Probabilistic Timing Estimates in Scenarios Under Testing Constraints

Sergi Vilardell[1], Francesco Rossi[2], Gabriele Giordana[2], Isabel Serra[3], Enrico Mezzetti[1], Jaume Abella[1] , and Francisco J. Cazorla[1]

[1]Barcelona Supercomputing Center (BSC)
[2]Aiko Space S.r.l
[3]Universitat Autònoma de Barcelona (UAB)

Email: sergi.vilardell@bsc.es

04/04/2025

The 40th ACM/SIGAPP Symposium On Applied Computing 2025 Catania, Sicily, Italy

Agenda

1. Probabilistic Timing Analysis (PTA): Motivation and Problem Statement
2. Measurement-Based PTA under Testing Constraints
3. Uncertainty in Extreme Value Theory (EVT)
4. Markov's Inequality for pWCET in Low Sample Scenarios
5. Results
6. Conclusions

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Motivation

Soft Real-Time Requirements in Modern MPSoCs

- Multi-Processor Systems on Chips (MPSoCs) are becoming more complex
- Timing Analysis (TA) is difficult due to complex software and hardware
- Measurement-Based Probabilistic Timing Analysis (MBPTA) provides solid TA

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

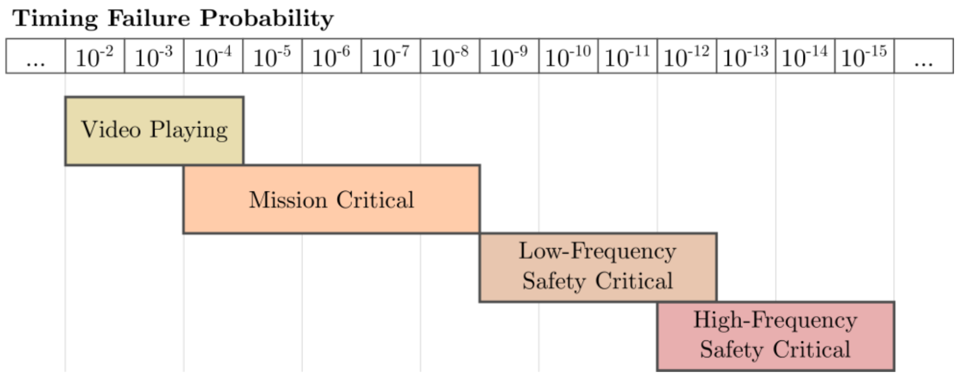
Soft Real-Time Requirements in Modern MPSoCs

- Multi-Processor Systems on Chips (MPSoCs) are becoming more complex
- Timing Analysis (TA) is difficult due to complex software and hardware
- Measurement-Based Probabilistic Timing Analysis (MBPTA) provides solid TA
- Risk tolerance shall be assessed against practicality
- More precise estimations for timing are in proportion to the time and effort required to apply MBPTA techniques

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Soft Real-Time Requirements in Modern MPSoCs

- Multi-Processor Systems on Chips (MPSoCs) are becoming more complex
- Timing Analysis (TA) is difficult due to complex software and hardware
- Measurement-Based Probabilistic Timing Analysis (MBPTA) provides solid TA
- Risk tolerance shall be assessed against practicality
- More precise estimations for timing are in proportion to the time and effort required to apply MBPTA techniques
- In MBPTA this is directly related to the number of test/runs (sample size)
  - **Costs:** increasing number of tests may be too costly (e.g. on-the-road testing)
  - **Benefits:** timing can range from quality, to economic loss or causalities
- However, low sample sizes can produce high uncertainty



Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Probabilistic Timing Analysis /1

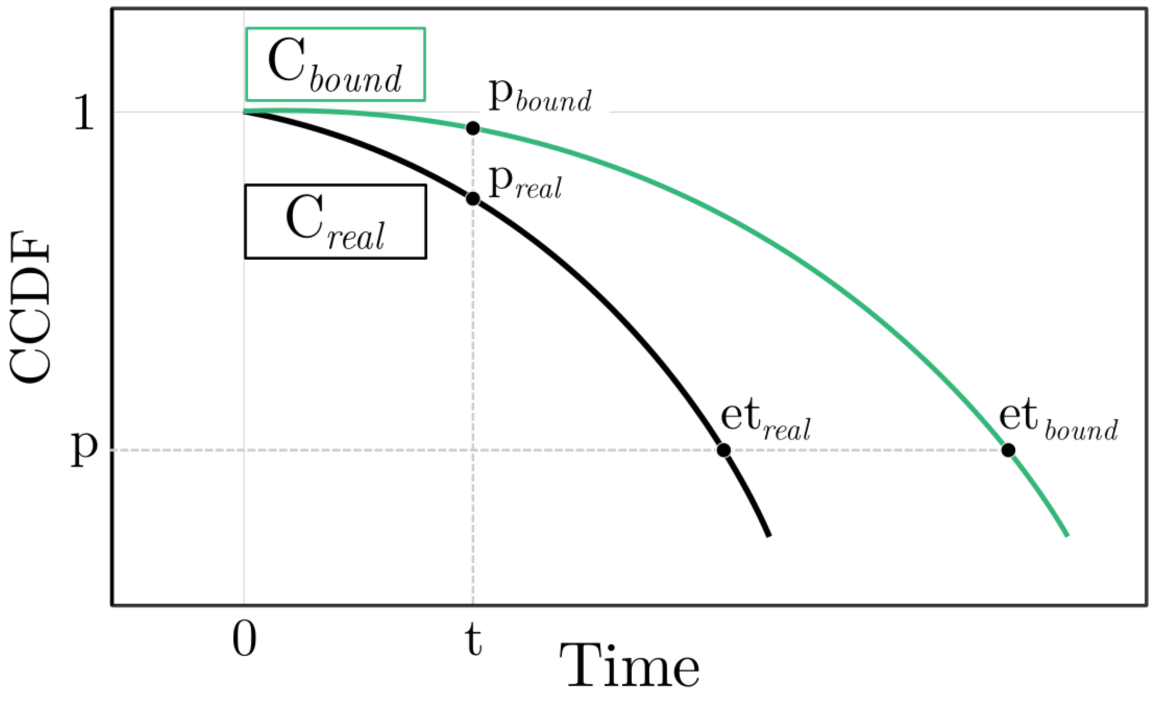
Measurement-Based Probabilistic
Timing Analysis

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Probabilistic Timing Analysis /1

Measurement-Based Probabilistic Timing Analysis

- Based directly on the test/runs on the hardware and application
- The goal is to derive a probabilistic upperbound on the probabilistic Worst-Case Execution Time (pWCET)
- We aim at low sample sizes (N = 100)
- The CCDF denotes the probability of exceeding a certain value (e.g. ET)
- The pWCET required properties
  - ✗ **Optimistic**:
    - ✗ $p_{bound} < p_{real}$
    - ✗ $et_{bound} < et_{real}$
  - ✓ **Conservative**:
    - ✓ $p_{bound} \geq p_{real}$
    - ✓ $et_{bound} \geq et_{real}$

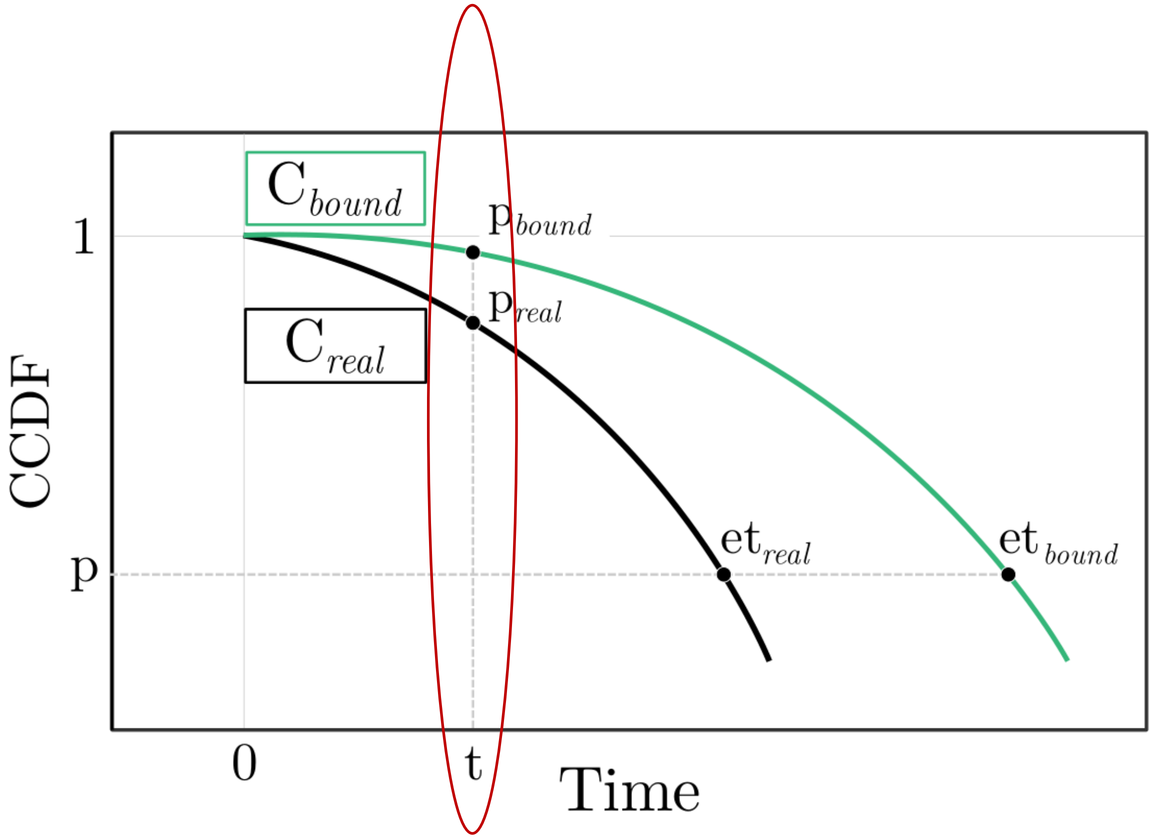


Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Probabilistic Timing Analysis /1

Measurement-Based Probabilistic Timing Analysis

- Based directly on the test/runs on the hardware and application
- The goal is to derive a probabilistic upperbound on the probabilistic Worst-Case Execution Time (pWCET)
- We aim at low sample sizes (N = 100)
- The CCDF denotes the probability of exceeding a certain value (e.g. ET)
- The pWCET required properties
  - ✕ **Optimistic**:
    - ✕ $p_{bound} < p_{real}$
    - ✕ $et_{bound} < et_{real}$
  - ✓ **Conservative**:
    - ✓ $p_{bound} \geq p_{real}$
    - ✓ $et_{bound} \geq et_{real}$

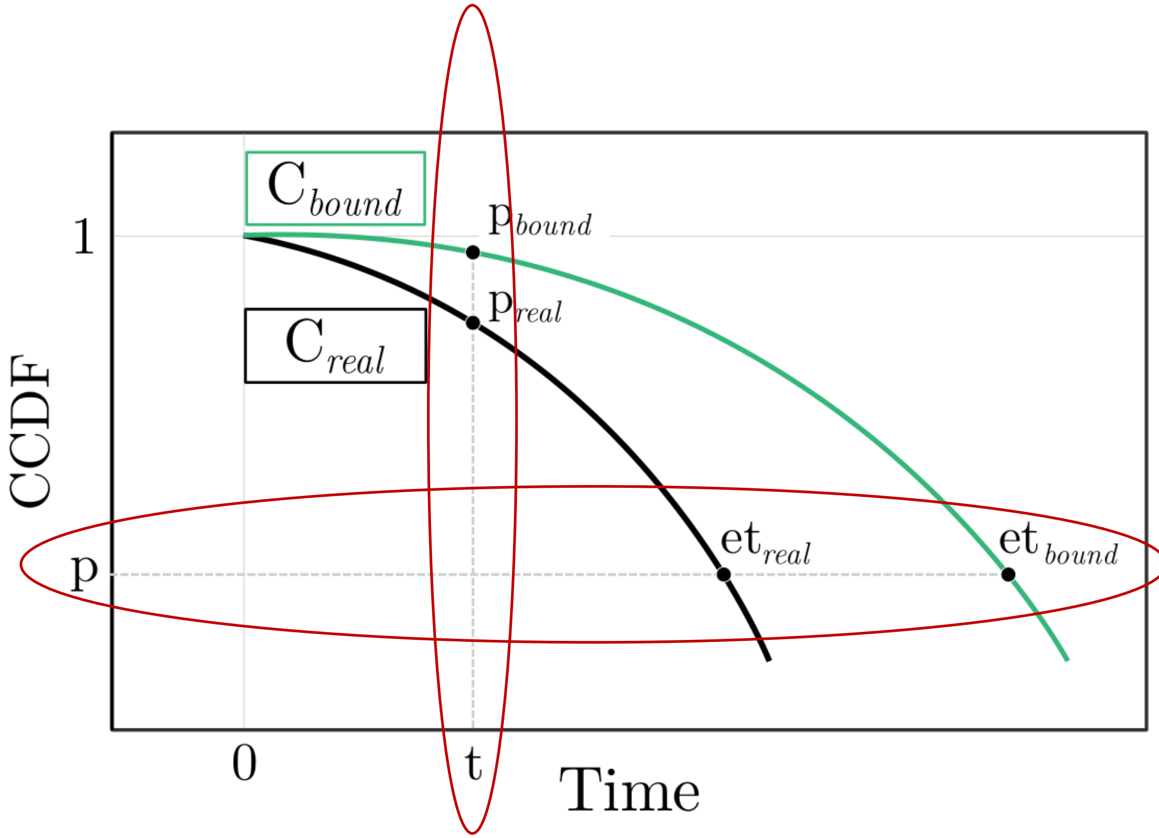


Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Probabilistic Timing Analysis /1

Measurement-Based Probabilistic Timing Analysis

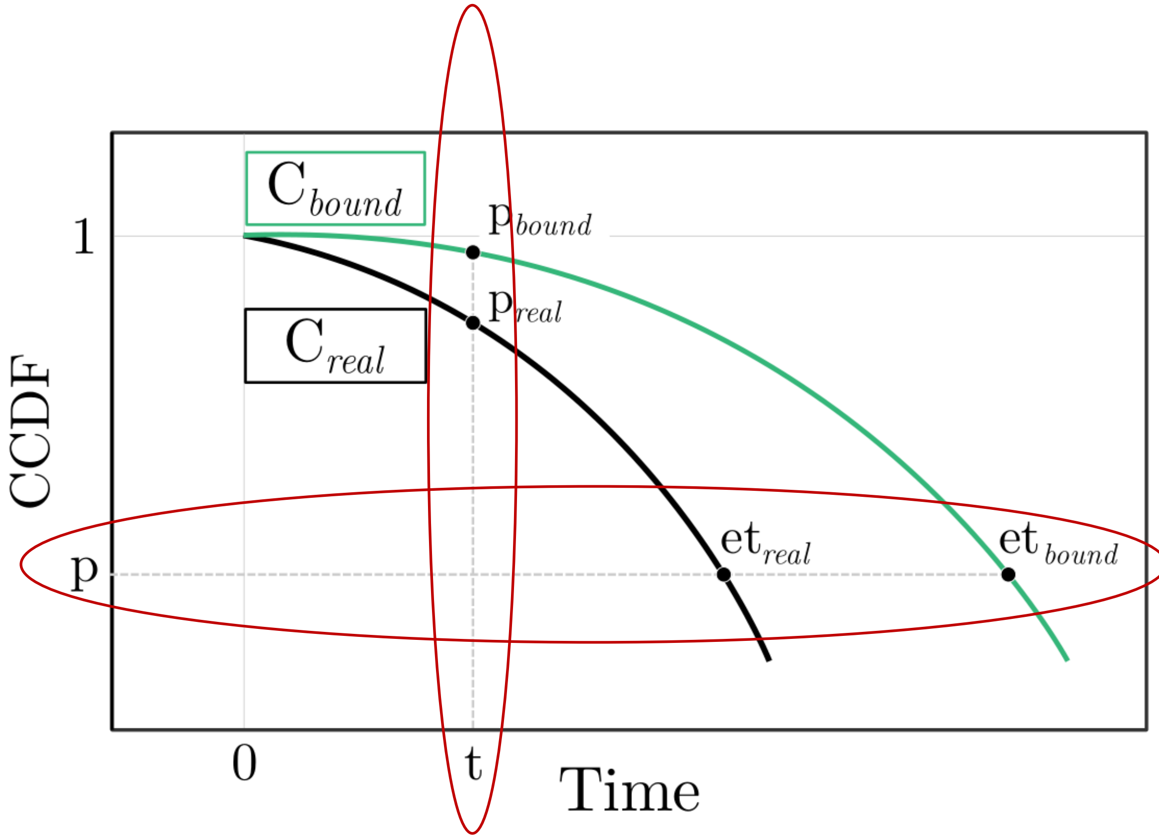
- Based directly on the test/runs on the hardware and application
- The goal is to derive a probabilistic upperbound on the probabilistic Worst-Case Execution Time (pWCET)
- We aim at low sample sizes (N = 100)
- The CCDF denotes the probability of exceeding a certain value (e.g. ET)
- The pWCET required properties
  - ✕ **Optimistic**:
    - ✕ $p_{bound} < p_{real}$
    - ✕ $et_{bound} < et_{real}$
  - ✓ **Conservative**:
    - ✓ $p_{bound} \geq p_{real}$
    - ✓ $et_{bound} \geq et_{real}$



Probabilistic Timing Analysis /1

Measurement-Based Probabilistic Timing Analysis

- Based directly on the test/runs on the hardware and application
- The goal is to derive a probabilistic upperbound on the probabilistic Worst-Case Execution Time (pWCET)
- We aim at low sample sizes (N = 100)
- The CCDF denotes the probability of exceeding a certain value (e.g. ET)
- The pWCET required properties
  - ✕ **Optimistic**:
    - ✕ $p_{bound} < p_{real}$
    - ✕ $et_{bound} < et_{real}$
  - ✓ **Conservative**:
    - ✓ $p_{bound} \geq p_{real}$
    - ✓ $et_{bound} \geq et_{real}$
- Exceedingly pessimistic pWCET are not useful
- pWCET estimates should be **tight** to the real distribution



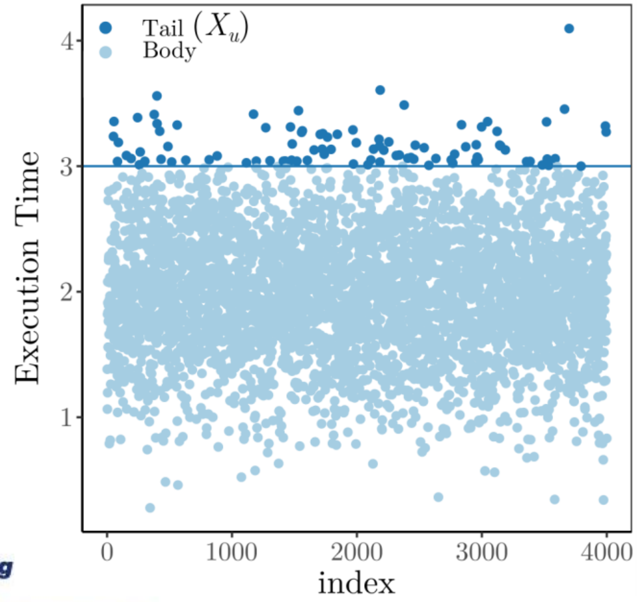
$$tightness(p) = \frac{et_{bound}(p)}{et_{real}(p)}$$

Probabilistic Timing Analysis /2

- Two main frameworks to derive pWCET are used in this work
  - **Extreme Value Theory:** based on the asymptotic distribution of the highest quantiles. Estimating the **extreme value index ξ is crucial**.

EVT

$$G(y;\sigma,\xi) = \begin{cases} 1-\left(1+\xi\frac{y}{\sigma}\right)^{-\frac{1}{\xi}} & \text{if} \quad \xi \neq 0, \\ 1-\exp\left(-\frac{y}{\sigma}\right) & \text{if} \quad \xi = 0, \end{cases}$$



Barcelona Supercomputing Center
Centro Nacional de Supercomputación

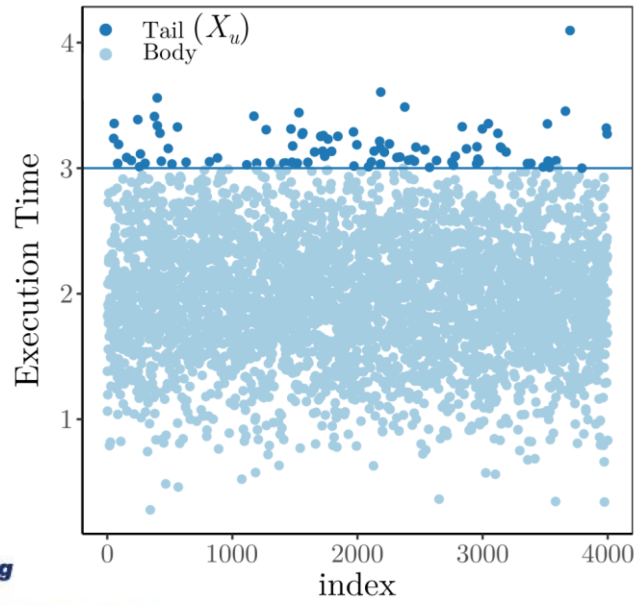
SAFEXPLAIN

Probabilistic Timing Analysis /2

- Two main frameworks to derive pWCET are used in this work
  - **Extreme Value Theory:** based on the asymptotic distribution of the highest quantiles. Estimating the **extreme value index ξ is crucial**.
  - **Markov's Inequality:** based on a probabilistic upperbound of the moments of a distribution

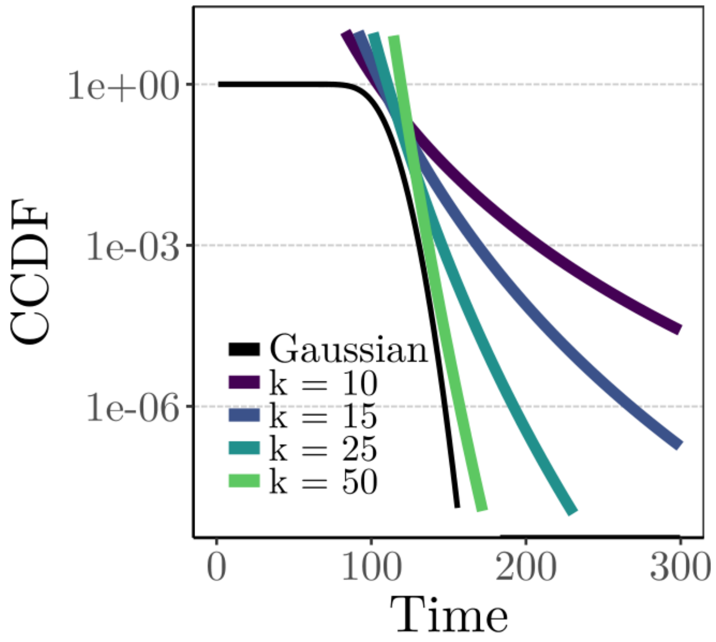
EVT

$$G(y;\sigma,\xi) = \begin{cases} 1-\left(1+\xi\frac{y}{\sigma}\right)^{-\frac{1}{\xi}} & \text{if} \quad \xi \neq 0, \\ 1-\exp\left(-\frac{y}{\sigma}\right) & \text{if} \quad \xi = 0, \end{cases}$$



Markov

$$P(X \geq b) \leq \frac{E(X^k)}{b^k}$$



Barcelona Supercomputing Center
Centro Nacional de Supercomputación

SAFEXPLAIN

EVT and Uncertainty Estimation

EVT Models

- We use three methods to **estimate the tail**, for EXP and GPD models, with different approaches for modelling diversity

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

EVT Models

- We use three methods to **estimate the tail**, for EXP and GPD models, with different approaches for modelling diversity

CV Plot

$$\widehat{\text{cv}}_u = \frac{\sqrt{\hat{V}(X - u|x > u)}}{\hat{E}(X - u|x > u)},$$

Residual Coefficient of Variation

EVT Models

- We use three methods to **estimate the tail**, for EXP and GPD models, with different approaches for modelling diversity

CV Plot

$$\widehat{\mathrm{cv}}_u = \frac{\sqrt{\hat{V}(X-u|x>u)}}{\hat{E}(X-u|x>u)},$$

Residual Coefficient of Variation

Quantile-Quantile

$$d_b(u) = \frac{1}{m}\sum_{j=1}^{m}\left|\frac{\hat{\sigma}_u^b}{\hat{\xi}_u^b}\left[(1-p_j)^{\hat{\xi}_u^b}-1\right]-Q(p_j,X_u^b)\right|,$$

Quantile distance

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

EVT Models

- We use three methods to **estimate the tail**, for EXP and GPD models, with different approaches for modelling diversity

CV Plot

$$\widehat{\mathrm{cv}}_u = \frac{\sqrt{\hat{V}(X-u|x>u)}}{\hat{E}(X-u|x>u)},$$

Residual Coefficient of Variation

Quantile-Quantile

$$d_b(u) = \frac{1}{m}\sum_{j=1}^{m}\left|\frac{\hat{\sigma}_u^b}{\hat{\xi}_u^b}\left[(1-p_j)^{\hat{\xi}_u^b}-1\right]-Q(p_j,X_u^b)\right|,$$

Quantile distance

Semi-Parametric

$$L_{glt}^{PoT}(X) = L_{bulk}(X)(1-H(u|\gamma))L_{glt}(X)$$

$$= \underbrace{\prod_{x_i \le u} h(x_i|\gamma)}_{\text{Semi-parametric}} \underbrace{\prod_{x_i > u} \frac{1-H(u|\gamma)}{\sigma}\left(1+\xi\frac{x_i-u}{\sigma}\right)^{-\frac{(1+\xi)}{\xi}}}_{\text{Parametric}},$$

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Uncertainty Analysis of EVT Models

- For the **Semi-parametric and the QQ models**, we need to assess the uncertainty of the estimator of the Extreme Value Index:

Uncertainty for the Extreme Value Index

$$\hat{\xi} = \sum_{i=1}^{N} \log(1 + y_i/\phi)$$

We use this formula to estimate the tail

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Uncertainty Analysis of EVT Models

- For the **Semi-parametric and the QQ models**, we need to assess the uncertainty of the estimator of the Extreme Value Index:

Uncertainty for the Extreme Value Index

$$\hat{\xi} = \sum_{i=1}^{N} \log(1 + y_i/\phi)$$

We use this formula to estimate the tail

Cramer-Rao Bound

$$\mathrm{var}(\theta) \geq \frac{1}{I(\theta)} \;\; ; \;\; \sqrt{N}[\hat{\theta} - \theta] \xrightarrow{d} \mathcal{N}(0, \sigma^2)$$

We obtain the estimator with least amount of variance

$$\mathrm{var}(\theta) \geq \frac{1}{I(\theta)} = \frac{1}{-(-\frac{1}{\hat{\xi}^2})} = \hat{\xi}^2 \longrightarrow \boxed{\sqrt{N}[\hat{\xi} - \xi] \xrightarrow{d} \mathcal{N}(0, \xi^2)}$$

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Uncertainty Analysis of EVT Models

- For the **Semi-parametric and the QQ models**, we need to assess the uncertainty of the estimator of the Extreme Value Index:

Uncertainty for the Extreme Value Index

$$\hat{\xi} = \sum_{i=1}^{N} \log(1 + y_i/\phi)$$

We use this formula to estimate the tail

Cramer-Rao Bound

$$\text{var}(\theta) \geq \frac{1}{I(\theta)} \; ; \; \sqrt{N}[\hat{\theta} - \theta] \xrightarrow{d} \mathcal{N}(0, \sigma^2)$$

We obtain the estimator with least amount of variance

$$\text{var}(\theta) \geq \frac{1}{I(\theta)} = \frac{1}{-(-\frac{1}{\hat{\xi}^2})} = \hat{\xi}^2 \longrightarrow \boxed{\sqrt{N}[\hat{\xi} - \xi] \xrightarrow{d} \mathcal{N}(0, \xi^2)}$$

Example Uncertainty Estimation

$\xi_1 = -0.5$

$N = 100, N_u = 10$

95% confidence interval

Uncertainty Analysis of EVT Models

- For the **Semi-parametric and the QQ models**, we need to assess the uncertainty of the estimator of the Extreme Value Index:

Uncertainty for the Extreme Value Index

$$\hat{\xi} = \sum_{i=1}^{N} \log(1 + y_i/\phi)$$

We use this formula to estimate the tail

Cramer-Rao Bound

$$\mathrm{var}(\theta) \geq \frac{1}{I(\theta)} \; ; \; \sqrt{N}[\hat{\theta} - \theta] \xrightarrow{d} \mathcal{N}(0, \sigma^2)$$

We obtain the estimator with least amount of variance

$$\mathrm{var}(\theta) \geq \frac{1}{I(\theta)} = \frac{1}{-(-\frac{1}{\hat{\xi}^2})} = \hat{\xi}^2 \longrightarrow \sqrt{N}[\hat{\xi} - \xi] \xrightarrow{d} \mathcal{N}(0, \xi^2)$$

Example Uncertainty Estimation

$$\xi_1 = -0.5$$

$$N = 100, N_u = 10$$

95% confidence interval

$$\hat{\xi}_1 \in [-0.65, -0.34]$$

$$\hat{q}_1 \in [176.7, 240.6]$$

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Markov’s Inequality In Low Sample Scenarios

Markov's Inequality for PTA

Markov's Inequality Theoretically

$$P(X \geq b) \leq \frac{E(X)}{b} \longrightarrow P(X \geq b) \leq \frac{E(X^k)}{b^k}$$

Markov's Inequality for PTA

Markov's Inequality Theoretically

$$P(X \geq b) \leq \frac{E(X)}{b} \rightarrow P(X \geq b) \leq \frac{E(X^k)}{b^k}$$

Markov's Inequality for Samples

$$P(X \geq b) \leq \frac{1}{N} \frac{\sum_i^N x_i^k}{b^k} \rightarrow$$

Find the max_k for each p that does not underestimate

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Markov's Inequality for PTA

Markov's Inequality Theoretically

$$P(X \geq b) \leq \frac{E(X)}{b} \rightarrow P(X \geq b) \leq \frac{E(X^k)}{b^k}$$

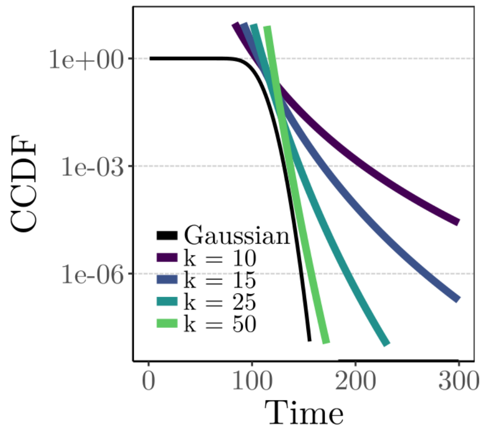
Markov's Inequality for Samples

$$P(X \geq b) \leq \frac{1}{N}\frac{\sum_i^N x_i^k}{b^k} \rightarrow$$

Find the max_k for each p that does not underestimate

MI for pWCET

- The main consideration is that, the **higher the target quantile, the higher the *k* needed** for tight estimations
- **Higher *k* need bigger sample sizes** to estimate the moments accurately



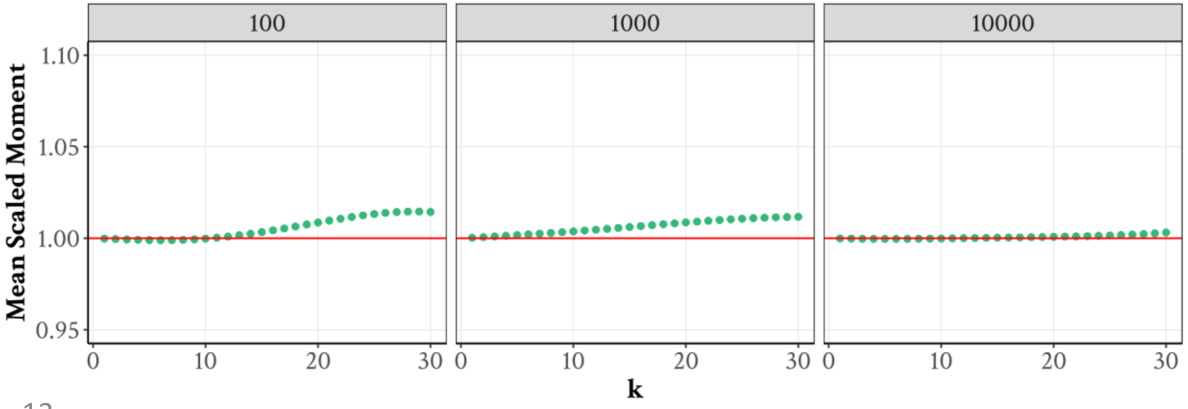
Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Markov's Inequality Considerations

- The estimations are done with the whole sample size, thus decreasing variance
- Markov's Inequality has been tested for sample sizes of N = 1000, but the mean scaled moment show similar results for N = 100

$$P(X \geq b) \leq \frac{\hat{E}(X^k)}{b^k} = \frac{1}{N}\frac{\sum_i^N x_i^k}{b^k}$$

$$\hat{E}(X^k)/E(X^k)$$



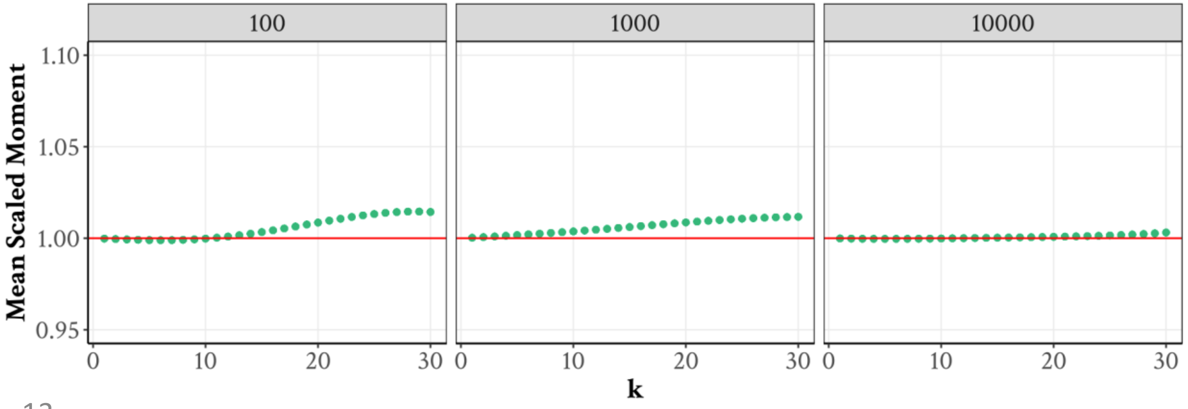
Markov's Inequality Considerations

$$P(X \geq b) \leq \frac{\hat{E}(X^k)}{b^k} = \frac{1}{N} \frac{\sum_i^N x_i^k}{b^k}$$

- The estimations are done with the whole sample size, thus decreasing variance
- Markov's Inequality has been tested for sample sizes of N = 1000, but the mean scaled moment show similar results for N = 100
- We aim at lower quantiles with probability $p = 10^{-6}$, thus smaller k are needed
- Need to estimate the max_k line which needs quantile estimation within the sample

$$max_k(p) = \alpha \log(p) + \beta$$

$\hat{E}(X^k)/E(X^k)$



Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Markov's Inequality Considerations

- The estimations are done with the whole sample size, thus decreasing variance

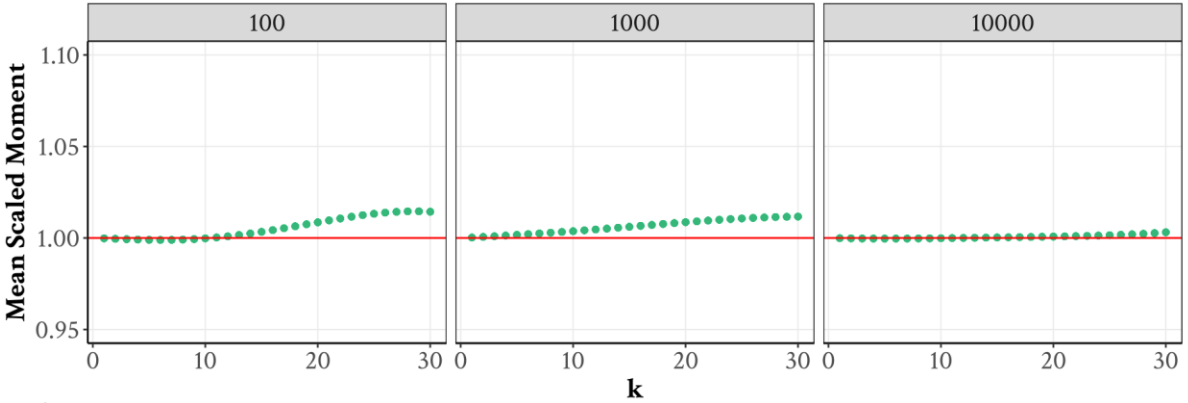
$$P(X \geq b) \leq \frac{\hat{E}(X^k)}{b^k} = \frac{1}{N}\frac{\sum_i^N x_i^k}{b^k}$$

- Markov's Inequality has been tested for sample sizes of N = 1000, but the mean scaled moment show similar results for N = 100
- We aim at lower quantiles with probability $p = 10^{-6}$, thus smaller k are needed
- Need to estimate the max_k line which needs quantile estimation within the sample

$$max_k(p) = \alpha \log(p) + \beta$$

- We make use of the binomial Lower Confidence Interval to increase robustness in the quantile estimation

$\hat{E}(X^k)/E(X^k)$



Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Uncertainty Analysis Summary

- In low sample size scenarios **EVT models** are using a **very small sample size ($\mathbf{N} \sim \mathbf{10}$)** to estimate the tail
- The **least amount of uncertainty** that we can estimate (EVI estimation) is **already very high**
- **Markov Inequality** is less affected by a small sample size if the target probability is not too extreme
- The addition of the Binomial Lower Confidence Interval reduces variance
- **Let us compare EVT models and Markov's Inequality** on pWCET estimations

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Results

Results for Parametric Distributions

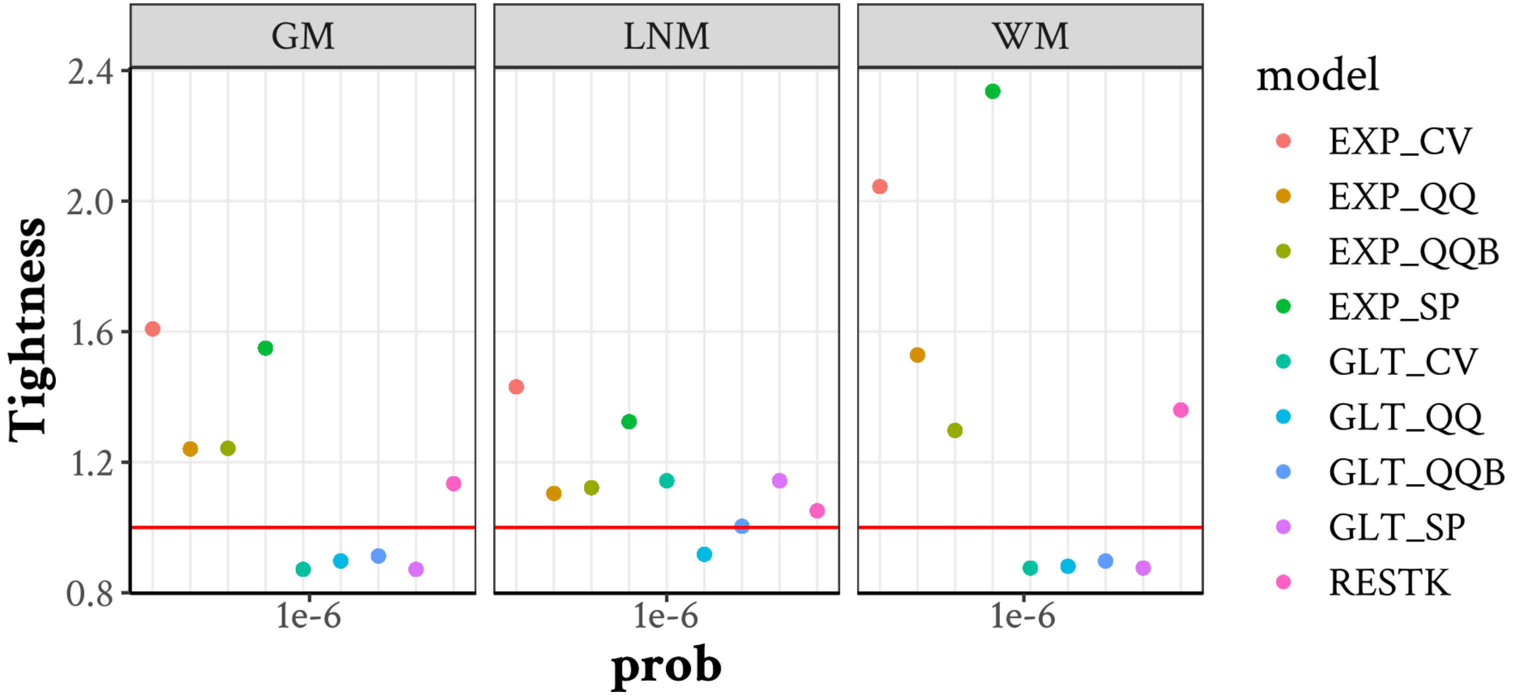
- **Generated synthetic mixture distributions**
- **Sample Size:** $N = 100$
- **Reference Runs:** $N_{ref} = 10^6$
- **Target Probability:** $p = 10^{-6}$

EXP/GLT

$$G(y;\sigma,\xi) = \begin{cases} 1 - \left(1 + \xi\frac{y}{\sigma}\right)^{-\frac{1}{\xi}} & \text{if} \quad \xi \neq 0, \\ 1 - \exp\left(-\frac{y}{\sigma}\right) & \text{if} \quad \xi = 0, \end{cases}$$

RESTK

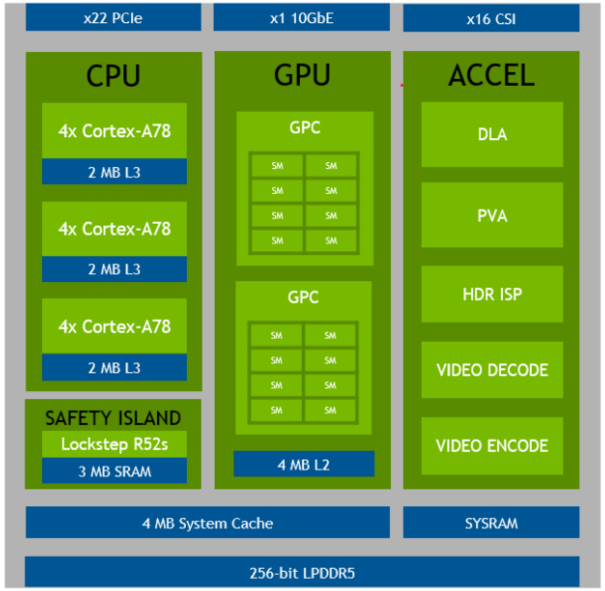
$$P(X \geq b) \leq \frac{E(X^k)}{b^k}$$



Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Results for Hardware Platform

- **Platform:** NVIDIA AGX Orin
- **Sample Size:** $N = 100$
- **Reference Runs:** $N_{ref} = 10^6$
- **Target Probability:** $p = 10^{-6}$

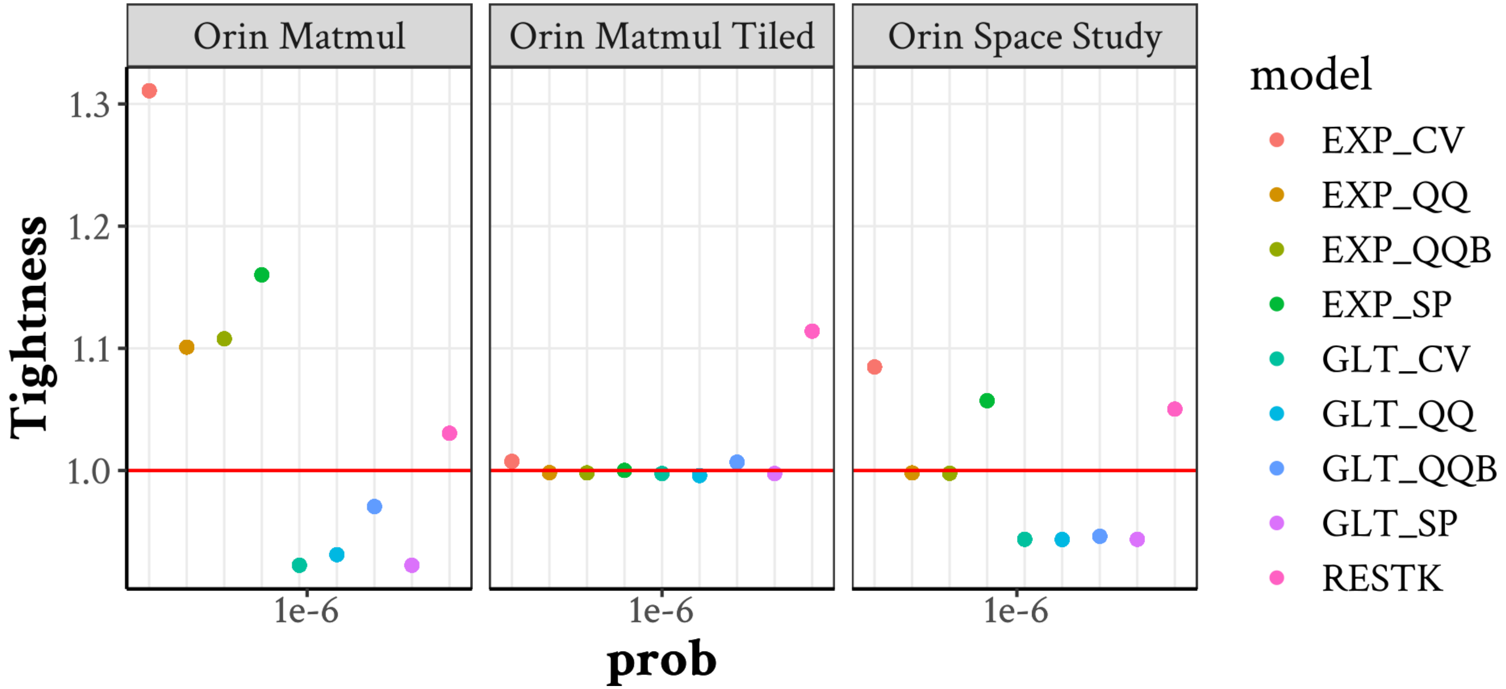


EXP/GLT

$$G(y;\sigma,\xi) = \begin{cases} 1-\left(1+\xi\frac{y}{\sigma}\right)^{-\frac{1}{\xi}} & \text{if } \xi \neq 0,\\ 1-\exp\left(-\frac{y}{\sigma}\right) & \text{if } \xi = 0,\end{cases}$$

RESTK

$$P(X \geq b) \leq \frac{E(X^k)}{b^k}$$



Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Conclusions

Conclusions

- In some domains the **risk assessment is less strict,** and costs could be reduced with few tests
- In others, risks can be **more critical** but obtaining a great number of runs is **too costly or unfeasible**
- We perform an analysis of the **effect of low sample sizes** in current MBPTA techniques

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Conclusions

- In some domains the **risk assessment is less strict,** and costs could be reduced with few tests
- In others, risks can be **more critical** but obtaining a great number of runs is **too costly or unfeasible**
- We perform an analysis of the **effect of low sample sizes** in current MBPTA techniques
- **EVT techniques use a fraction** of the sample for extreme modelling, thus increasing variance
- **Markov's Inequality can use the whole sample**, thus reducing variance
- We introduce the use of Binomial Lower Confidence Interval to **increase robustness** for Markov

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Conclusions

- In some domains the **risk assessment is less strict,** and costs could be reduced with few tests
- In others, risks can be **more critical** but obtaining a great number of runs is **too costly or unfeasible**
- We perform an analysis of the **effect of low sample sizes** in current MBPTA techniques
- **EVT techniques use a fraction** of the sample for extreme modelling, thus increasing variance
- **Markov's Inequality can use the whole sample**, thus reducing variance
- We introduce the use of Binomial Lower Confidence Interval to **increase robustness** for Markov
- EVT techniques can produce **tight upperbounds**, but the results are **less consistent** due to increased variance
- The results show that Markov's Inequality can keep **consistently tight upperbounds** for a variety of scenarios in synthetic and hardware data.

Thank You

Sergi Vilardell[1], Francesco Rossi[2], Gabriele Giordana[2], Isabel Serra[3], Enrico Mezzetti[1], Jaume Abella[1] , and Francisco J. Cazorla[1]

[1]Barcelona Supercomputing Center (BSC)
[2]Aiko Space S.r.l
[3]Universitat Autònoma de Barcelona (UAB)

Email: sergi.vilardell@bsc.es

04/04/2025