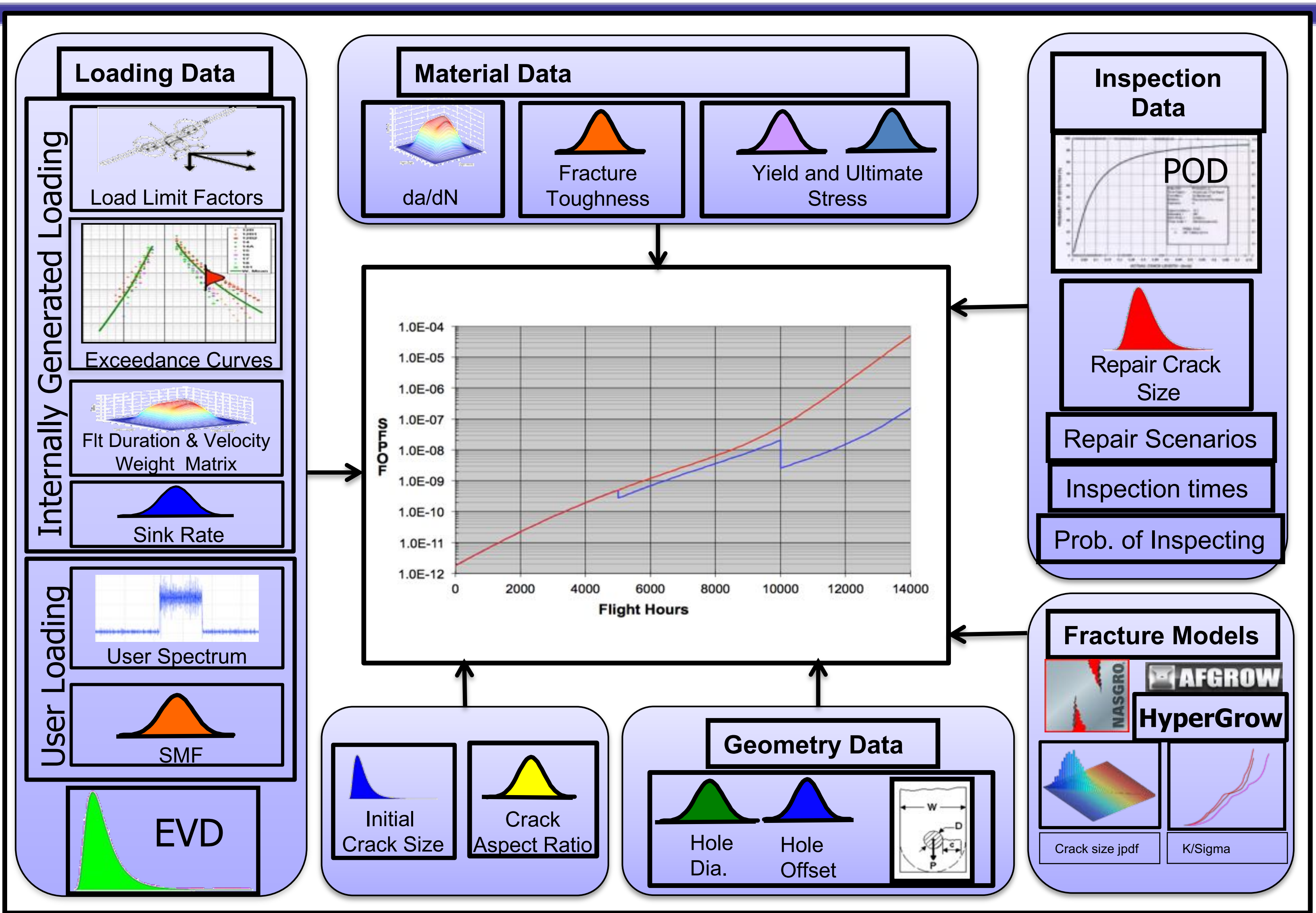
Adaptive Importance Sampling for Probabilistic Damage Tolerance Analysis

Nathan Crosby
Harry Millwater
University of Texas at San Antonio

AA&S 2020
VIRTUAL
AUGUST 25-28, 2020

SMART|DT

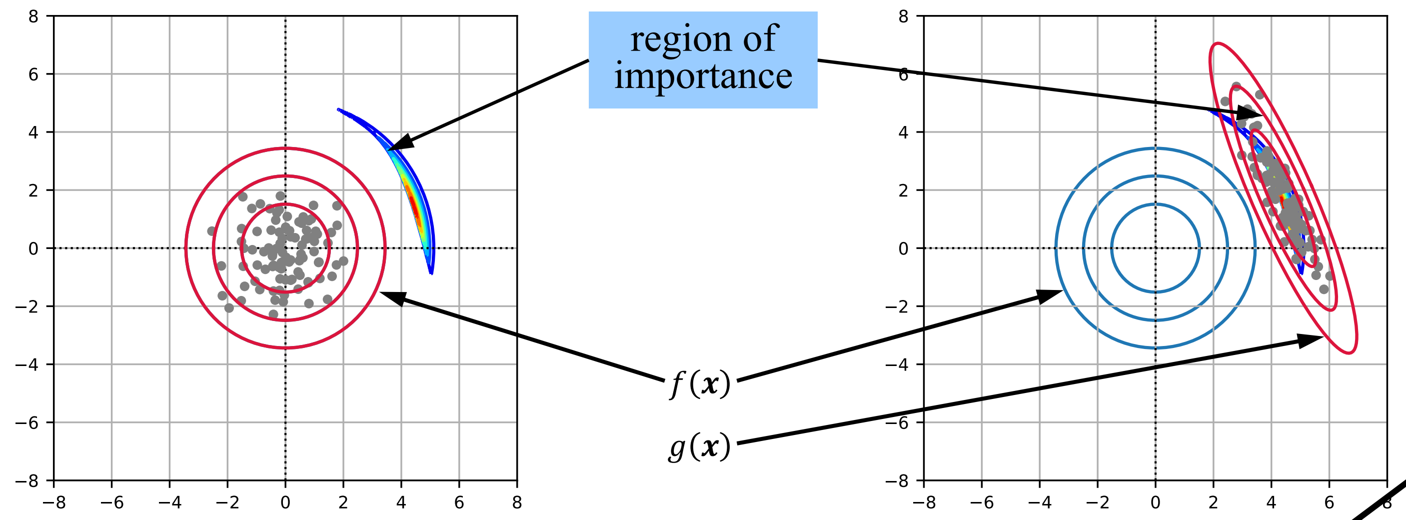


Loading Data
Internally Generated Loading
Load Limit Factors
Exceedance Curves
Flt Duration & Velocity
Weight Matrix
Sink Rate
User Loading
User Spectrum
SMF
EVD
Material Data
da/dN
Fracture
Toughness
Yield and Ultimate
Stress
Inspection
Data
POD
Repair Crack
Size
Repair Scenarios
Inspection times
Prob. of Inspecting
SFPOF
1.0E-04
1.0E-05
1.0E-06
1.0E-07
1.0E-08
1.0E-09
1.0E-10
1.0E-11
1.0E-12
0
2000
4000
6000
8000
10000
12000
14000
Flight Hours
Fracture Models
NASGRO
AFGROW
HyperGrow
Crack size jpdf
K/Sigma
Initial
Crack Size
Crack
Aspect Ratio
Geometry Data
Hole
Dia.
Hole
Offset

Overview

- Importance Sampling
- Adaptive Importance Sampling
- Cross Entropy (CE) Method
- Application to PDTA
- Examples
- Conclusion

Importance Sampling



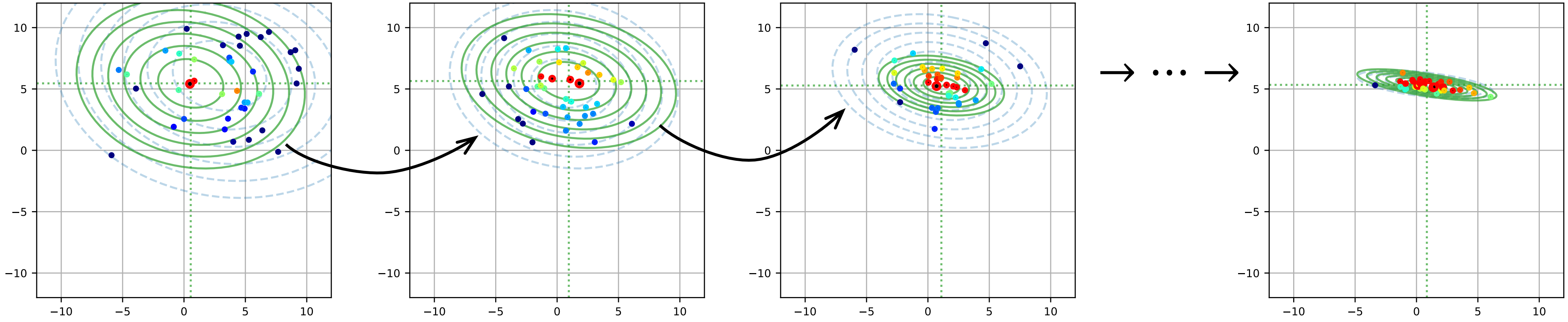
For illustration, $H(\boldsymbol{x};t)$ is the conditional expectation that $\sigma_{\mathrm{RS}}(\boldsymbol{x};t) < \sigma_{\mathrm{EVD}}(t)$:

$$H(\boldsymbol{x};t) = 1 - F_{\mathrm{EVD}}\big(\sigma_{\mathrm{RS}}(\boldsymbol{x};t)\big)$$

Standard Monte Carlo Sampling	Importance Sampling	
$\mathbb{E}[H(\boldsymbol{x};t)] = \int H(\boldsymbol{x})\, f(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x}$	$\mathbb{E}[H(\boldsymbol{x})] = \int H(\boldsymbol{x};t) \frac{f(\boldsymbol{x})}{g(\boldsymbol{x})} g(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x}$	Expectation integral for $H(\boldsymbol{x};t)$, not analytically solvable
$\hat{\mu} \approx \frac{1}{N}\sum_i H(\boldsymbol{x}_i;t)$	$\hat{\mu} \approx \frac{1}{N}\sum_i H(\boldsymbol{x}_i;t)\frac{f(\boldsymbol{x}_i)}{g(\boldsymbol{x}_i)}$	Sampling estimator of expected value of $H(\boldsymbol{x};t)$
$\mathrm{Var}(\hat{\mu}) \approx \frac{1}{N^2}\sum_i (H(\boldsymbol{x}_i;t) - \hat{\mu})^2$	$\mathrm{Var}(\hat{\mu}) \approx \frac{1}{N^2}\sum_i \left(H(\boldsymbol{x};t)\frac{f(\boldsymbol{x}_i)}{g(\boldsymbol{x}_i)} - \hat{\mu}\right)^2$	Variance of the estimator of expected value of $H(\boldsymbol{x};t)$

Likelihood Ratio: $\frac{f(\boldsymbol{x})}{g(\boldsymbol{x})}$

Adaptive Importance Sampling



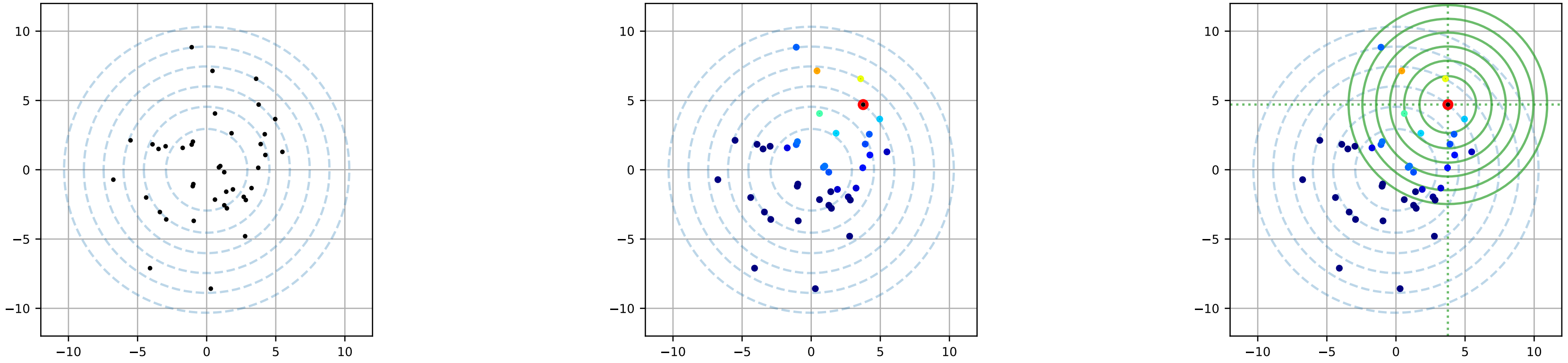
- Adaptive sampling splits the estimation problem into 2 tasks:
  - Find the optimal sampling density for $P_f(t_i)$
  - Estimate $P_f(t_i)$ using the optimal sampling density
- Cross-Entropy (CE) Method – find the optimal sampling density
  - Estimating the optimal sampling density
  - Kullback-Leibler Divergence

Estimating the Optimal Sampling Density

Generate Samples from $g_{(j)}(\boldsymbol{x})$

Estimate $g_{\text{opt}}(\boldsymbol{x})$

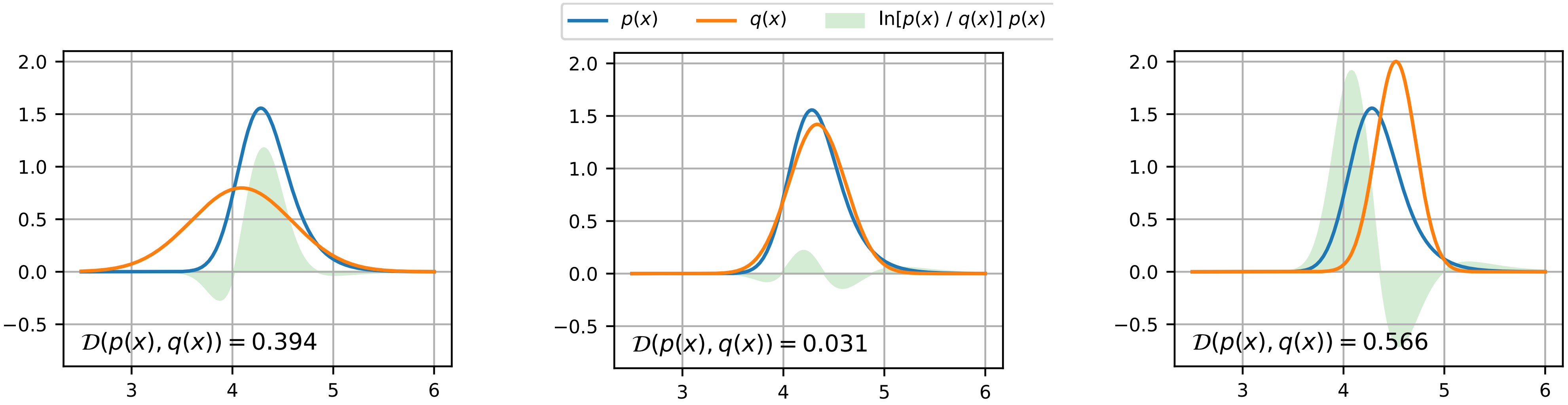
Find $g_{(j+1)}(\boldsymbol{x})$ that most closely matches $\hat{g}_{\text{opt}}(\boldsymbol{x}_i)$



- The estimator variance, $\frac{1}{N^2}\sum_i \left(\frac{H(\boldsymbol{x}_i)\,f(\boldsymbol{x}_i)}{g(\boldsymbol{x}_i)} - \mu\right)^2$, is minimized when $\frac{H(\boldsymbol{x}_i)\,f(\boldsymbol{x}_i)}{g(\boldsymbol{x}_i)} - \mu = 0$
- This implies the optimal sampling density is $g_{\text{opt}}(\boldsymbol{x}) = \frac{H(\boldsymbol{x})\,f(\boldsymbol{x})}{\mu}$
- μ is unknown, but can be estimated from the samples: $\hat{\mu} = \frac{1}{N}\sum_i \frac{H(\boldsymbol{x}_i)\,f(\boldsymbol{x}_i)}{g(\boldsymbol{x}_i)}$
- Leads to an estimated optimal sampling density: $\hat{g}_{\text{opt}}(\boldsymbol{x}_i) = \frac{H(\boldsymbol{x}_i)\,f(\boldsymbol{x}_i)}{\hat{\mu}}$

Kullback-Leibler Divergence

$$\mathcal{D}\big(p(\boldsymbol{x}), q(\boldsymbol{x})\big) = \int \ln\left(\frac{p(\boldsymbol{x})}{q(\boldsymbol{x})}\right) p(\boldsymbol{x})\, dx = \int \ln\big(p(\boldsymbol{x})\big)\, p(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} - \int \ln\big(q(\boldsymbol{x})\big)\, p(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x}$$



- $\mathcal{D}$ is a metric – a measure of difference between two PDFs
  - $\mathcal{D} \geq 0$, $\mathcal{D} = 0$ when the PDFs are identical

CE Method

Fixed reference density

Parametric density

$$\mathcal{D}\left(\hat{g}_{\text{opt}}(\boldsymbol{x}), f(\boldsymbol{x};\boldsymbol{v})\right) = \underbrace{\int \ln\left(\hat{g}_{\text{opt}}(\boldsymbol{x})\right) \hat{g}_{\text{opt}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}}_{\text{constant}} - \underbrace{\int \ln\left(f(\boldsymbol{x};\boldsymbol{v})\right) \hat{g}_{\text{opt}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}}_{\text{cross entropy}}$$

Optimal sampling density parameters

Drop additive constant term (1st integral), drop minus sign of cross entropy term (2nd integral)

Drop normalizing constant (denominator) of $\hat{g}_{\text{opt}}(\boldsymbol{x})$, original parametric density is $f(\boldsymbol{x};\boldsymbol{u})$

Introduce importance sampling density with parameters $\boldsymbol{w}$ (from previous iteration)

$$\begin{aligned}
\boldsymbol{v}_{\text{opt}} &= \underset{\boldsymbol{v}}{\operatorname{argmin}} \left(\mathcal{D}\left(\hat{g}_{\text{opt}}(\boldsymbol{x}), f(\boldsymbol{x};\boldsymbol{v})\right)\right) \\
&= \underset{\boldsymbol{v}}{\operatorname{argmax}} \left(\int \ln\left(f(\boldsymbol{x};\boldsymbol{v})\right) \hat{g}_{\text{opt}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\right) \\
&= \underset{\boldsymbol{v}}{\operatorname{argmax}} \left(\int \ln\left(f(\boldsymbol{x};\boldsymbol{v})\right) H(\boldsymbol{x};t)\, f(\boldsymbol{x};\boldsymbol{u})\,\mathrm{d}\boldsymbol{x}\right) \\
&= \underset{\boldsymbol{v}}{\operatorname{argmax}} \left(\int \ln\left(f(\boldsymbol{x};\boldsymbol{v})\right) H(\boldsymbol{x};t)\, \frac{f(\boldsymbol{x};\boldsymbol{u})}{f(\boldsymbol{x};\boldsymbol{w})} f(\boldsymbol{x};\boldsymbol{w})\,\mathrm{d}\boldsymbol{x}\right) \\
&= \underset{\boldsymbol{v}}{\operatorname{argmax}}\, \mathbb{E}_{\boldsymbol{w}}\left[\ln\left(f(\boldsymbol{x};\boldsymbol{v})\right) H(\boldsymbol{x};t)\, \frac{f(\boldsymbol{x};\boldsymbol{u})}{f(\boldsymbol{x};\boldsymbol{w})}\right]
\end{aligned}$$

- minimizes the difference between a parametric sampling density, $f(\boldsymbol{x};\boldsymbol{v})$, and the estimated optimal sampling density, $\hat{g}_{\text{opt}}(\boldsymbol{x})$
- by finding parameters, $\boldsymbol{v}_{\text{opt}}$, that minimize $\mathcal{D}\left(\hat{g}_{\text{opt}}(\boldsymbol{x}), f(\boldsymbol{x};\boldsymbol{v})\right)$

CE Method – Solving the Parameter Optimization Problem

- Sampling estimator

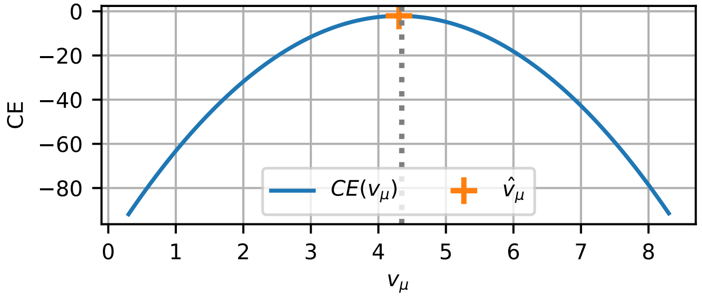
$$\boldsymbol{v}_{\text{opt}} = \max_{\boldsymbol{v}} \left(\frac{1}{N} \sum_{\boldsymbol{i}} H(\boldsymbol{x}_i; t) \frac{f(\boldsymbol{x}_i; \boldsymbol{u})}{f(\boldsymbol{x}_i; \boldsymbol{w})} \ln\big(f(\boldsymbol{x}_i; \boldsymbol{v})\big) \right)$$

- Take the derivative with respect to and solve the resulting system of equations

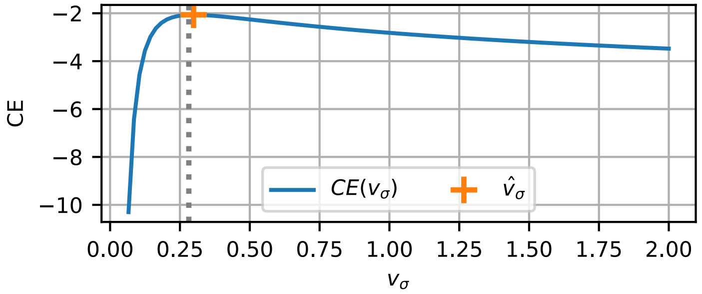
$$\frac{1}{N} \sum_{i} H(\boldsymbol{x}_i; t) \frac{f(\boldsymbol{x}_i; \boldsymbol{u})}{f(\boldsymbol{x}_i; \boldsymbol{w})} \nabla_{\boldsymbol{v}} \ln\big(f(\boldsymbol{x}_i; \boldsymbol{v})\big) = 0$$

- Natural Exponential Family distributions have closed form solutions
  - Exponential
  - Normal
  - Weibull
  - Gamma

Multivariate Normal Solution



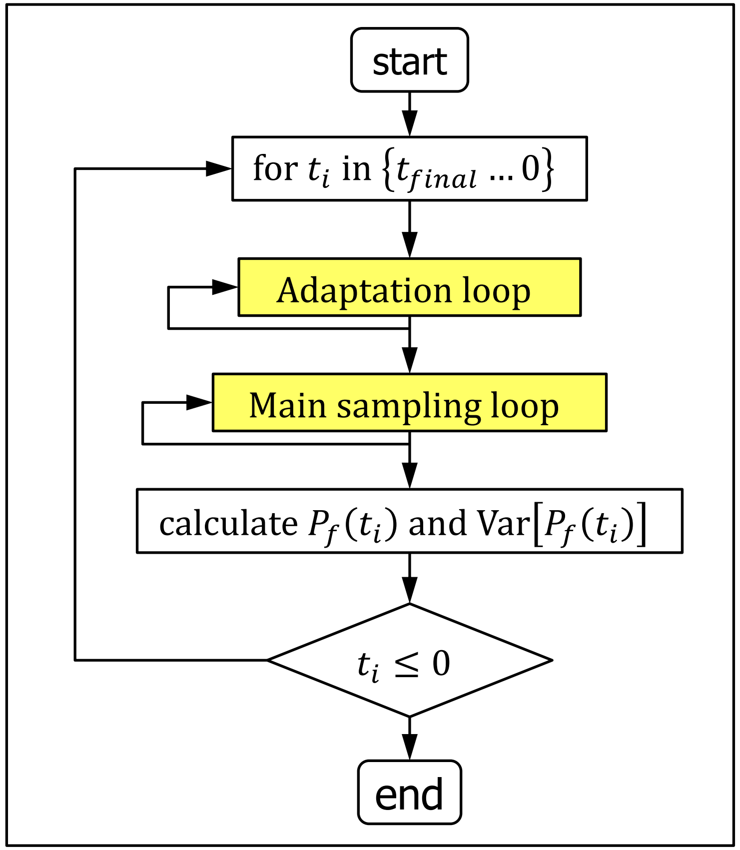
$$\widehat{\boldsymbol{v}}_\mu = \frac{\sum_i H(\boldsymbol{x}_i; t) \frac{f(\boldsymbol{x}_i; \boldsymbol{u})}{f(\boldsymbol{x}_i; \boldsymbol{w})} \boldsymbol{x}_i}{\sum_i H(\boldsymbol{x}_i; t) \frac{f(\boldsymbol{x}_i; \boldsymbol{u})}{f(\boldsymbol{x}_i; \boldsymbol{w})}}$$



$$\widehat{\boldsymbol{v}}_\Sigma = \frac{\sum_i (\boldsymbol{x}_i - \widehat{\boldsymbol{v}}_\mu) \left(H(\boldsymbol{x}_i; t) \frac{f(\boldsymbol{x}_i; \boldsymbol{u})}{f(\boldsymbol{x}_i; \boldsymbol{w})} \right) (\boldsymbol{x}_i - \widehat{\boldsymbol{v}}_\mu)^{\mathrm{T}}}{\sum_i H(\boldsymbol{x}_i; t) \frac{f(\boldsymbol{x}_i; \boldsymbol{u})}{f(\boldsymbol{x}_i; \boldsymbol{w})}}$$

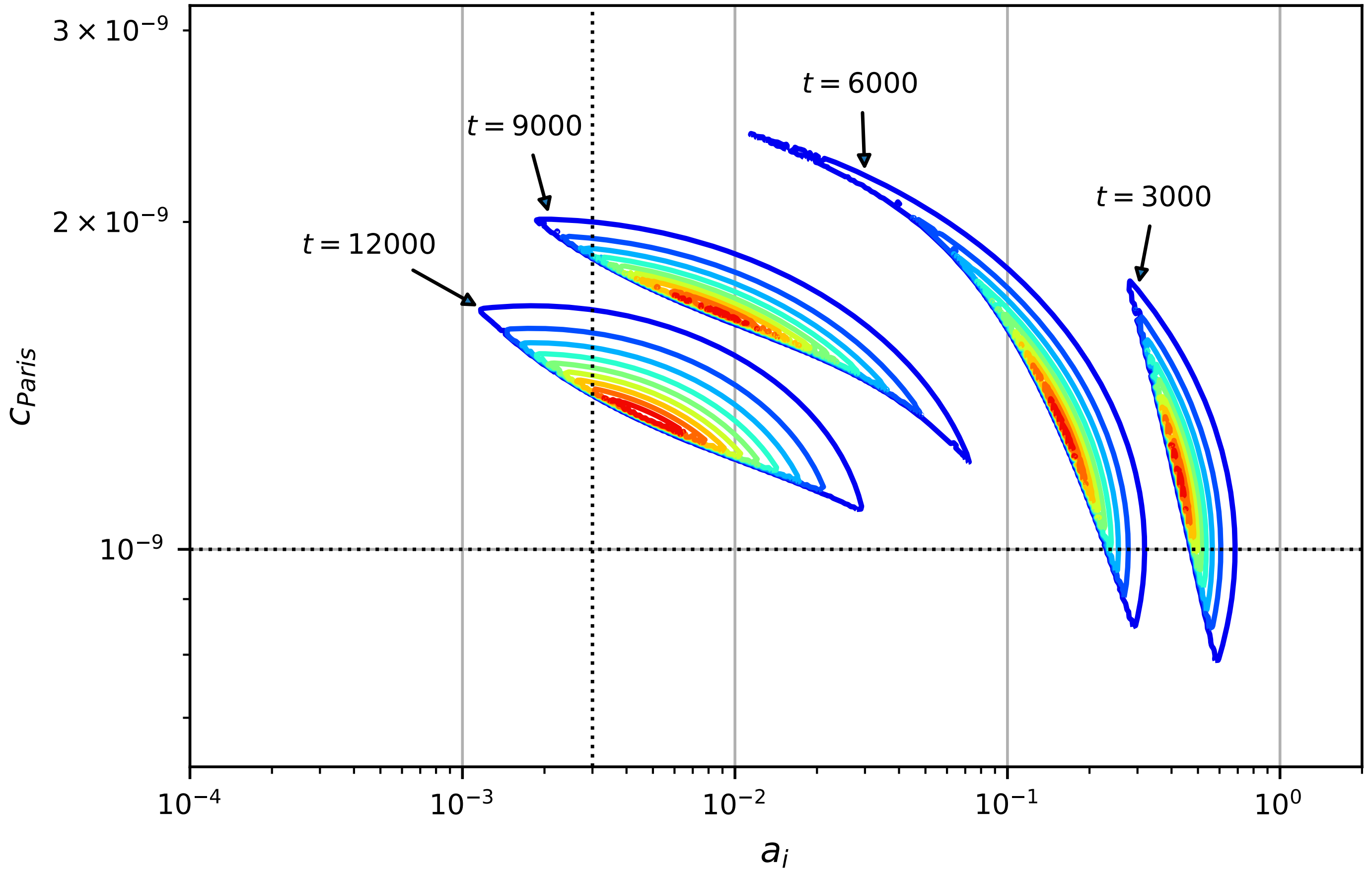
Application to PDTA

Evaluate $P_f(\{0 \ldots t_{final}\})$



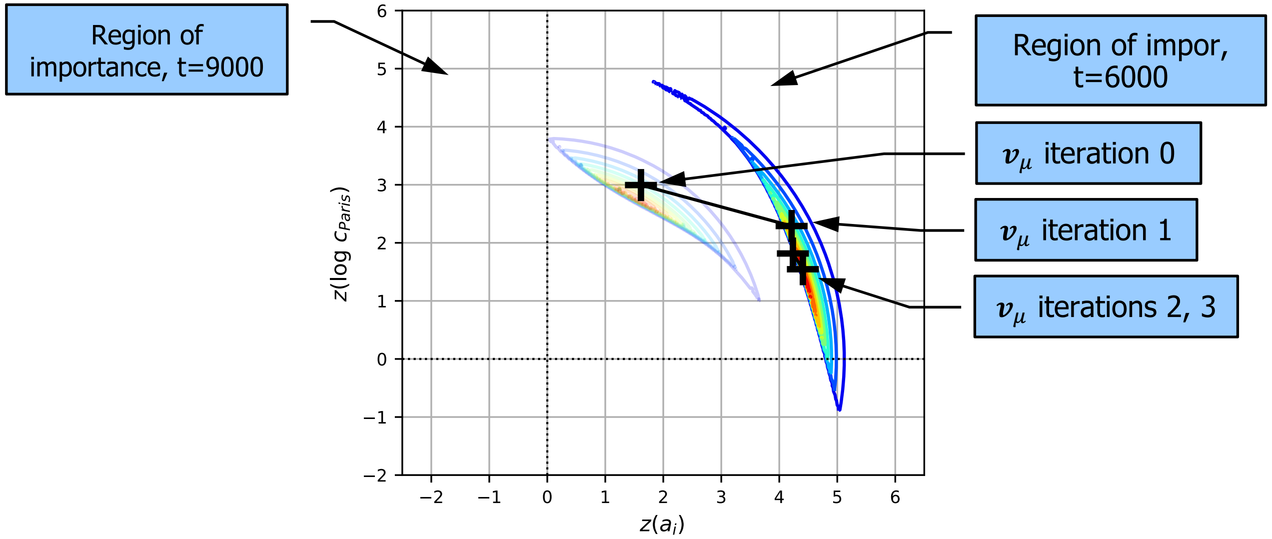
- Algorithm overview
  - Adaptation loop iterates until the sampling density converges to the optimal sampling density
    - n_{adp} is the number of samples per loop
  - Main sampling loop iterates until the covariance threshold is met
    - ϵ_{COV} covariance threshold
    - n_{main} number of samples per loop

PDTA Challenges



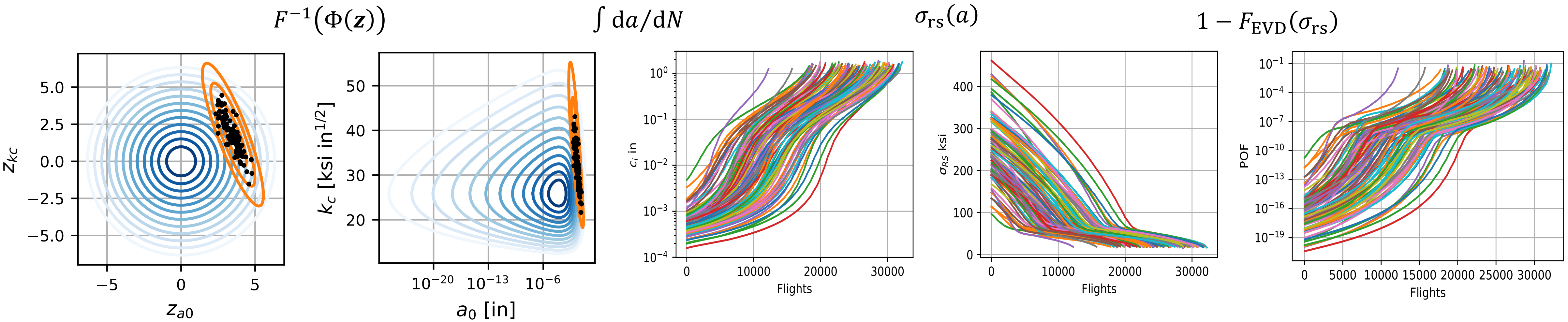
- Region of interest changes with t
- Variety of probability distributions with large differences in magnitude
- Convergence control

Changing Region of Importance



- Warm start: start the adaptation process for t_{i+1} using the optimal v_μ from t_i
  - reset v_Σ to I
  - reduces adaptation loop iterations ~10-20% fewer samples
- Increases chance of a sample in the new region of importance
  - Improves convergence with small sample sizes
- Start from t_{final} and work back to 0
  - The sampling density at t_{final} is closest to the nominal density mean

Handling Mixture of Input Density Functions

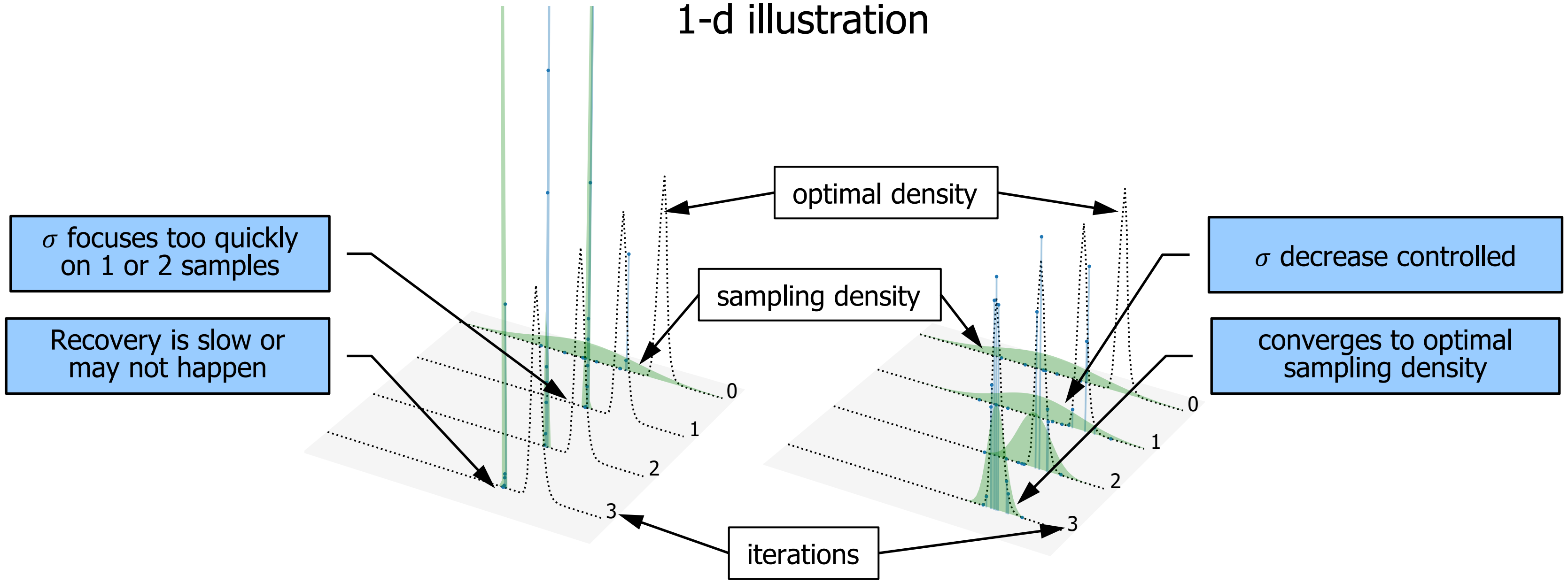


- The Transform Likelihood Ratio method is used to simplify working with multiple distributions
  - Samples are generated in standard multivariate normal space
  - Nominal variable density is independent standard normal
  - Sampling density is multivariate normal
  - Estimated optimal sampling density, likelihood ratio, and parameter updates use normal space
  - Samples are transformed to original space using inverse CDF transform (Nataf for correlated variables)

Kroese, D.P. and Rubenstein, R.Y., The transform likelihood ratio method for rare event simulation with heavy tails, Queueing Systems, 2004

Convergence Control

1-d illustration

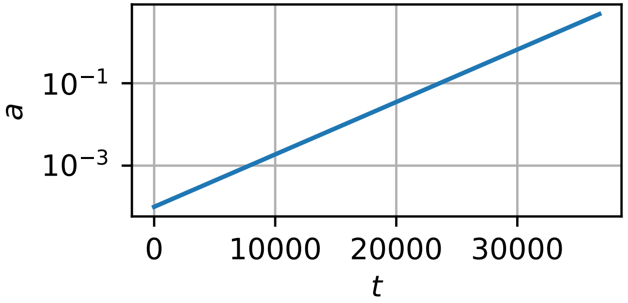


- Small sample size increases likelihood of just 1 or 2 samples outweighing
- Algorithm has to control contraction of the sampling density covariance
  - While in the adaptation loop, inflate the covariance by a factor in the range 1.6 to 3
  - Prevent large changes until the mean starts to converge at the new optimal sampling density
- Smoothing
  - Weights parameter changes so that they do not jump too far from a good value due to a bad sample set
  - $\boldsymbol{v}_{new} = \alpha\, \boldsymbol{v}_{CE} + (1-\alpha)\, \boldsymbol{v}_{old}$, where
    - α is a weight between 0 and 1 (0.8)
    - $\boldsymbol{v}_{CE}$ is the CE solution for the current estimated optimal sampling density

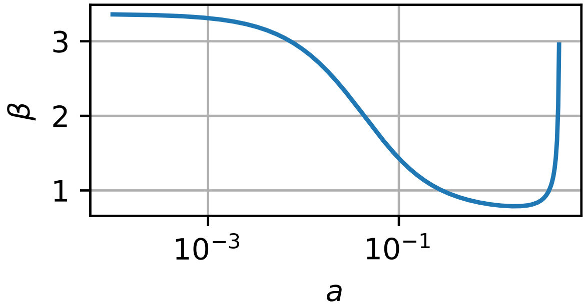
Examples

- Handbook Problem Example
- General Aviation Example
- NASGRO Example

Handbook Problem



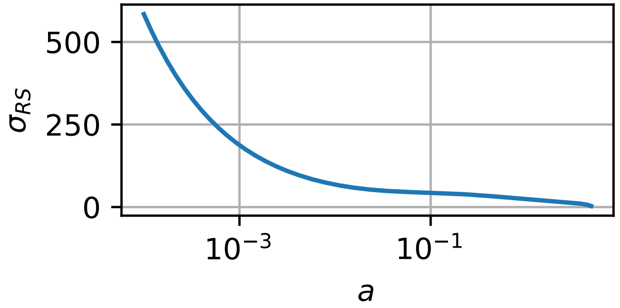
$$a(t) = a_0 \cdot \exp(2.93\times10^{-4}\, t)$$



$$\beta(a) = \beta_{\text{hole}}(a) \cdot \beta_{\text{width}}(a)$$

$$\beta_{\text{hole}}(a) = 0.6762 + \frac{0.8734}{0.3254 + a\,/\,R}$$

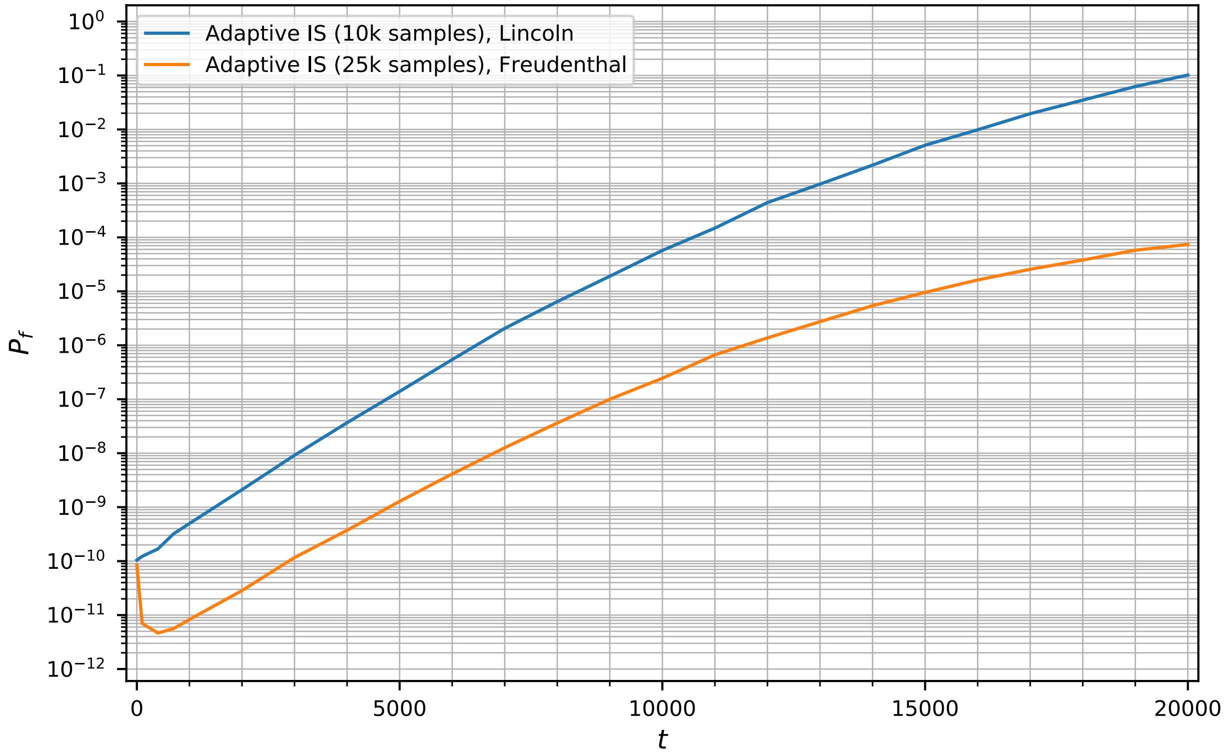
$$\beta_{\text{width}}(a) = \sqrt{\sec(\pi\,(R + a)\,/\,W)}$$

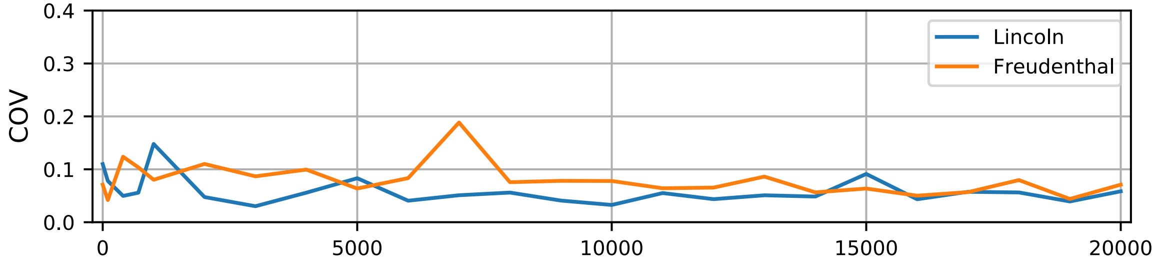


$$\sigma_{\text{rs}}(a) = K_c\,/\,(\beta(a)\,\sqrt{\pi\, a})$$

Parameter	Value
Width	Deterministic 10 in
Thickness	Deterministic 0.125 in
Initial Crack Size	$LN(0.0032, 0.0047)$ in
Fracture Toughness	$N(34.8, 3.90)$ ksi $\sqrt{\text{in}}$
Hole Diameter	Deterministic 0.25 in
Maximum Stress per Flight	$W(5.0, 10.0, 5.0)$ ksi

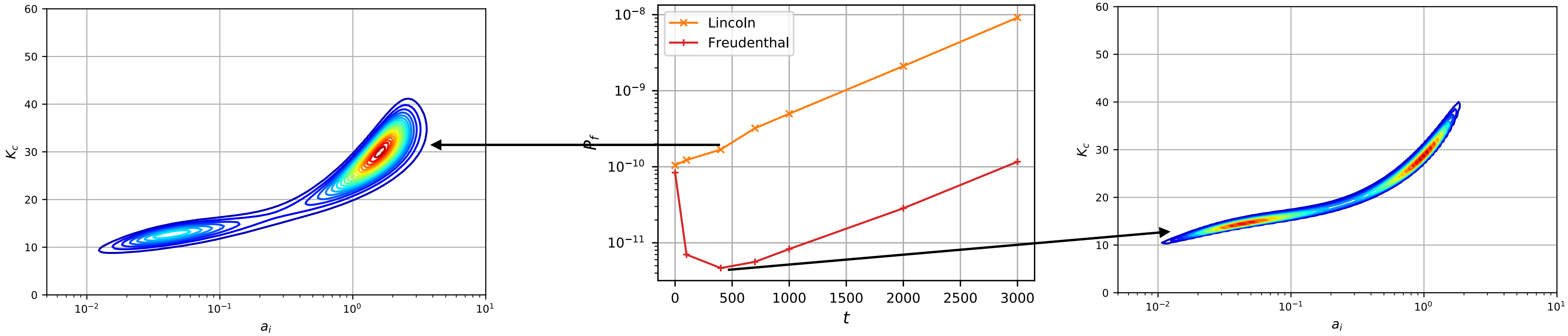
Handbook Problem POF





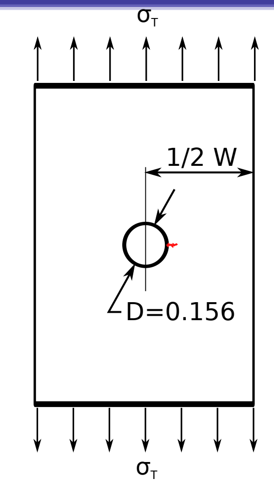
- Adaptive importance sampling parameters
  - $\epsilon_{\mathrm{COV}} = 0.2$
  - $n_{\mathrm{main}} = 100$
  - Adaptation samples
    - Lincoln: $n_{\mathrm{adp}} = 20$
    - Freudenthal: $n_{\mathrm{adp}} = 60$
- Runtimes (serial)
  - Lincoln: 2 seconds
  - Freudenthal: 7 seconds

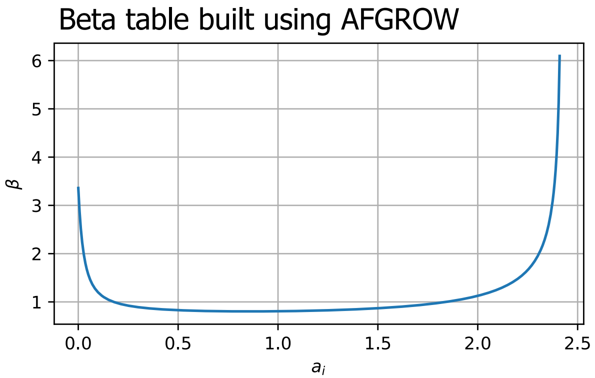
Handbook Problem POF Integrand Contours

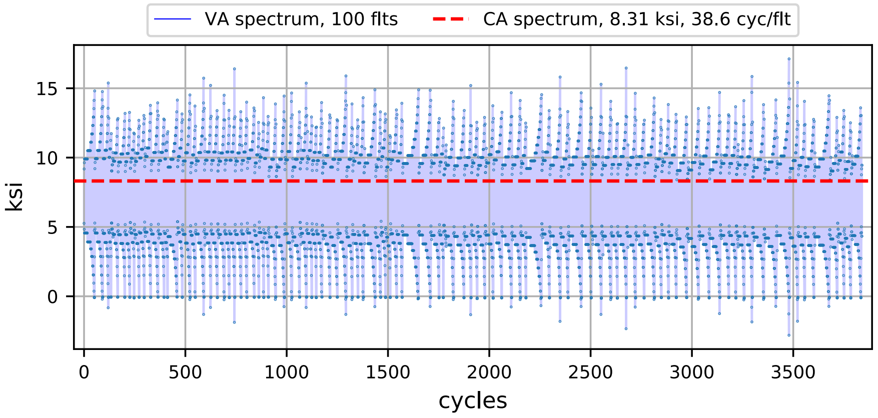


- Freudenthal region of importance is much smaller
  - More susceptible to unlucky sets of samples that miss the region of importance even when the sampling density is close to the optimal density
  - The sampling density will not conform to a long and narrow shape, so the main sampling will have higher variance

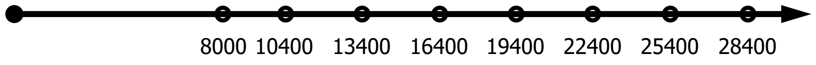
General Aviation Example Problem





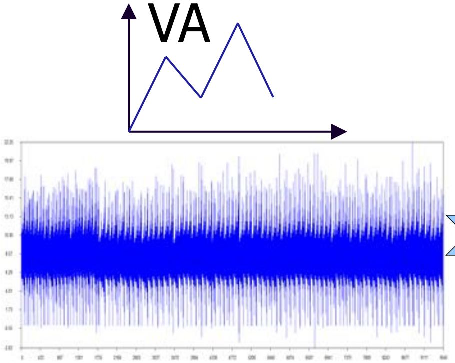


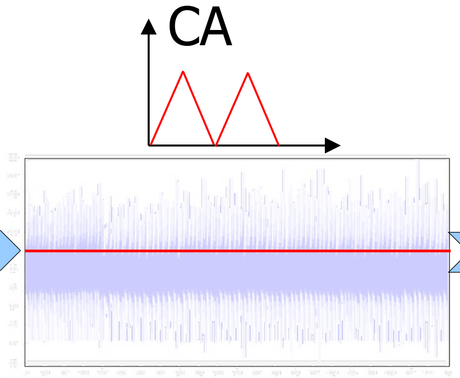
Inspection Schedule



Parameter	Values
Width	Deterministic 5 in
Thickness	Deterministic 0.125 in
Log Paris Constant	$N(-9.0, 0.08)$
Paris Exponent	Deterministic 3.8
Initial Crack Size	$W(0.45, 4.17\times10^{-5})$ in
Fracture Toughness	$N(35.0, 3.5)$ ksi $\sqrt{\text{in}}$
Maximum Stress per Flight	$EVD(13.4, 1.3, 0.07)$ ksi
Probability of Detection	$LN(0.05, 0.065)$ in
Repair Quality (Crack Size)	Perfect

Hypergrow (SMART|DT Internal CG Code)





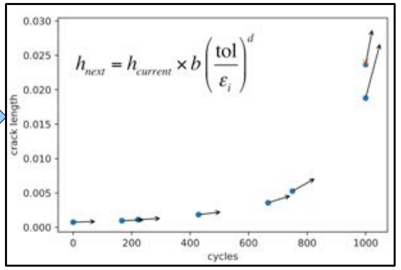
ODE Formulation

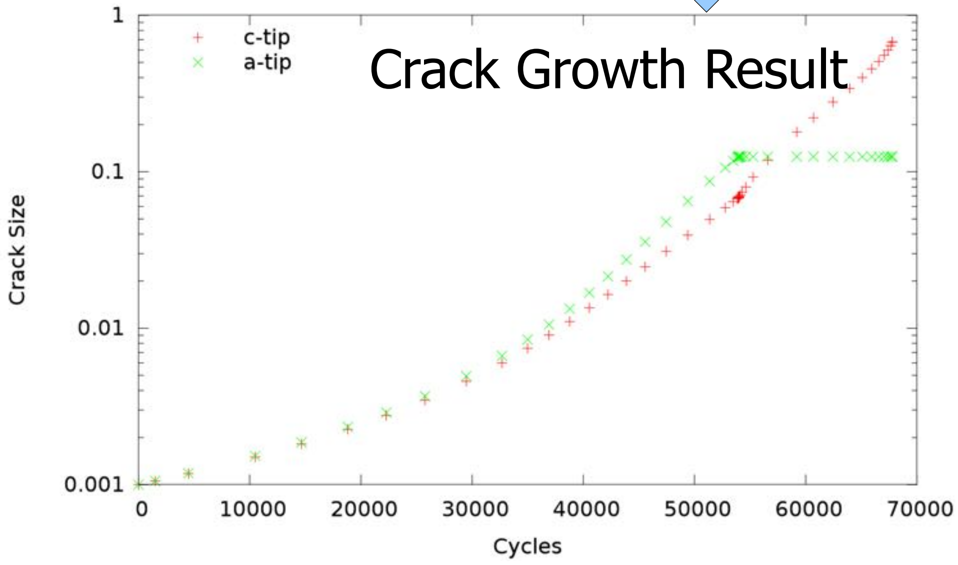
$$\frac{da}{dN} - C(\Delta K(a,c))^n = 0$$

$$\frac{dc}{dN} - C(\Delta K(a,c))^n = 0$$

$$Initial\ Conditions: a(0) = a_i, c(0) = c_i$$

RK ODE Solver

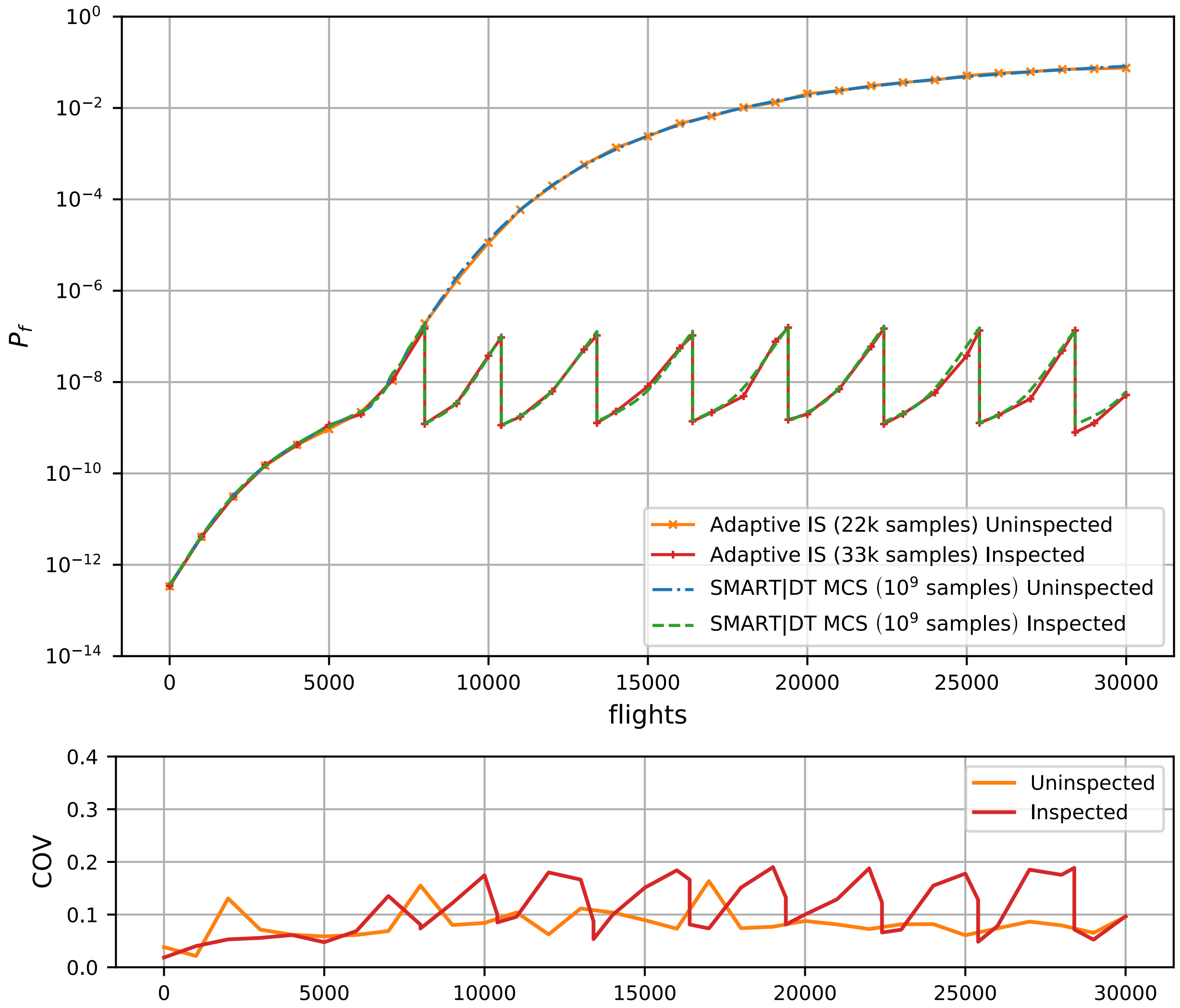




ICG Capabilities	
Method	4-5th order Runge-Kutta
Accuracy	Error controlled by user tolerance
Speed	~7000/sec single proc.
Parallel	95% speedup on 8 proc.
K solutions	Newman-Raju, beta tables

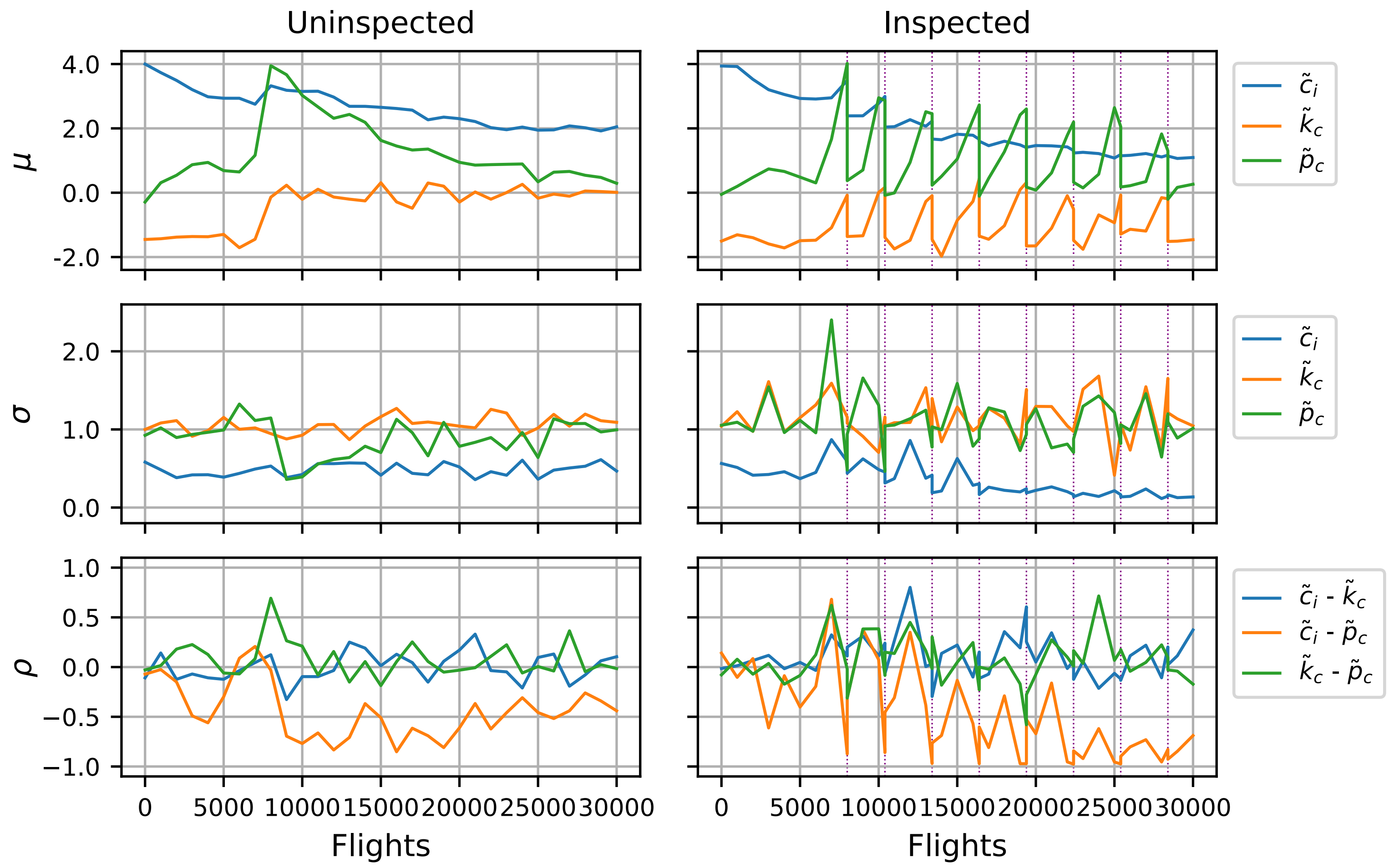
An Ultrafast Crack Growth Lifing Algorightm for Probabilistic Damage Tolerance Analysis, Millwater et al., AA&S 2018

General Aviation Example Problem POF



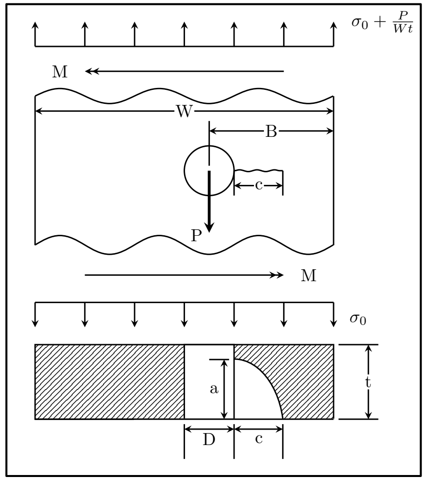
- Adaptive importance sampling parameters
  - $\epsilon_{\mathrm{COV}} = 0.2$
  - $n_{\mathrm{main}} = 100$
  - $n_{\mathrm{adp}} = 30$
- Runtimes (serial)
  - Uninspected: 7 seconds
  - Inspected: 14 seconds

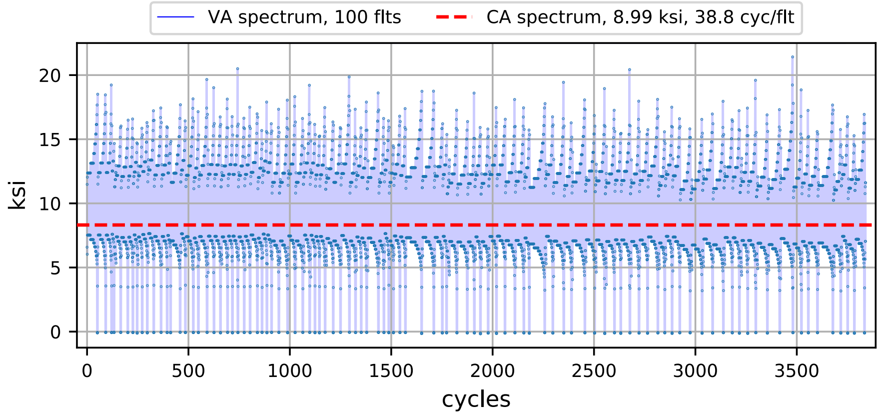
General Aviation Example Problem

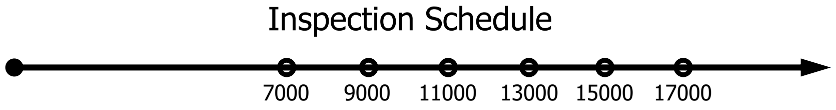


- Uninspected:
  - c_i and K_c important up to $t = 6000$ (inflection on POF)
  - c_i and p_c important after $t = 6000$ with high correlation
- Inspected:
  - Step change at each inspection
  - Switches from c_i and p_c significant before inspection to c_i and k_c significant after
  - Higher correlation between c_i and p_c than uninspected

NASGRO Example

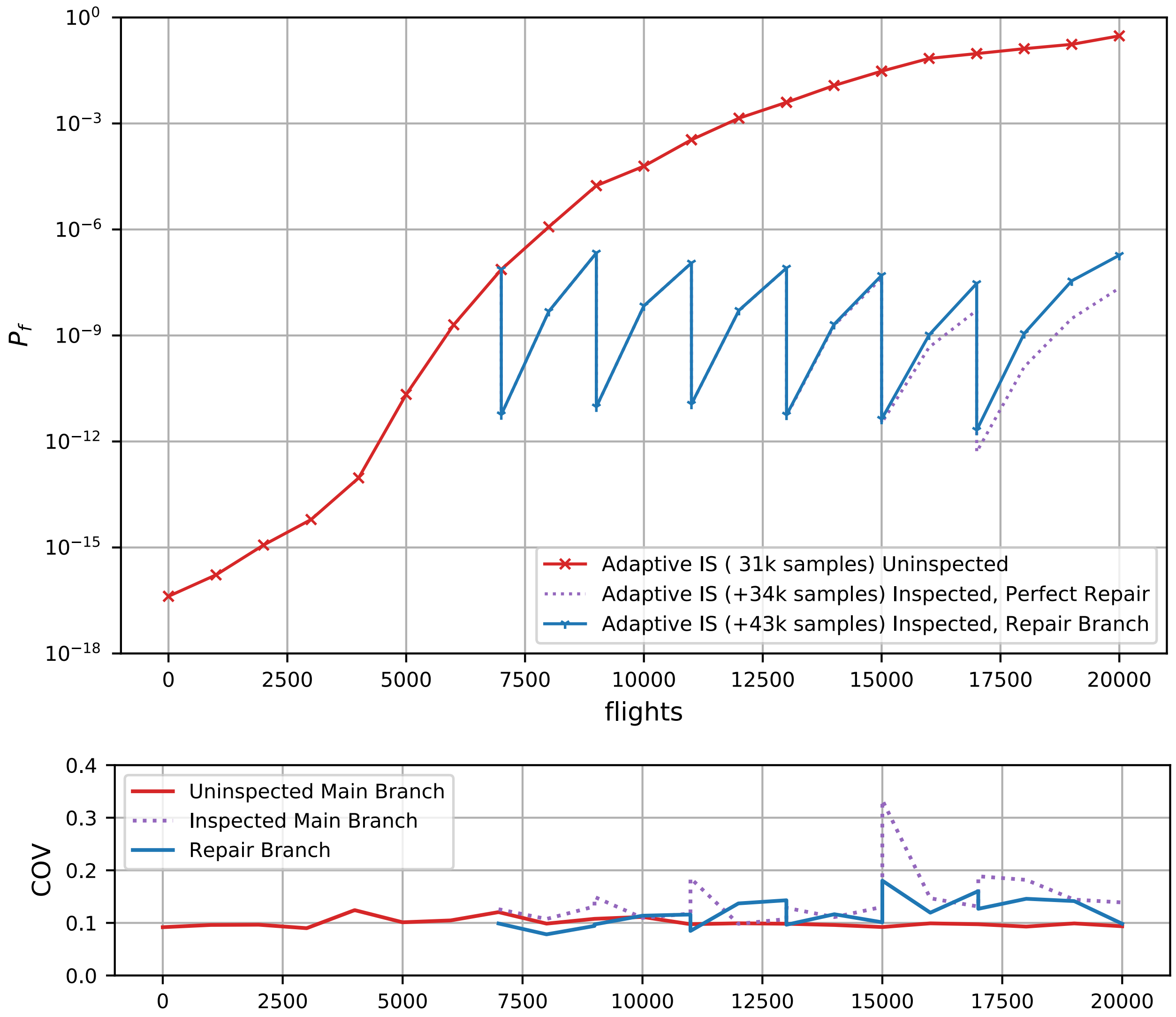






Parameter	Value
Width	Deterministic 0.1562 in
Thickness	Deterministic 0.1562 in
Initial Crack Size	$LN(0.005, 0.002)$ in
Aspect Ratio (A/C)	$N(1.5, 0.14)$
Fracture Toughness	$N(34.8, 3.90)$ ksi $\sqrt{\text{in}}$
Log Paris Constant	$N(-8.777, 0.08)$
Paris Exponent	Deterministic 3.273
Hole Diameter	Deterministic 0.1562 in
Hole Offset	$N(0.05, 0.05)$ in
Maximum Stress per Flight	$EVD(16.74, 2.08, 0.0)$ ksi
Probability of Detection	$LN(0.021, 0.028)$ in
Repair Quality (crack size)	$LN(0.01, 0.004)$ in

NASGRO Example
POF



- Adaptive importance sampling parameters
  - $\epsilon_{\mathrm{COV}} = 0.2$
  - $n_{\mathrm{main}} = 100$
  - $n_{\mathrm{adp}} = 40$

- Runtime (parallel 12 processors)
  - Uninspected: 1 hr 24 min
  - Inspected: 2 hrs 50 min
  - Repair Branch: 1 hr 36 min
  - Total: 5 hrs 50 min

Conclusions

- Adaptive importance sampling increases the sampling efficiency by 5 orders of magnitude
- The optimal sampling density must adapt to region of importance change over time in order to achieve high efficiency
- Importance sampling density parameters give an idea of the parameter sensitivities
- Adaptive importance sampling algorithm applied to variety of POF problems with different sets of random variables, distributions and POF formulations
- Integration into SMART|DT (expected Fall 2020)

Acknowledgments

- Probabilistic Fatigue Management Program for General Aviation, Federal Aviation Administration, Grant 16-G-005
  - Sohrob Mattaghi (FAA Tech Center) – Program Manager
  - Michael Reyer (Kansas City) – Sponsor

Thank you