

A DISTRIBUTION-AGNOSTIC AND CORRELATION-AWARE ANALYSIS OF PERIODIC TASKS



RTSS 2024
December 12

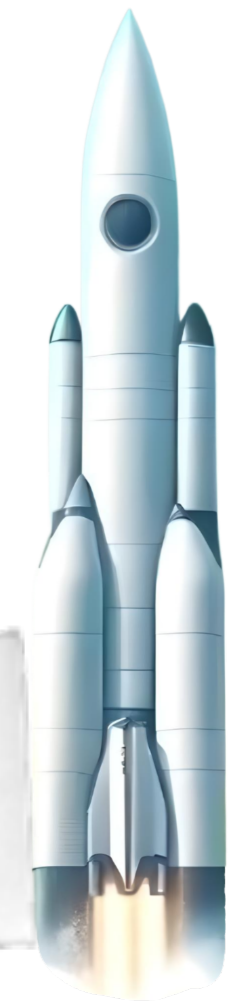


Filip Marković, Georg von der Brüggen, Mario Günzel,
Jian-Jia Chen, and Björn B. Brandenburg



THE RISE OF PROBABILISTIC ANALYSIS

Many modern systems **fail** to meet their timing requirements



Expert Opinion on Tesla Model S with Autopilot, 2016

NHTSA Recall notice 1

NHTSA Recall notice 2

NHTSA Recall notice 3

NHTSA Recall notice 4

VW/Audi / Nov. 2022



Surviving an In-Flight Anomaly: What Happened on Ingenuity's Sixth Flight,

NASA, Håvard Grip. 2021

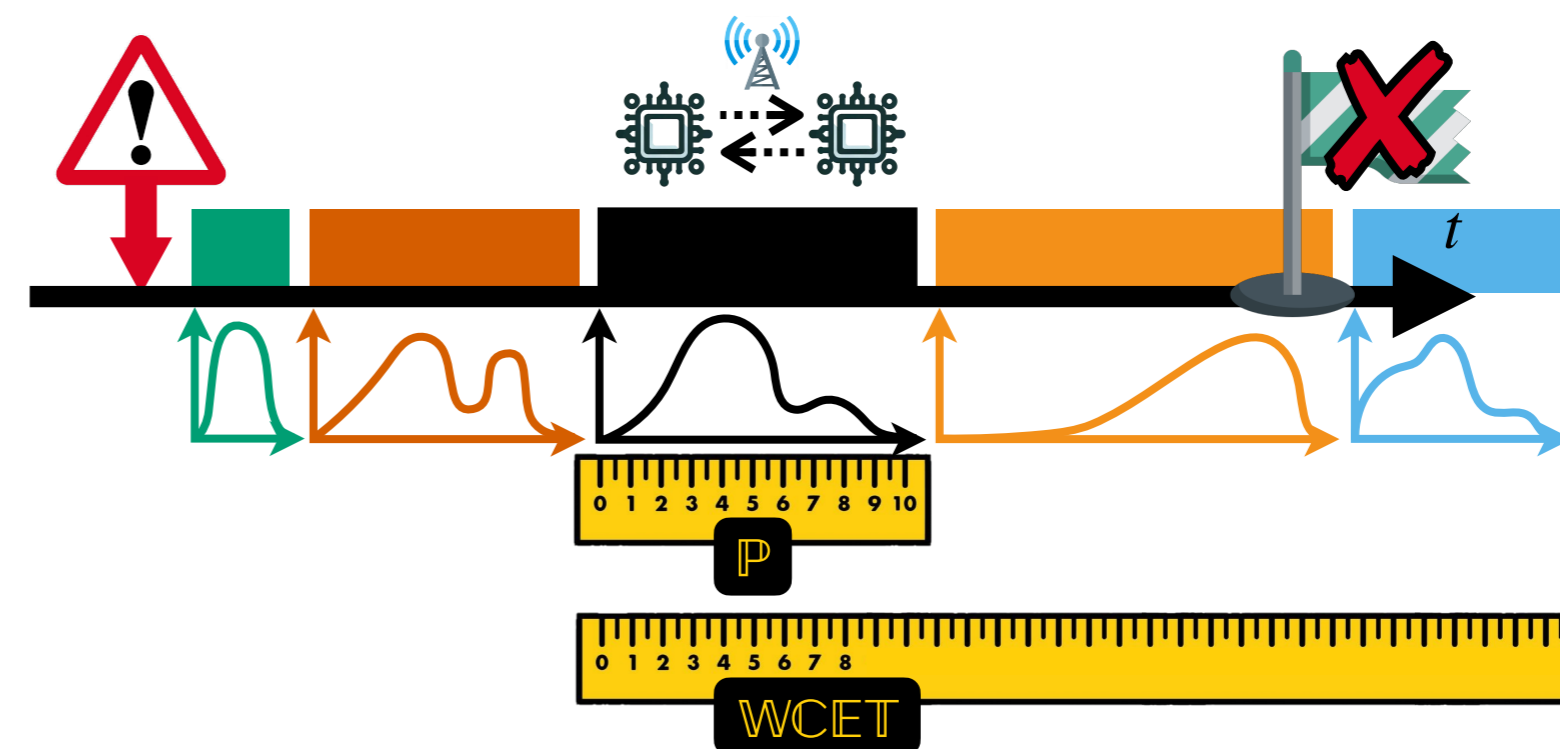


Many safety standards are defined in terms of **failure probability**

Many modern systems are **not statically** analyzable but rather **statistically**

SIL	Low demand mode prob. failure on demand	Continuous/High demand mode prob. failure per hour
1	$\geq 10^{-2}$ to $< 10^{-1}$	$\geq 10^{-6}$ to $< 10^{-5}$
2	$\geq 10^{-3}$ to $< 10^{-2}$	$\geq 10^{-7}$ to $< 10^{-6}$
3	$\geq 10^{-4}$ to $< 10^{-3}$	$\geq 10^{-8}$ to $< 10^{-7}$
4	$\geq 10^{-5}$ to $< 10^{-4}$	$\geq 10^{-9}$ to $< 10^{-8}$

Table 1: IEC 61508: Permitted Failure Probabilities [1]



[1] "The safe and effective application of probabilistic techniques in safety-critical systems", Agrawal et al. ICCAD (2020)

DEPENDENCE – A MAJOR OBSTACLE

There are several **open problems** in the field of probabilistic analysis

HOW TO MODEL DEPENDENCE?

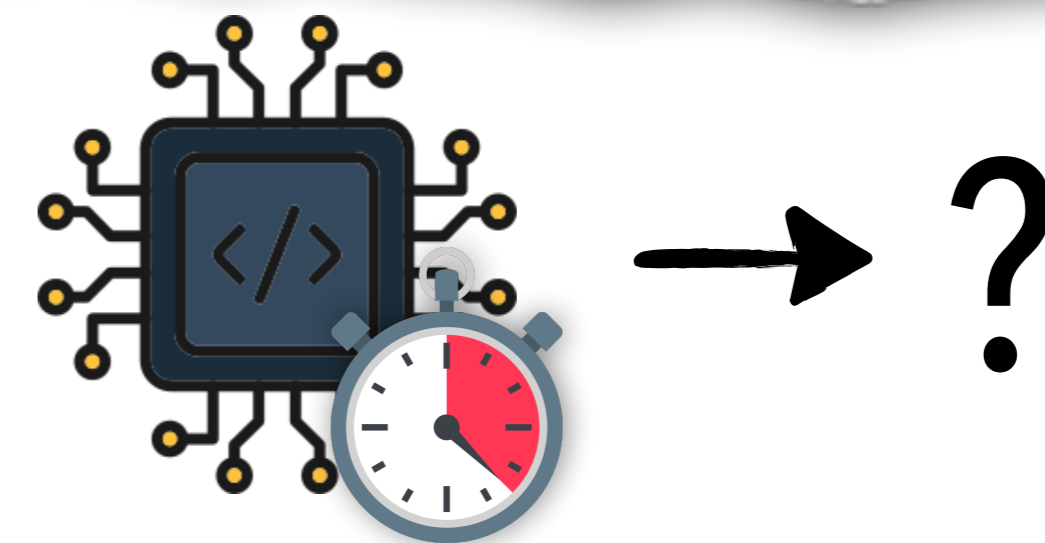
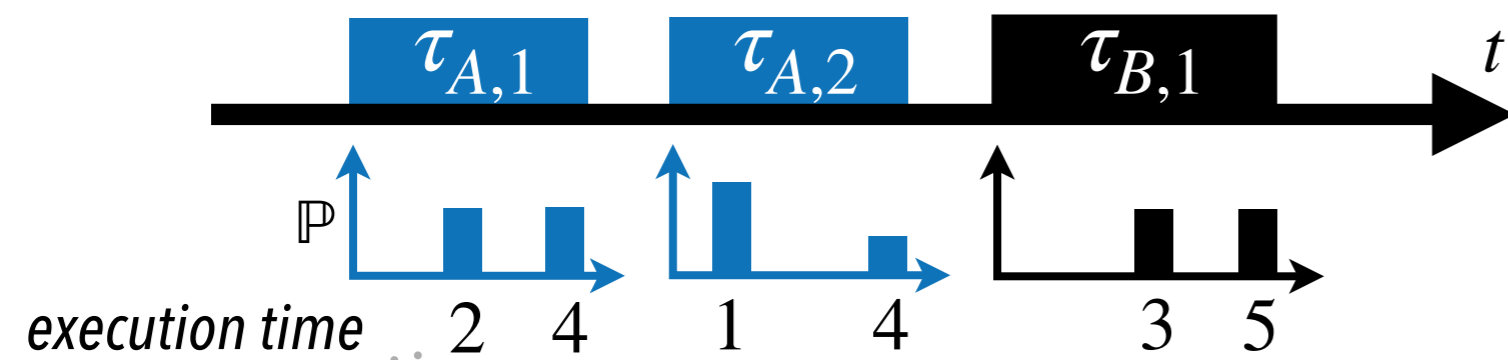
“How to handle ... dependences between the execution times of
(i) jobs of the same task, and
(ii) jobs of different tasks?”

[2]

HOW TO STATISTICALLY INFER DEPENDENCE?

“Appropriate statistical studies are needed to investigate the
types of dependences and their impact on probabilistic
schedulability analysis”

[2]



HOW TO QUANTIFY DEPENDENCE?

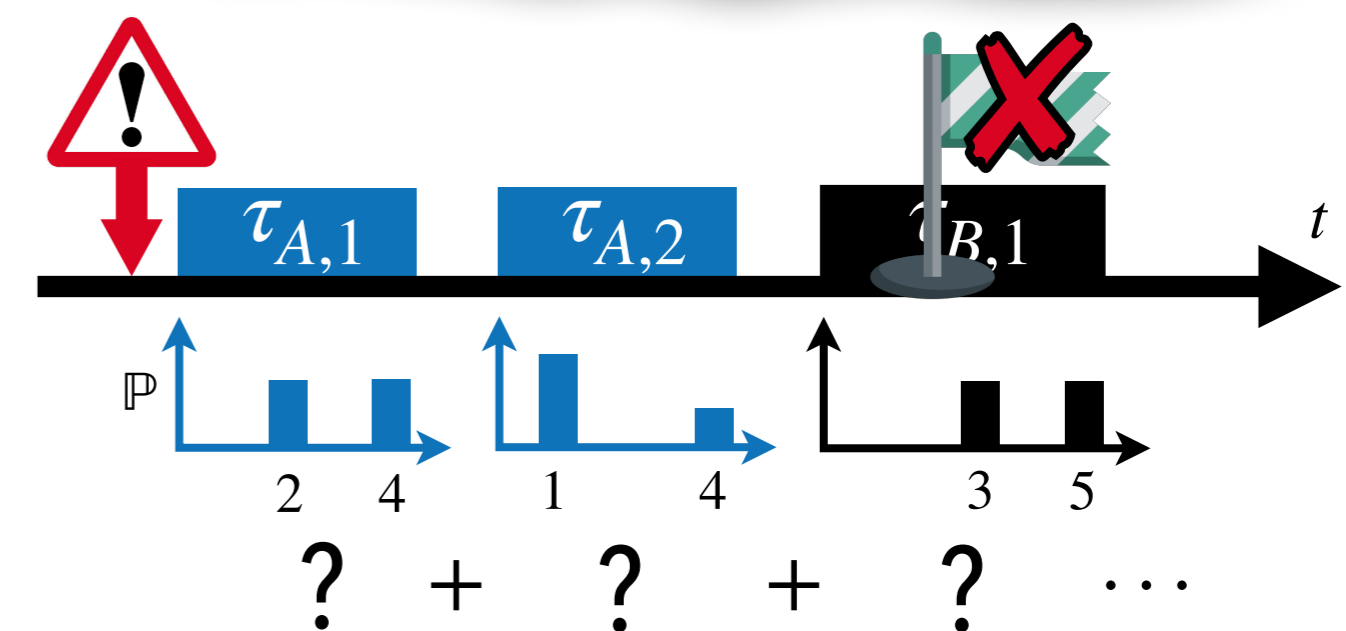
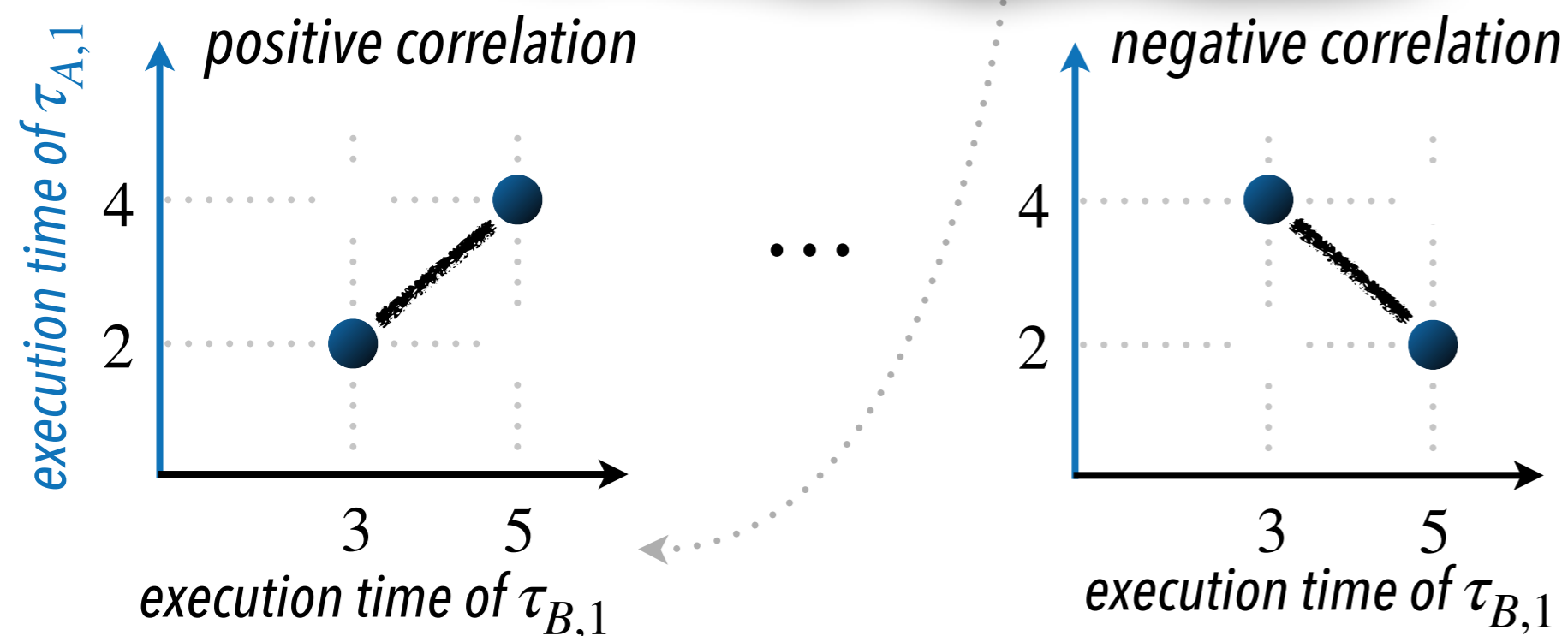
“The impact of these dependences may vary
based on how strong they are.”

[2]

HOW TO USE IT IN RT ANALYSIS?

“Analyses are needed that can address
dependencies.”

[2]



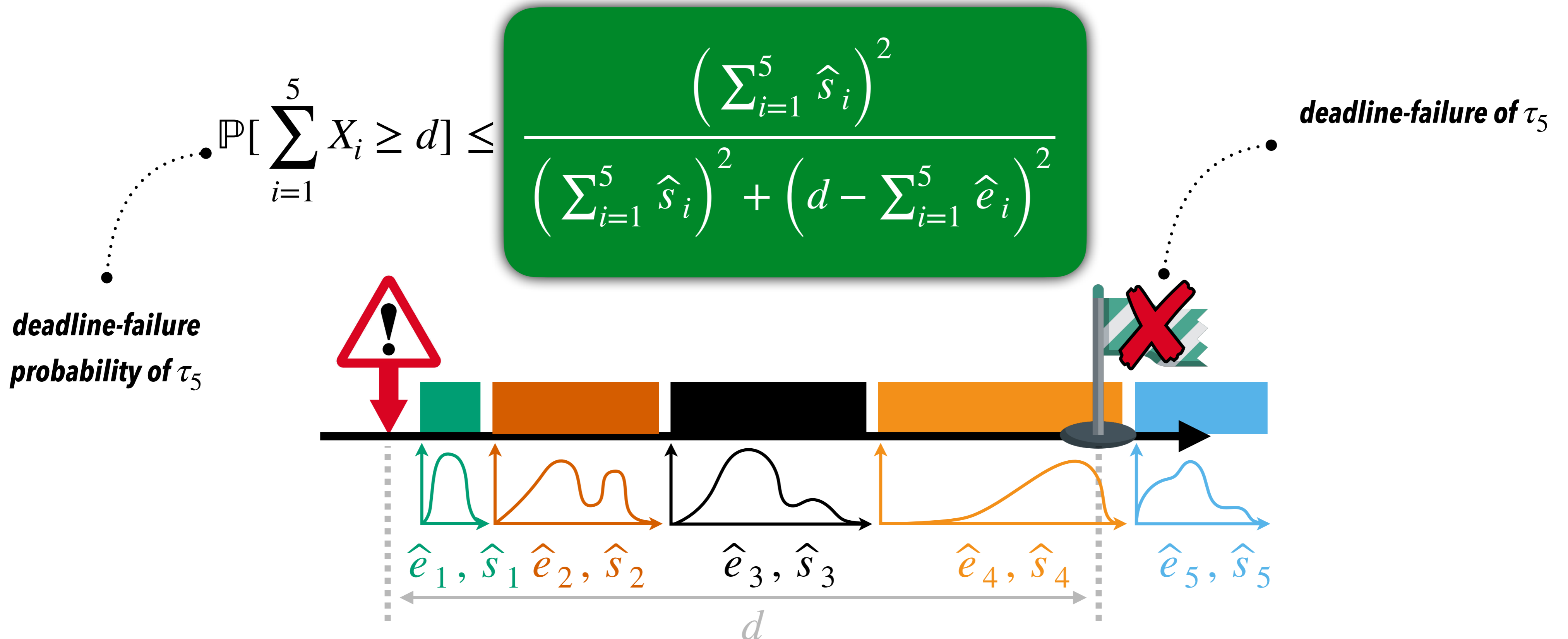
[2] “A survey of probabilistic schedulability analysis techniques for real-time systems” Davis and Cucu-Grosjean. LITES (2019)

CORRELATION-TOLERANT ANALYSIS [3]

Sound, irrespective of **any** potential dependence

Inputs:

- ▶ \hat{e}_i : upper bound on the **mean** execution time (ET) of any job of task τ_i
- ▶ \hat{s}_i : upper bound on the **standard deviation** of the ET of any job of τ_i



[3] "CTA: A Correlation-Tolerant Analysis of the Deadline-Failure Probability of Dependent Tasks" Marković et al. *RTSS* (2023)

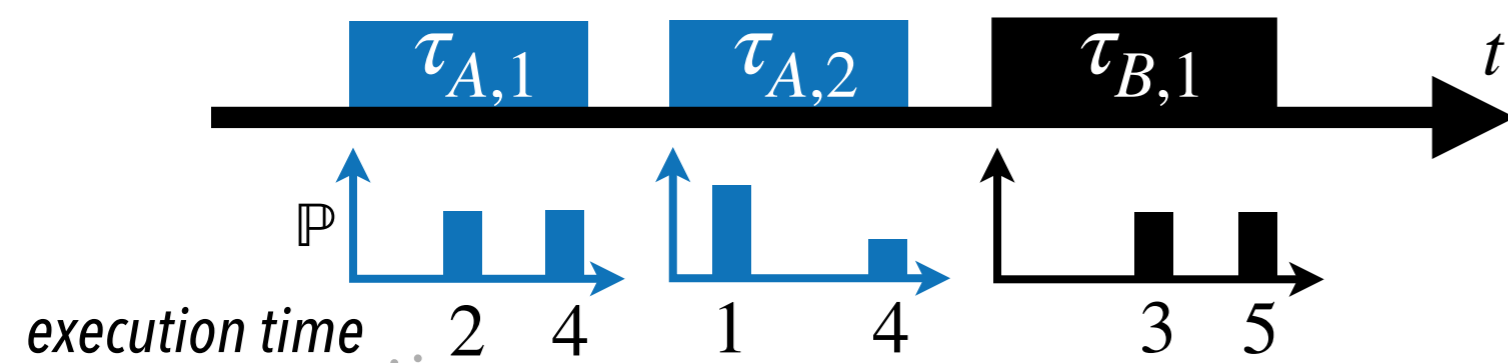
DEPENDENCE – A MAJOR OBSTACLE

However, these open problems remain **unaddressed**

HOW TO MODEL DEPENDENCE?

“How to handle ... dependences between the execution times of
(i) jobs of the same task, and
(ii) jobs of different tasks?”

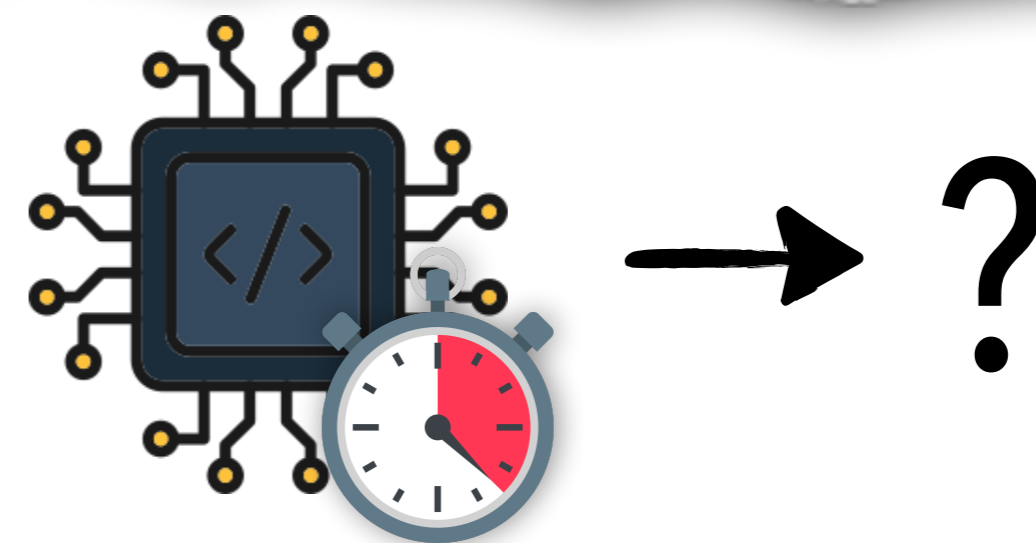
[2]



HOW TO STATISTICALLY INFER DEPENDENCE?

“Appropriate statistical studies are needed to investigate the
types of dependences and their impact on probabilistic
schedulability analysis”

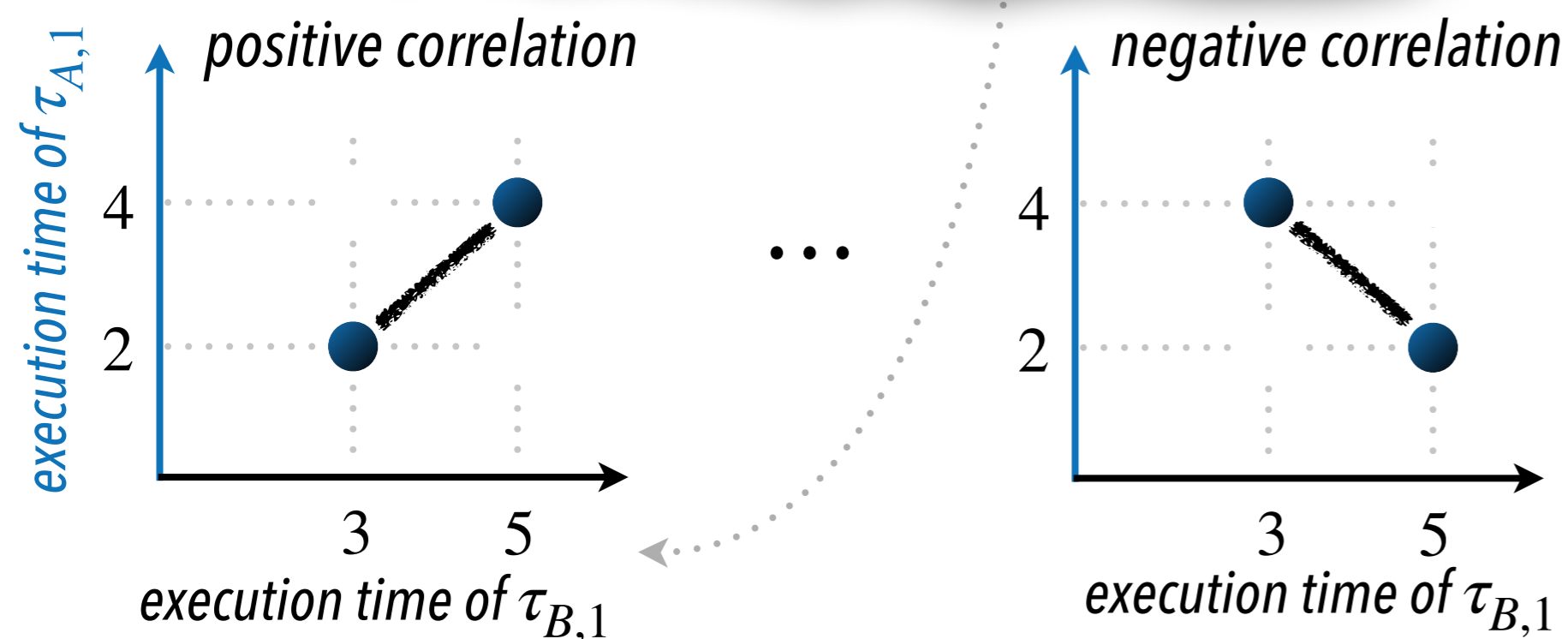
[2]



HOW TO QUANTIFY DEPENDENCE?

“The impact of these dependences may vary
based on how strong they are.”

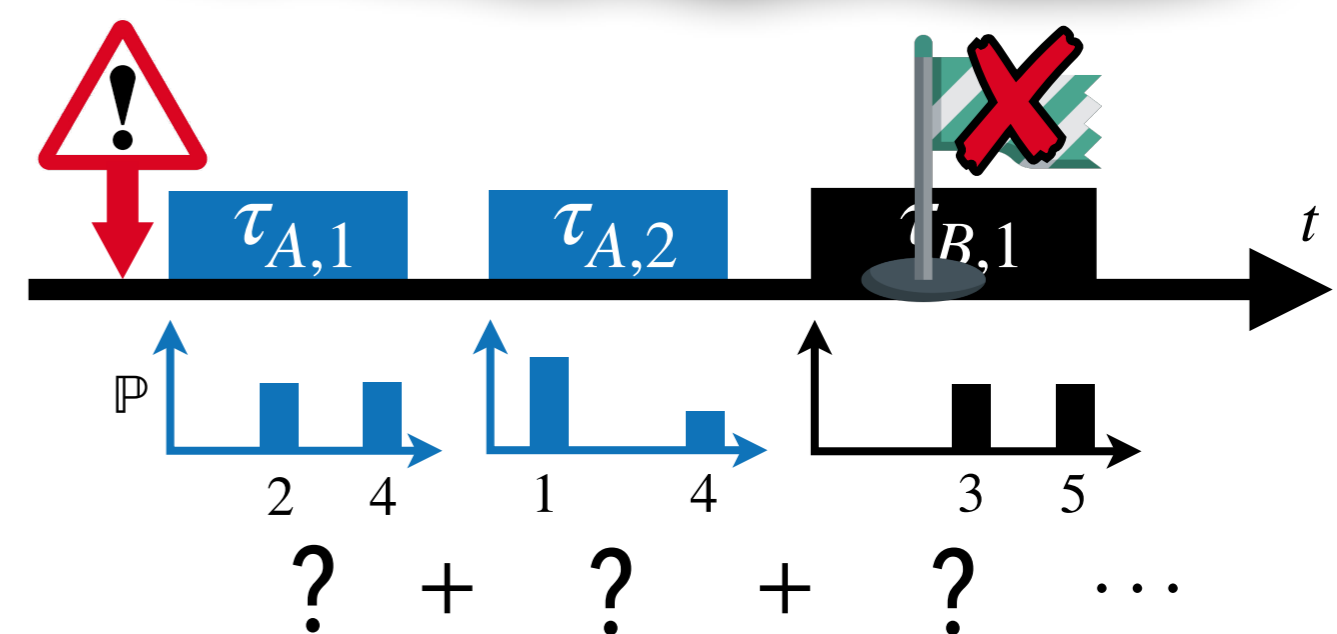
[2]



HOW TO USE IT IN RT ANALYSIS?

“Analyses are needed that can address
dependencies.”

[2]



[2] “A survey of probabilistic schedulability analysis techniques for real-time systems” Davis and Cucu-Grosjean. LITES (2019)

THIS PAPER (PART 1)

A CORRELATION-AWARE ANALYSIS

HOW TO MODEL DEPENDENCE?

*“How to handle ... dependences between the execution times of
(i) jobs of the same task, and
(ii) jobs of different tasks?”*

[2]

HOW TO QUANTIFY DEPENDENCE?

*“The impact of these dependences may vary
based on how strong they are.”*

[2]

HOW TO USE IT IN RT ANALYSIS?

*“Analyses are needed that can address
dependencies.”*

[2]

CORRELATION AWARENESS

How do we model various types of dependence?

HOW TO QUANTIFY DEPENDENCE?

“The impact of these dependences may vary based on how strong they are.”

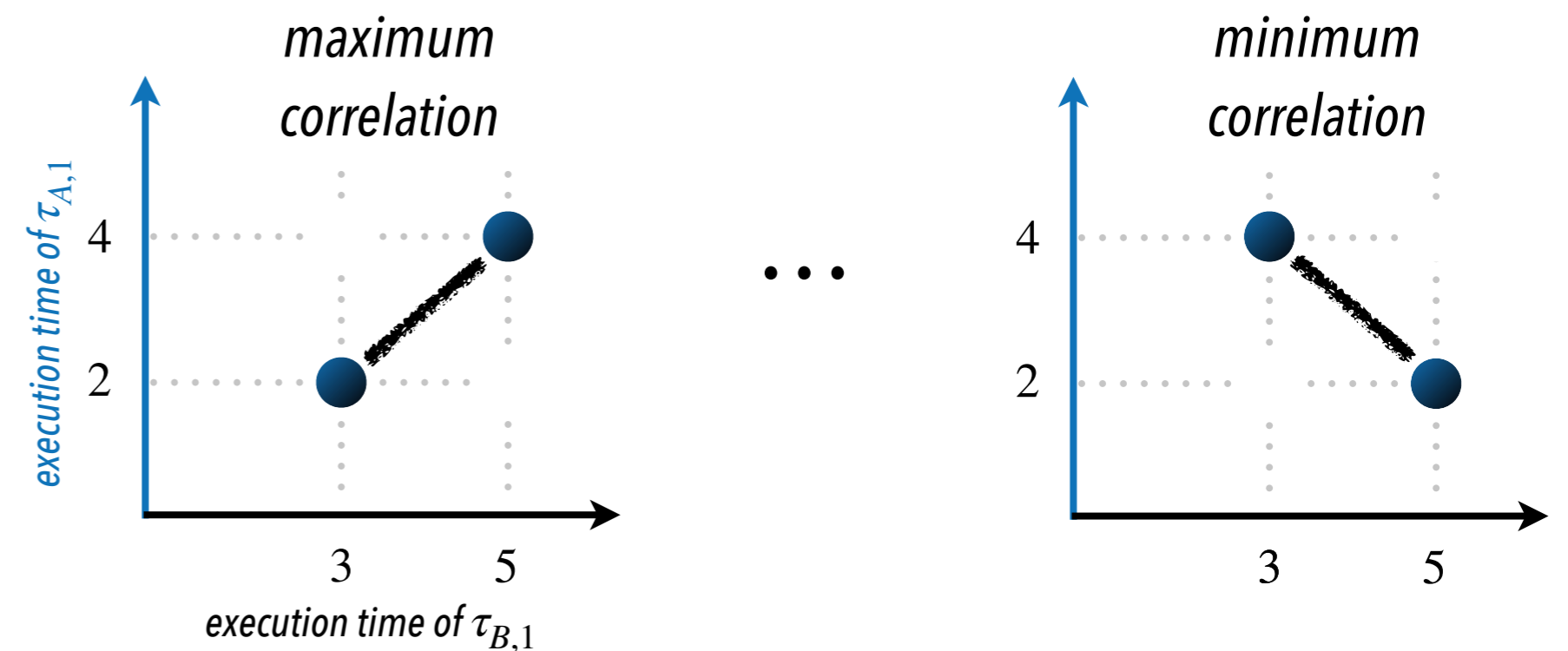
[2]

Covariance – a measure of the degree to which two random variables fluctuate together.

HOW TO MODEL DEPENDENCE?

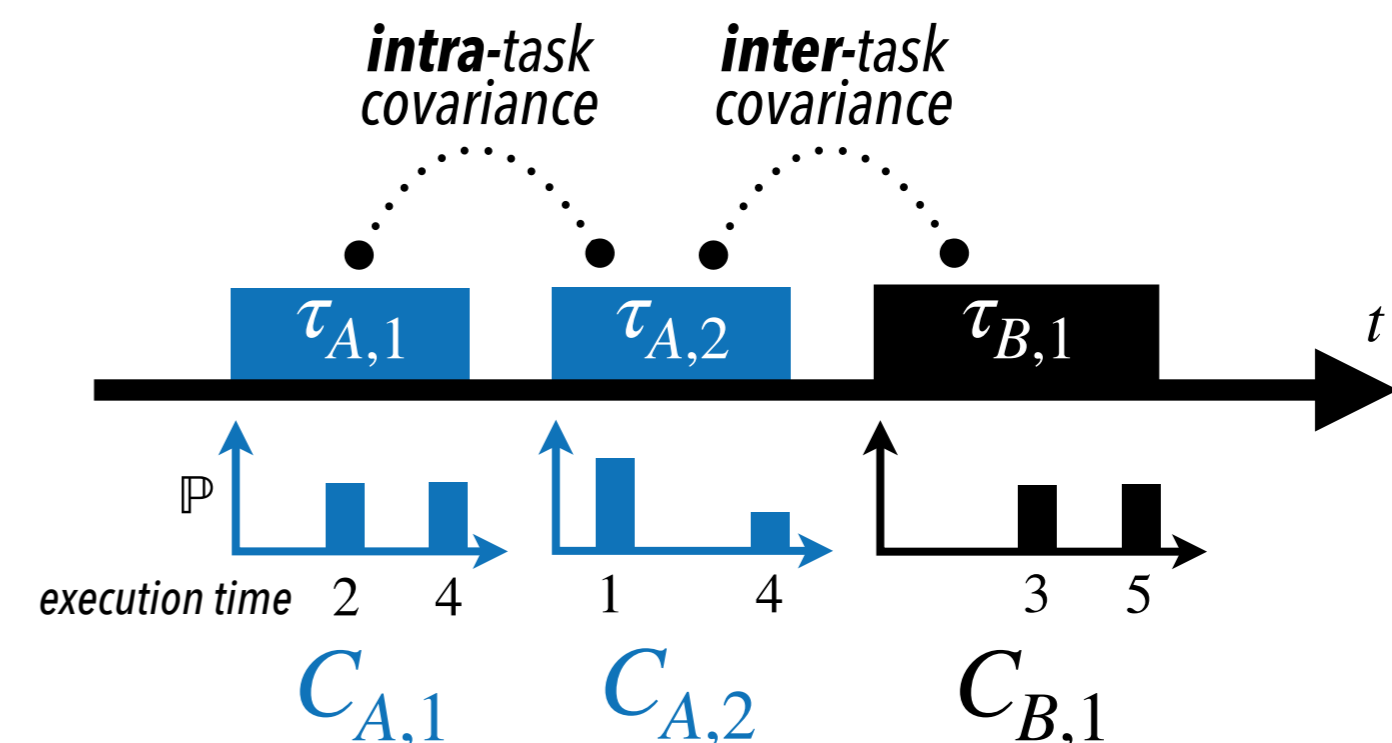
“How to handle ... dependences between the execution times of (i) jobs of the same task, and (ii) jobs of different tasks?”

[2]



Covariance of $C_{A,1}$ and $C_{B,1}$

$$\text{Cov}[C_{A,1}, C_{B,1}] = 1 \quad \dots \quad \text{Cov}[C_{A,1}, C_{B,1}] = -1$$



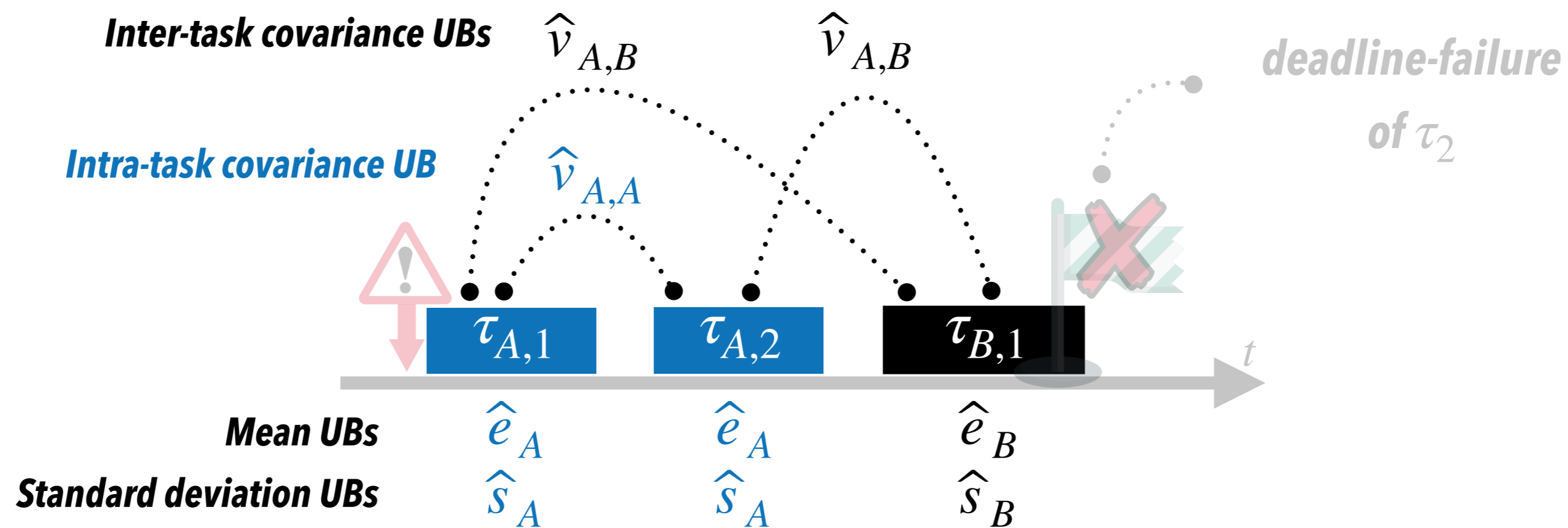
[2] “A survey of probabilistic schedulability analysis techniques for real-time systems” Davis and Cucu-Grosjean LITES (2019)

CORRELATION-AWARE ANALYSIS

An **efficient** way to analyse deadline-failure probability using **covariances**

Inputs:

- ▶ \hat{e}_i : upper bound on the **mean** execution time (ET) of any job of task τ_i
- ▶ \hat{s}_i : upper bound on the **standard deviation** of ET of any job of τ_i
- ▶ $\hat{v}_{i,i}$: upper bound on the ET **intra-task covariance** between any two jobs of τ_i
- ▶ $\hat{v}_{i,k}$: upper bound on the ET **inter-task covariance** between any two jobs of two distinct tasks τ_i and τ_k



HOW TO USE IT IN RT ANALYSIS?

“Analyses are needed that can address dependencies.”

[2]

[2] “A survey of probabilistic schedulability analysis techniques for real-time systems” Davis and Cucu-Grosjean LITES (2019)

CAA IN A NUTSHELL

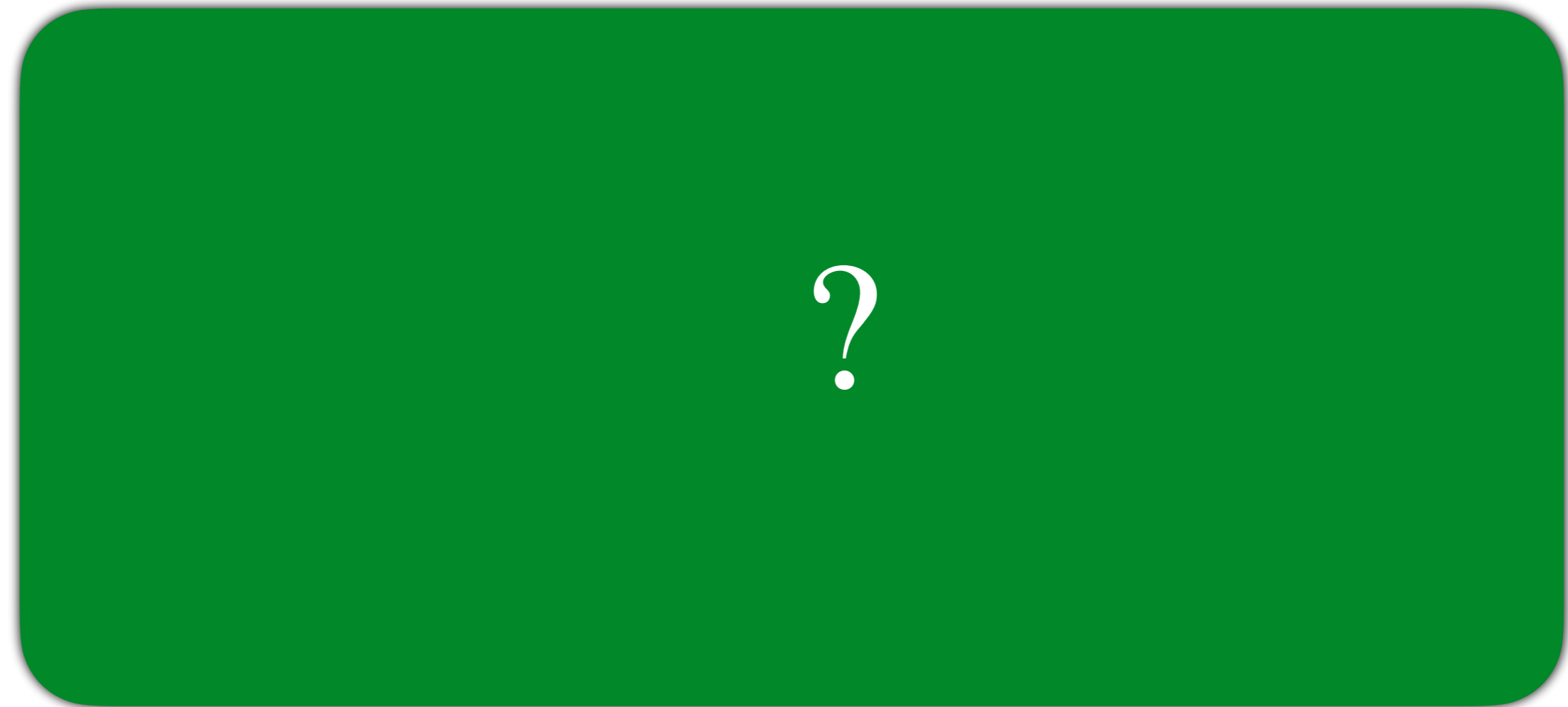
The **goal** is to express everything in terms of CAA inputs

*deadline-failure
probability of τ_B*

$$\mathbb{P}[R_B \geq d] \leq$$

*unknown response-time
distribution of τ_B*

*relative
deadline*



CAA IN A NUTSHELL

CAA rests on ...

deadline-failure
probability of τ_B

unknown response-time
distribution of τ_B

$$\mathbb{P}[R_B \geq d] \leq$$

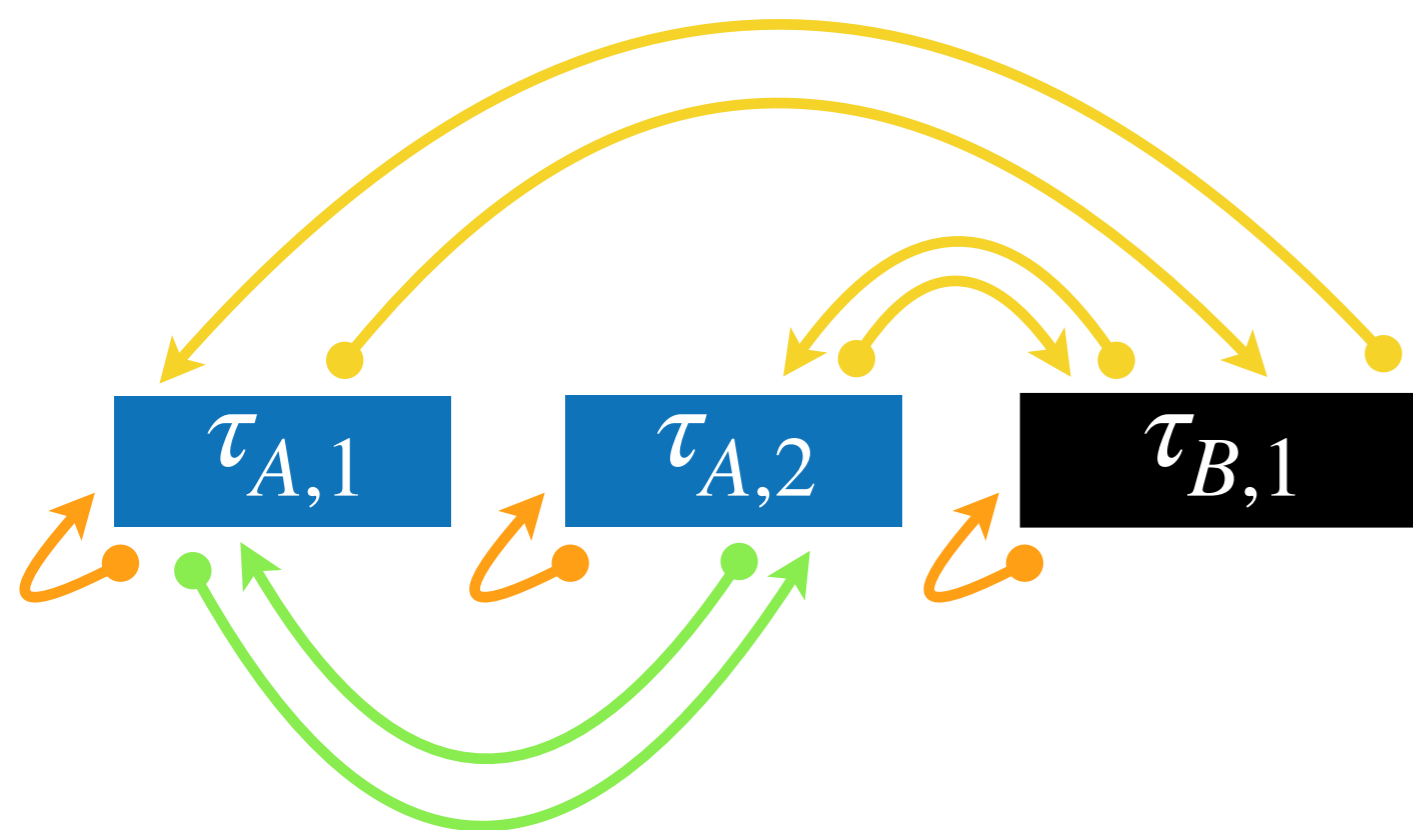
relative
deadline

$$\frac{\widehat{\mathbb{V}}[R_B]}{\widehat{\mathbb{V}}[R_B] + \left(d - \widehat{\mathbb{E}}[R_B]\right)^2}$$

upper bound on the variance of R_B

relative deadline of τ_B

upper bound on the mean of R_B



all covariance pairs

$$\widehat{\mathbb{E}}[R_B] = 2 \cdot \hat{e}_A + \hat{e}_B$$

$$\widehat{\mathbb{V}}[R_B] = 2 \cdot \hat{s}_A^2 + \hat{s}_B^2 + 2 \cdot \hat{v}_{A,A} + 4 \cdot \hat{v}_{A,B}$$

variance UBs Intra-task cov UB Inter-task cov UB

EVALUATION

How do **CAA** results compare to **CTA** in general?

We compared **CAA** and **CAA**.

A Distribution-Agnostic and Correlation-Aware Analysis of Periodic Tasks

Filip Marković¹ Georg von der Brüggen² Mario Günzel² Jian-Jia Chen² Björn B. Brandenburg¹

¹Max Planck Institute for Software Systems, Germany
²TU Dortmund, Department of Computer Science, Germany

Abstract—Real-time tasks often exhibit correlated execution-time distributions due to common factors such as shared caches, resources, and inputs. Yet state-of-the-art probabilistic analysis still overlooks the impact of correlation, a gap that has been highlighted as a major open problem in the field. This paper responds to the open problem with the first *correlation-aware analysis* (CAA) of periodic tasks with stochastic execution times. The proposed analysis, which derives response-time distributions to infer upper bounds on deadline-failure probabilities, applies to a novel task model that incorporates information about both *intra-* and *inter-task* dependencies. In addition, the paper shows how to statistically infer the two model parameters using confidence intervals obtained via nonparametric bootstrapping. Notably, the inference method described is *distribution-agnostic*, meaning that it does not assume any particular probability distribution *a priori*, thereby eliminating a major risk of misclassifying the ground-truth execution behavior. By design, CAA dominates state-of-the-art correlation-tolerant analysis (CTA). The significantly better accuracy of CAA is demonstrated via experiments with synthetically generated workloads, while a case study based on the WATERS'17 industrial challenge provides a proof-of-concept of the statistical inference method.

I. INTRODUCTION

A major challenge in the analysis of modern real-time systems is that many state-of-the-art methods, including *worst-case execution time* (WCET) analysis, prove ineffective when applied to complex software and hardware stacks. This is mainly due to unpredictable components and effects such as hardware accelerators [50], thermal noise [35], extrinsic nondeterminism in network communication protocols [30], and complex hardware and software stacks designed with *developer productivity and code reuse* as the primary goals, as opposed to *performance predictability* [52]. New approaches to analyze and mitigate the timing uncertainties inherent in modern real-time systems are thus urgently needed. In this context, probabilistic analysis has emerged as the most promising direction [16].

The two overarching (and often conflicting) goals in probabilistic analysis for real-time systems are *soundness and accuracy*. On the one hand, to ensure soundness, an analysis must not underestimate the probability of adverse events (e.g., a missed deadline). On the other hand, if the estimated probability of failure is much greater than a system's actual, ground-truth risk of failure, the analysis's excessive pessimism will result in overallocation of resources, reduced system efficiency, and ultimately increased costs and environmental impact.

A key and still largely unresolved issue at the core of the tension between analytical soundness and accuracy is the challenge of correlated execution times [16]. While it has long been

recognized that ignoring potential correlations among execution-time distributions (i.e., incorrectly assuming that all tasks are independent) can lead to optimistic (i.e., unsound) results [54], it is only recently that research has developed techniques that address the issue in a provably sound manner [5, 46].

The conventional method for circumventing correlation issues uses the notion of a task's *probabilistic worst-case execution time* (pWCET) [2, 5, 16]. This approach assumes that each task's pWCET distribution includes sufficient *padding* to account for and mask any potentially harmful dependencies on the behavior of other tasks. Although correctly padded pWCETs in principle allow the use of independence-assuming analysis methods, it has recently been observed that even just defining the concept of a pWCET is not trivial [5], let alone determining the correct amount of padding. In addition, even correctly padded pWCETs can be a challenge to use properly [12].

Recently, *correlation-tolerant analysis* (CTA) [46] has emerged as a more direct solution to the challenge of analyzing dependent tasks without resorting to false independence assumptions or relying on pWCET-based models. Notably, CTA accommodates *arbitrary* dependencies among tasks while requiring only upper bounds on the expectation and standard deviation of otherwise unknown execution-time distributions.

However, while both CTA and the careful use of padded pWCETs can ensure soundness in the presence of correlated execution times [5, 46], neither is ideal when it comes to accuracy. As we illustrate with an example in Sec. II, CTA, and even more so analyses built on the pWCET abstraction, can suffer from significant inherent pessimism because they only *tolerate or mask* correlation, rather than treating it as a first-order feature of the task model being analyzed.

Thus, recent advances [5, 46] notwithstanding, the problems of (i) statistically inferring dependencies among execution times and (ii) using this data in sound analysis remain largely open. As Davis and Cucu-Grosjean [16] highlight in their list of open issues and key challenges in probabilistic analysis:

- “How to handle issues relating to dependencies between the execution times of jobs of (i) the same task, and (ii) jobs of different tasks? The impact of these dependencies may vary based on how strong they are.” [16]
- “Appropriate statistical studies are needed to investigate the types of dependencies and their impact on probabilistic schedulability analysis. Analyses are needed that can address dependencies.” [16]

We propose the first solutions to both problems in this paper.

[4] Marković et al. *RTSS* (2024)

CTA: A Correlation-Tolerant Analysis of the Deadline-Failure Probability of Dependent Tasks

Filip Marković¹ Pierre Roux² Sergey Bozhko^{1,3} Alessandro V. Papadopoulos⁴ Björn B. Brandenburg¹

¹Max Planck Institute for Software Systems (MPI-SWS), Germany
²ONERA/DTIS, Université de Toulouse, France
³Saarbrücken Graduate School of Computer Science, Saarland University, Germany
⁴Mälardalen University (MDU), Sweden

Abstract—Estimating the *worst-case deadline failure probability* (WCDFP) of a real-time task is notoriously difficult, primarily because a task's execution time typically depends on prior activations (i.e., history dependence) and the execution of other tasks (e.g., via shared inputs). Previous analyses have either assumed that execution times are probabilistically independent (which is unrealistic and unsafe), or relied on complex upper-bounding abstractions such as *probabilistic worst-case execution time* (pWCET), which mask dependencies with pessimism. Exploring an analytically novel direction, this paper proposes the first closed-form upper bound on WCDFP that accounts for dependent execution times. The proposed *correlation-tolerant analysis* (CTA), based on Cantelli's inequality, targets fixed-priority scheduling and requires only two basic summary statistics of each task's ground-truth execution time distribution: upper bounds on the mean and standard deviation (for any possible job-arrival sequence). Notably, CTA does not use pWCET, nor does it require the full execution-time distribution to be known. Core parts of the analysis have been verified with the Coq proof assistant. Empirical comparison with state-of-the-art WCDFP analyses reveals that CTA can yield significantly improved bounds (e.g., a lower WCDFP than any pWCET-based method for $\approx 70\%$ of the workloads tested at 90% pWCET utilization and 60% average utilization). Beyond accuracy gains, the favorable results highlight the potential of the previously unexplored analytical direction underlying CTA.

I. INTRODUCTION

Probabilistic analysis of real-time systems holds the promise of addressing the central challenge of modern hardware and software architectures: *unavoidable uncertainty* in the execution behavior of real-time tasks. Such uncertainty, deeply embedded in the fabric of modern computing systems, more often than not precludes meaningful (classical) worst-case analysis, leaving a stochastic perspective as the only viable option.

One of the most pressing open problems in this space is the issue of *dependent* execution times (also referred to as *execution-time correlation*). Specifically, when bounding a task's *worst-case deadline-failure probability* (WCDFP), it is crucial to account for possible dependencies on both previous activations (*intra-task dependence*) and other tasks in the system (*inter-task dependence*). If such dependencies are ignored, the WCDFP may be severely under-approximated.

These observations are not new: the lack of independence in practice was recognized as a safety problem already more than 25 years ago by Tia et al. [49] in one of the first works on probabilistic schedulability analysis. Unfortunately, only little progress has been made on this issue since Tia et al.'s

observation, with Davis and Cucu-Grosjean noting in the closing remarks of their recent survey [19]: “*Issues of dependence are of great importance in probabilistic schedulability analysis [...] Analyses are needed that can address dependencies*”.

Prior attempts at tackling dependence in state-of-the-art WCDFP analyses have relied on over-approximation. The common idea in this line of work is to “pad” the ground-truth execution-time distributions with “sufficient pessimism,” to the point that task behavior can be safely assumed to be independent. The primary mechanism for realizing such an analysis in a sound manner is the concept of a *probabilistic worst-case execution time* (pWCET) distribution [5, 8, 14, 17, 18], which can be determined for each task either via static analyses [e.g., 4, 6, 16, 31] or with measurement-based techniques such as *extreme value theory* (EVT) [e.g., 32, 33, 46, 47].

Specifically, the pWCET approach promises that the analysis may model execution times with independent random variables following the pWCET distribution, provided the pWCET distribution is suitably determined [19]. However, a significant limitation of such *independence-assuming analysis* (IAA) lies in its inherent over-approximation of the ground truth, which can lead to considerable pessimism compared to actual behavior.

This paper. Exploring a fundamentally different direction, we propose a novel *correlation-tolerant analysis* (CTA) of WCDFP under fixed-priority scheduling. CTA is based on Cantelli's inequality [9] and departs from the state of the art in three major ways: first, CTA does not use pWCET, nor does it otherwise require ground-truth distributions to be pessimistically padded; second, unlike traditional methods, CTA does not require full knowledge of the ground-truth distributions, as it uses only bounds on their means and standard deviations (under any possible job-arrival sequence); and last but not least, CTA is safe in the presence of arbitrarily dependent execution times. Notably, CTA also does not require the degree of inter- or intra-task correlation to be quantified, which is desirable in practice.

In developing CTA, we make the following contributions:

- We convey the core idea with a simple example (Sec. II).
- From Cantelli's inequality [9], we derive, and verify with Coq [13, 41], an upper bound on the sum of random variables with unknown degrees of correlation (Sec. IV).
- We formally model the execution of a stochastic sporadic real-time workload under preemptive uniprocessor fixed-

[3] Marković et al. *RTSS* (2023)

EVALUATION

*Synthetic task sets were randomly generated to highlight differences between **CTA** and **CAA**.*

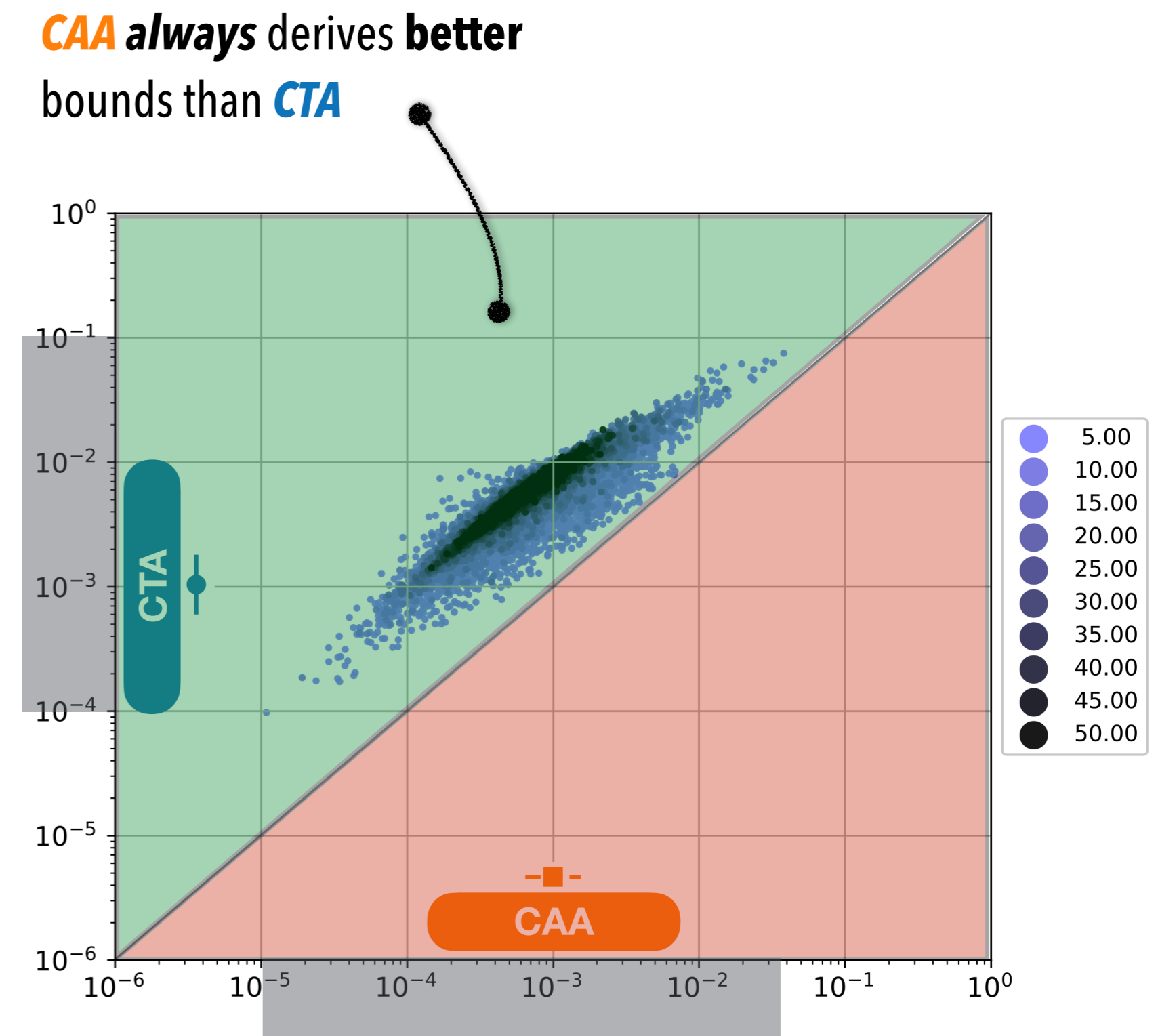
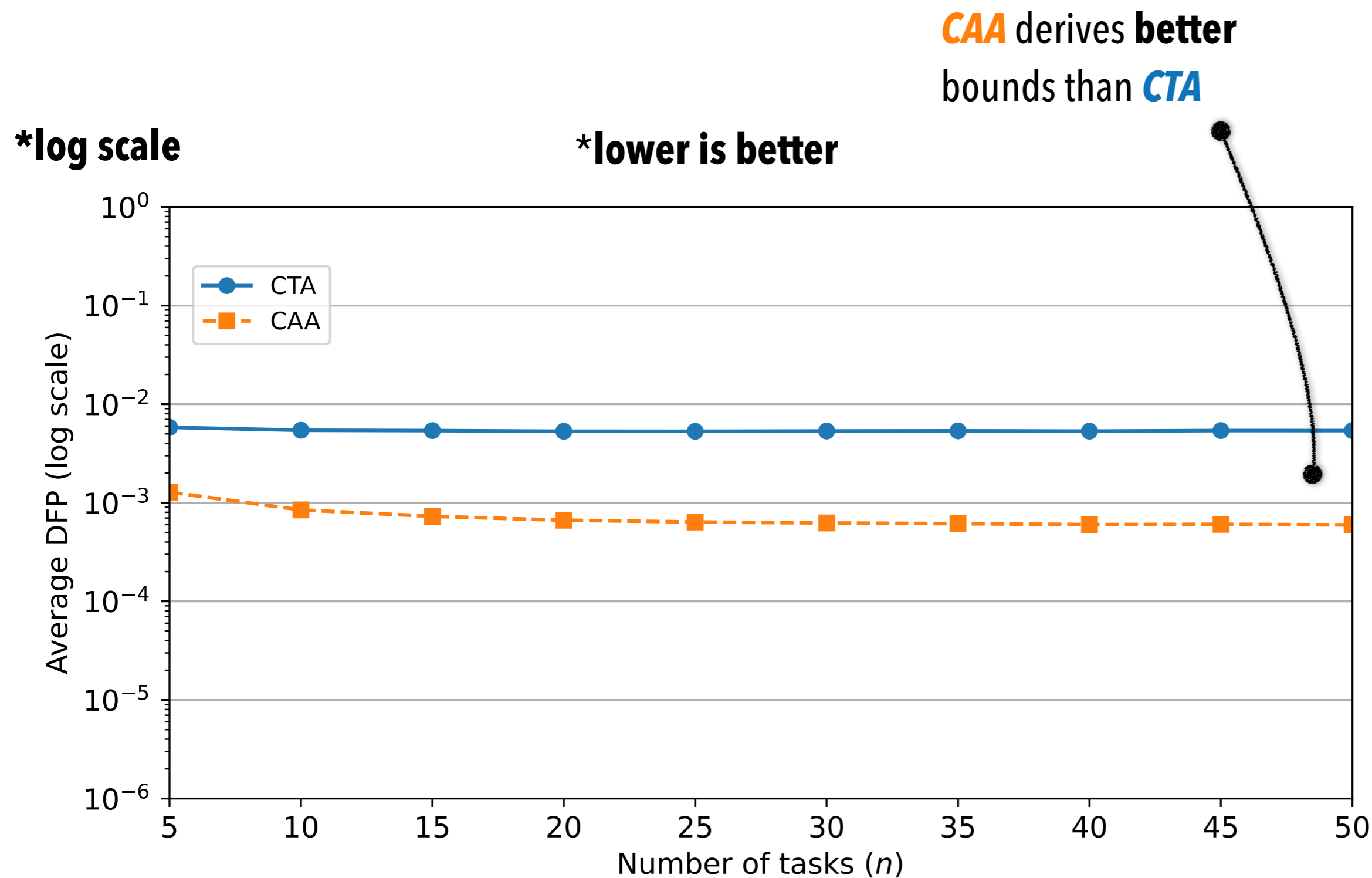
Four experiments were conducted to investigate:

1. Influence of the **task set size** on DFP,
2. The influence of the **total mean utilization** on DFP,
3. The influence of the **maximum standard deviation** on DFP,
4. The influence of the **maximum correlation** on DFP.

In this talk, we focus on (1).

EVALUATION

Investigating the influence of the task-set size



CAA dominates CTA, i.e., it always computes better or, at worst, equal DFP estimates

The gap becomes slightly more pronounced as the number of tasks increases due to **Bienaymé's** identity

$$\widehat{\mathbb{V}}[R_B] = 2 \cdot \widehat{\sigma}_A^2 + \widehat{\sigma}_B^2 + 2 \cdot \widehat{\nu}_{A,A} + 4 \cdot \widehat{\nu}_{A,B}$$

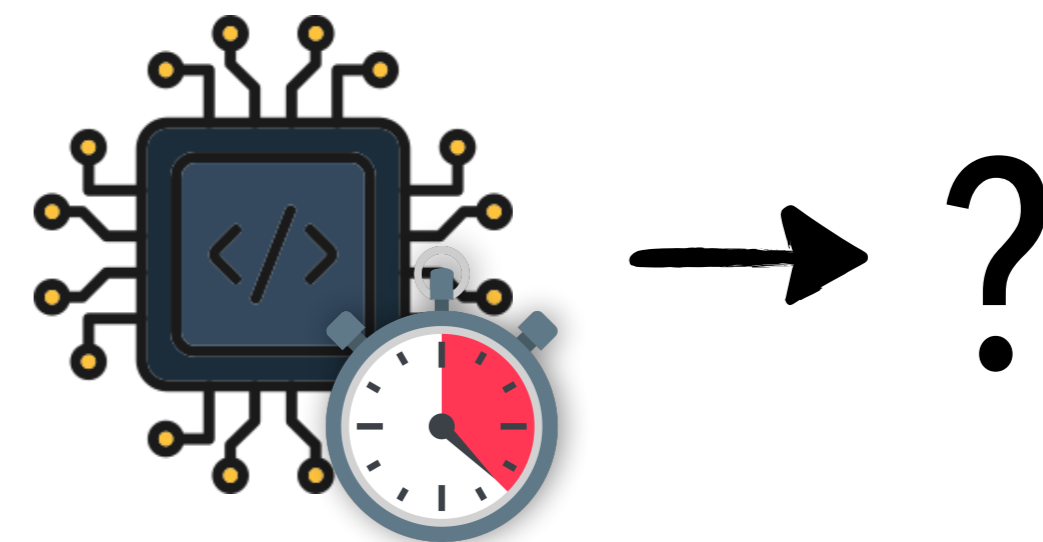
$$\widehat{\mathbb{V}}^{CTA}[R_B] = 2 \cdot \widehat{\sigma}_A^2 + \widehat{\sigma}_B^2 + 2 \cdot \widehat{\sigma}_A^2 + 4 \cdot \widehat{\sigma}_A \cdot \widehat{\sigma}_B$$

BUT, HOW DO WE DERIVE CAA (AND CTA) INPUTS? THIS PAPER (PART 2)

HOW TO STATISTICALLY INFER DEPENDENCE?

“Appropriate statistical studies are needed to investigate the types of dependences and their impact on probabilistic schedulability analysis”

[2]



ANOTHER PROBLEM: DISTRIBUTION MISCLASSIFICATION

Incorrectly assuming the *wrong* underlying distribution can lead to *unsound* results

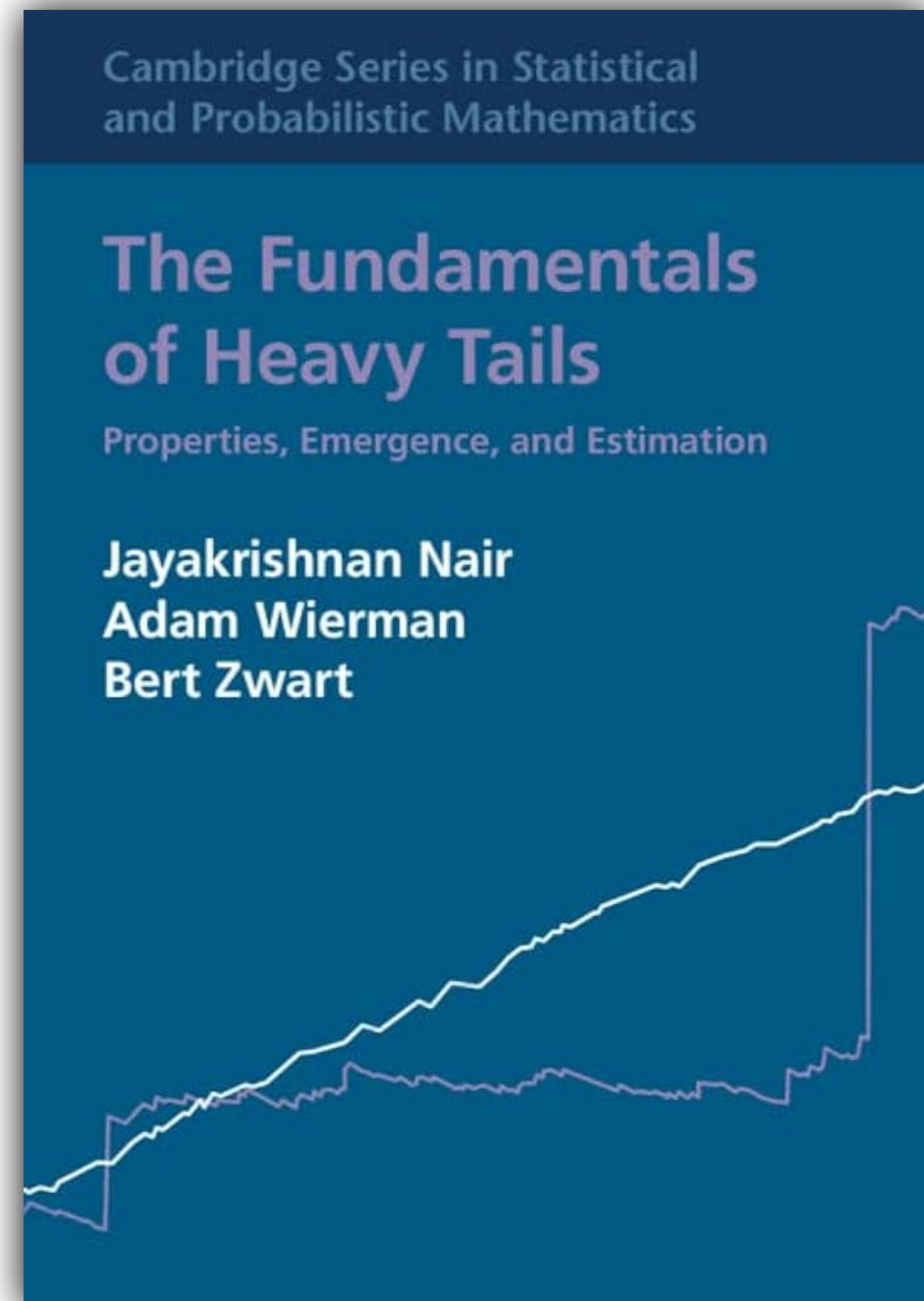
Many **refuted** results!

◇ **heavy-tailed distributions**

As explained in [The Fundamentals of Heavy Tails](#), 2022, Jayakrishnan et al.



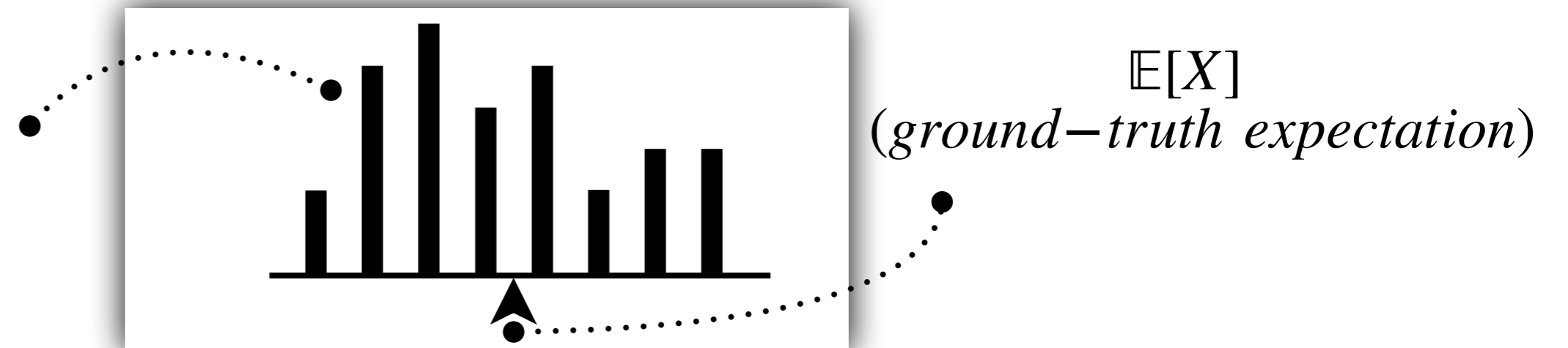
Since our research focuses on safety properties, it is crucial to **eliminate the risk of **distribution misestimation**.**



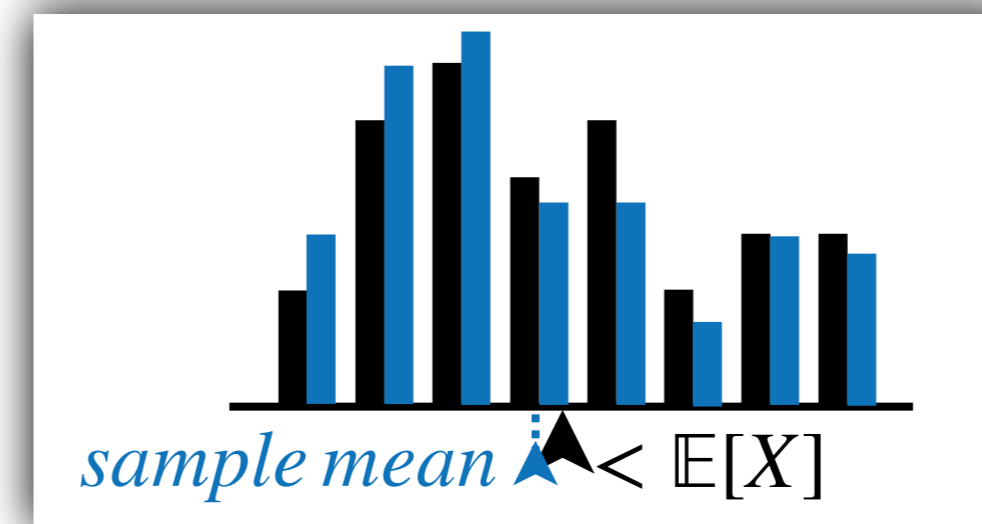
GOAL: DISTRIBUTION-AGNOSTIC STATISTICAL INFERENCE

We use Nonparametric Bootstrap

A random variable with an **unknown** distribution and **unknown** expectation



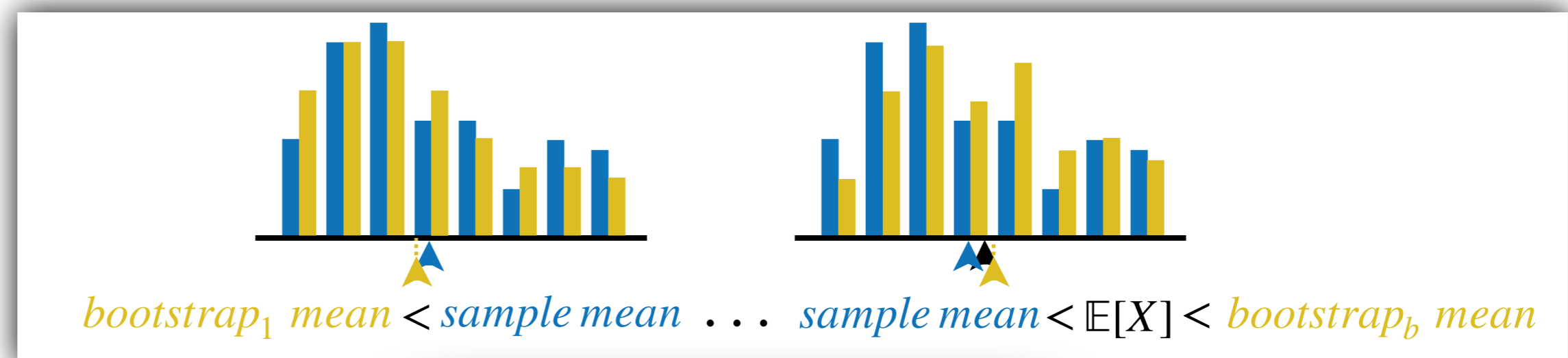
Step 1. Draw an initial sample I of n independent observations of X



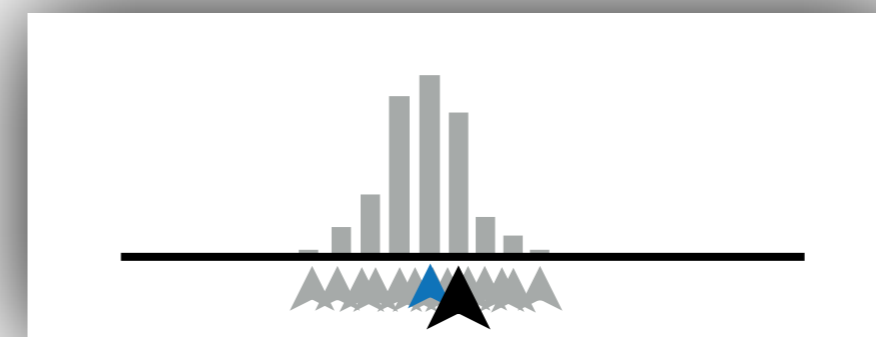
Step 2. Generate b bootstrap samples (by randomly resampling with replacement from the initial sample I)



Step 3. Compute the bootstrap statistic on each bootstrap sample B



Output: Bootstrap distribution of the expectation

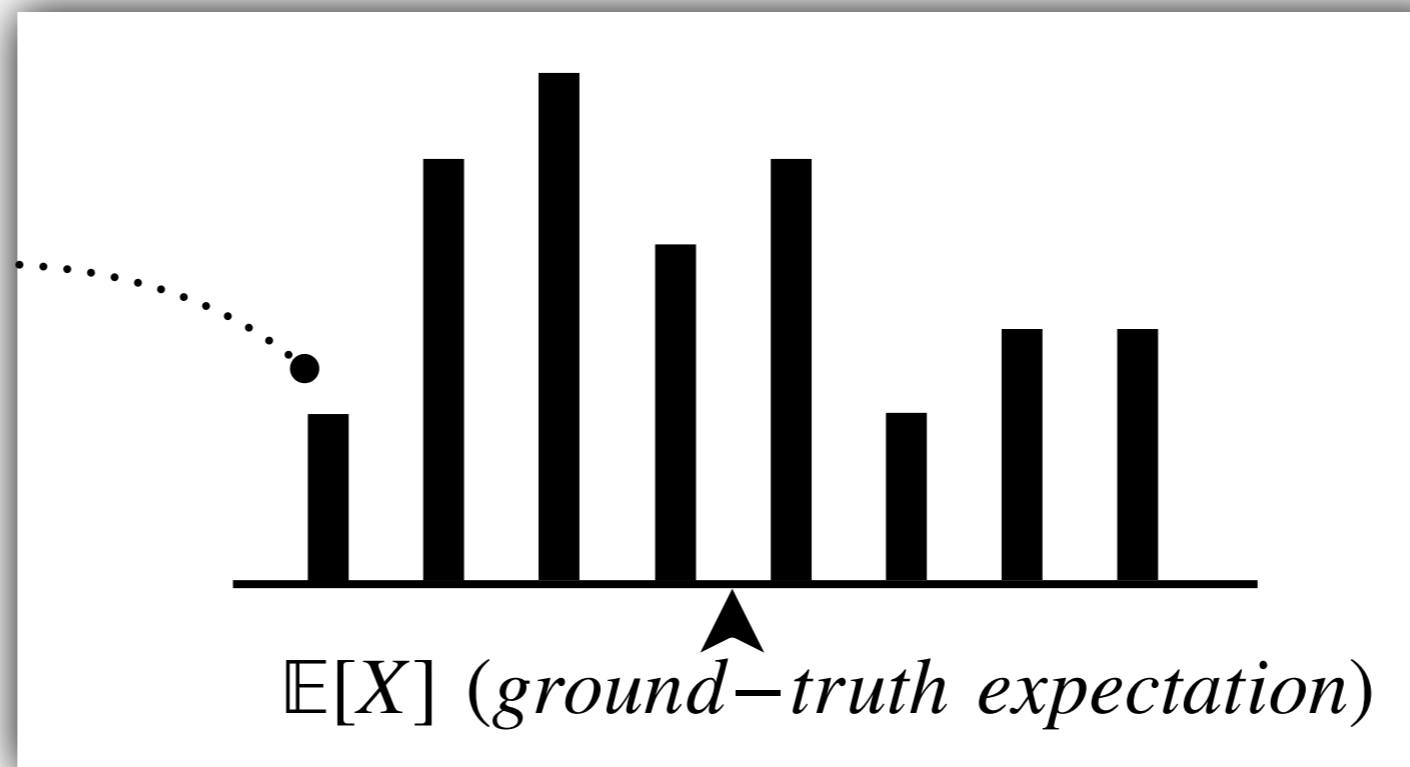


But how do we obtain sound upper bounds from the bootstrap distribution?

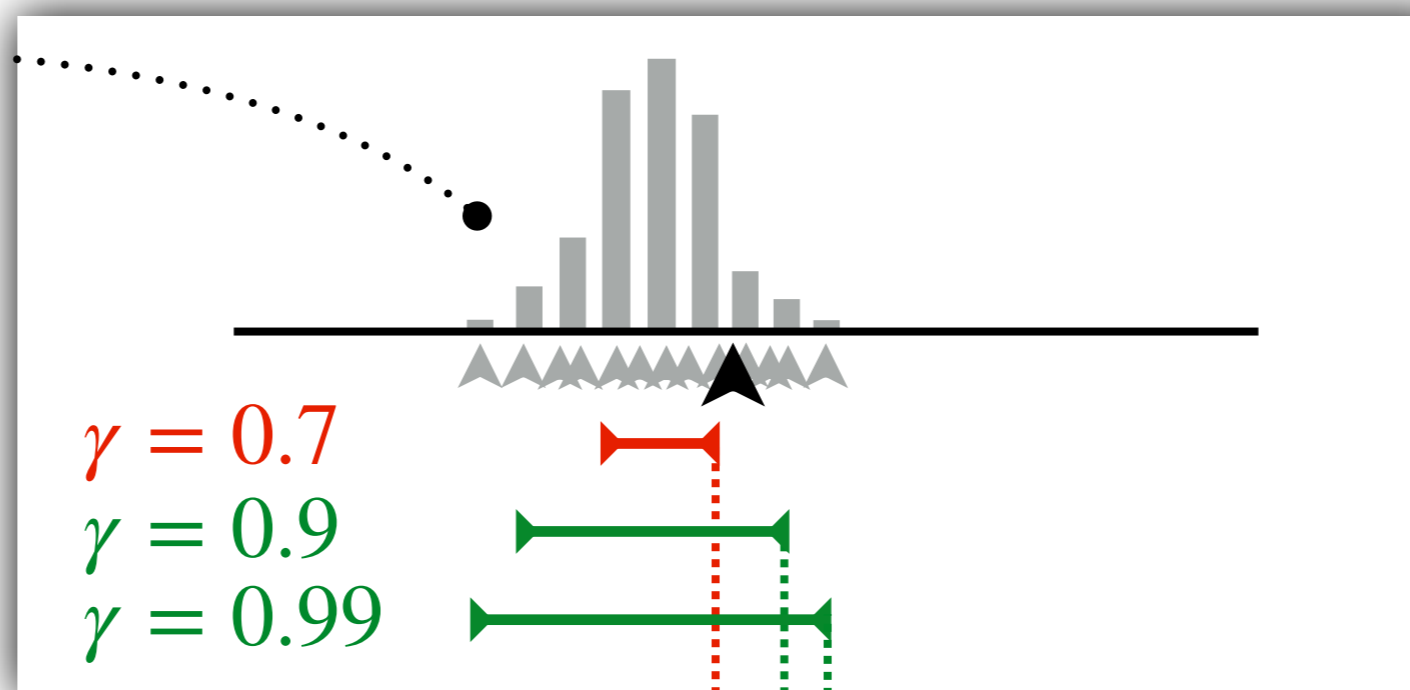
CONFIDENCE INTERVALS

The key to deriving **upper bounds** on parameters from an unknown distribution

A random variable X with an **unknown** distribution and **unknown** expectation

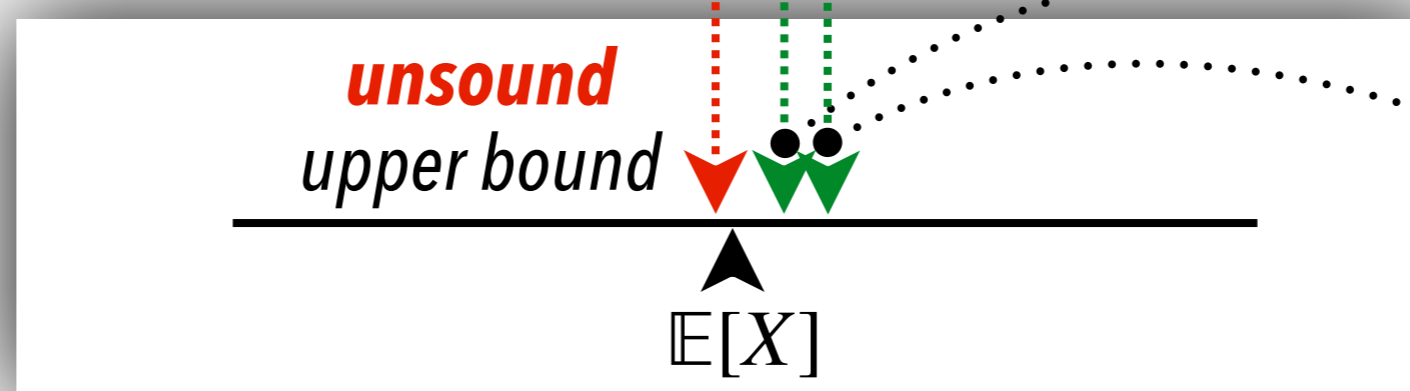


Bootstrap distribution of the expectation



Step 4. Compute the confidence interval with a given level of confidence γ

Output: The **upper bound** of the confidence interval.



sound upper bound

sound but overly pessimistic upper bound

WHAT ARE THE GUARANTEES?

How should we interpret the derived results?

Wrong confidence interval interpretation

- ▶ “A common misunderstanding about CIs is that for say a 95% CI (A to B), there is a 95% probability that the true population mean lies between A and B.”

[5]

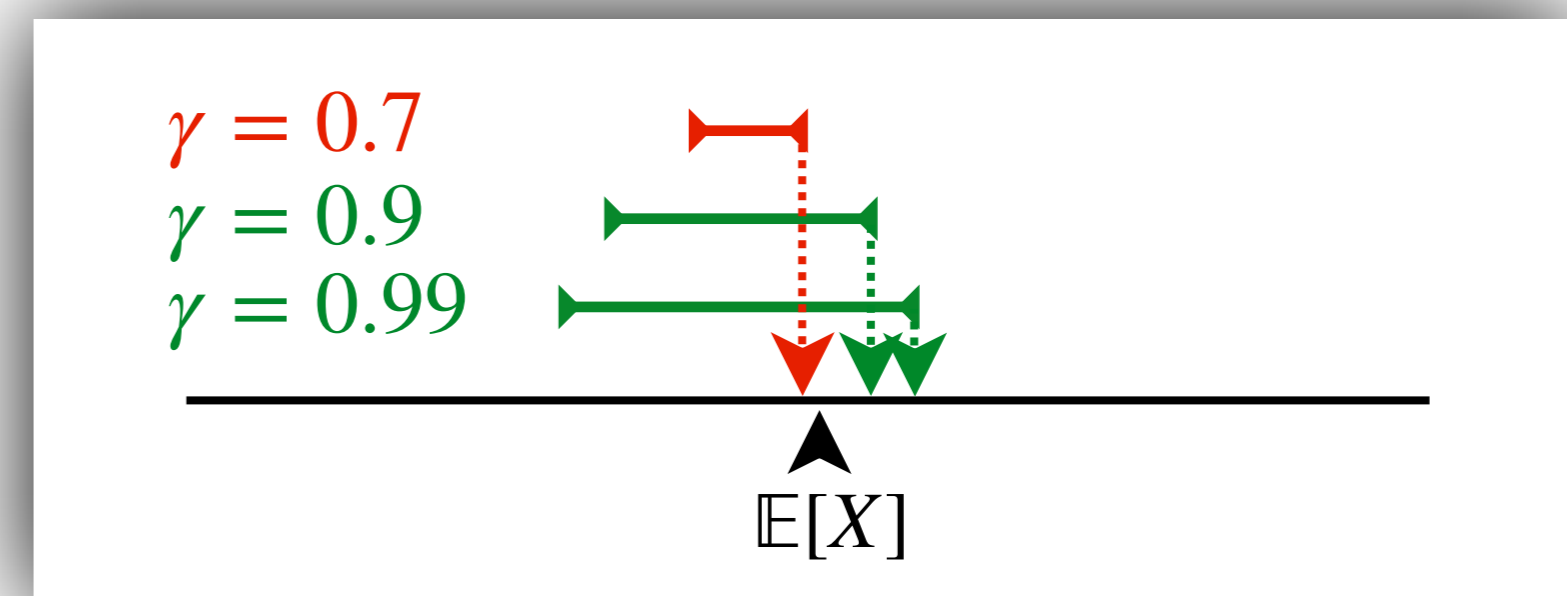
Correct confidence interval interpretation

- ▶ “A 95% CI simply means that if the study is conducted multiple times (multiple sampling from the same population) with corresponding 95% CI for the mean constructed, we expect 95% of these CIs to contain the true population mean”

[5]

- ▶ **No** statistical inference method can provide **absolute certainty**
- ▶ There is always a minuscule but **non-zero chance** that a ground-truth parameter **lies outside** the statistically estimated range

- ▶ Bootstrapped CIs provide an excellent means of estimating ground truth that is
 - ▶ **statistically rigorous**
 - ▶ **distribution-agnostic**
 - ▶ **sample-efficient**
 - ▶ **mathematically well-understood**



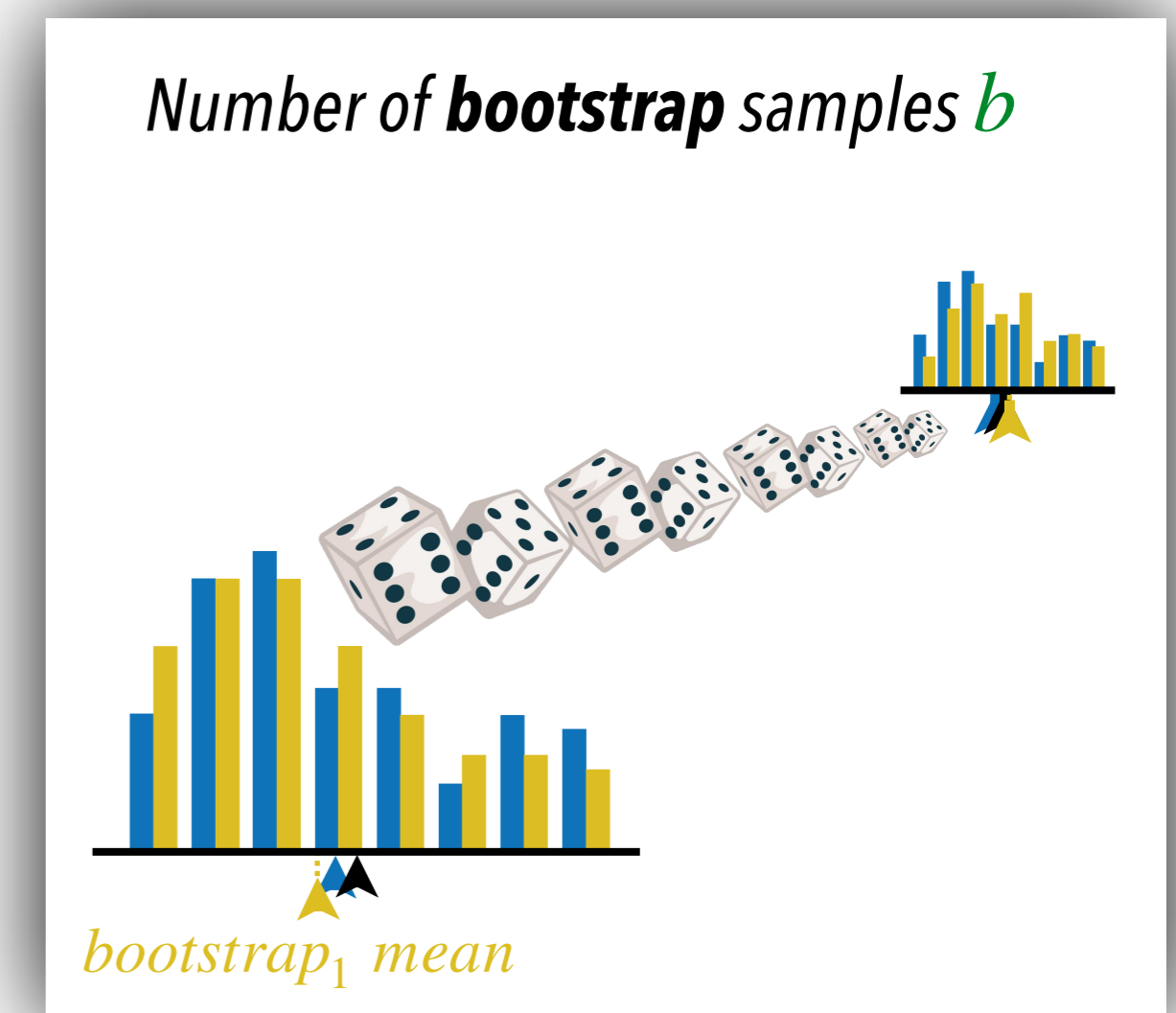
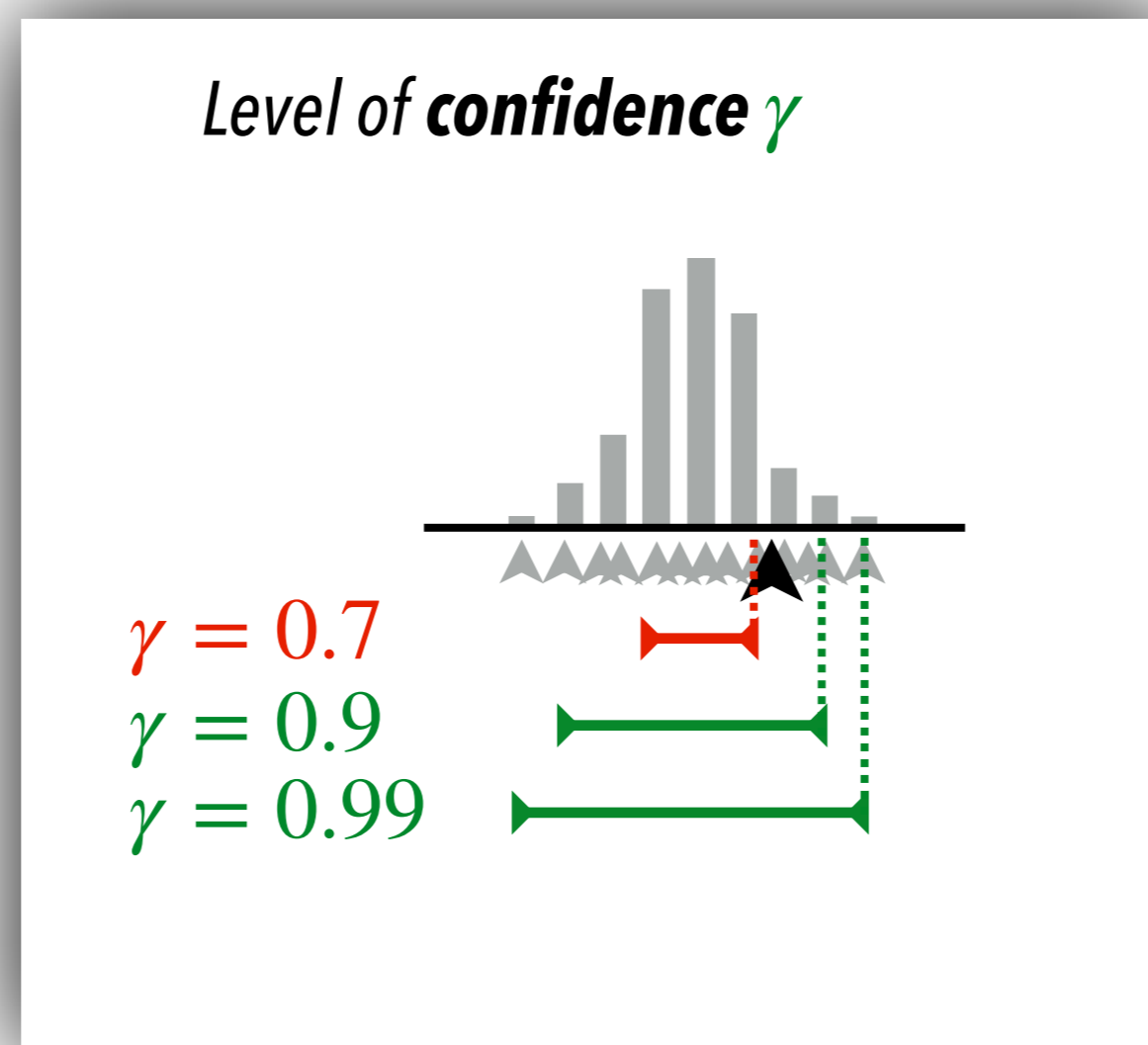
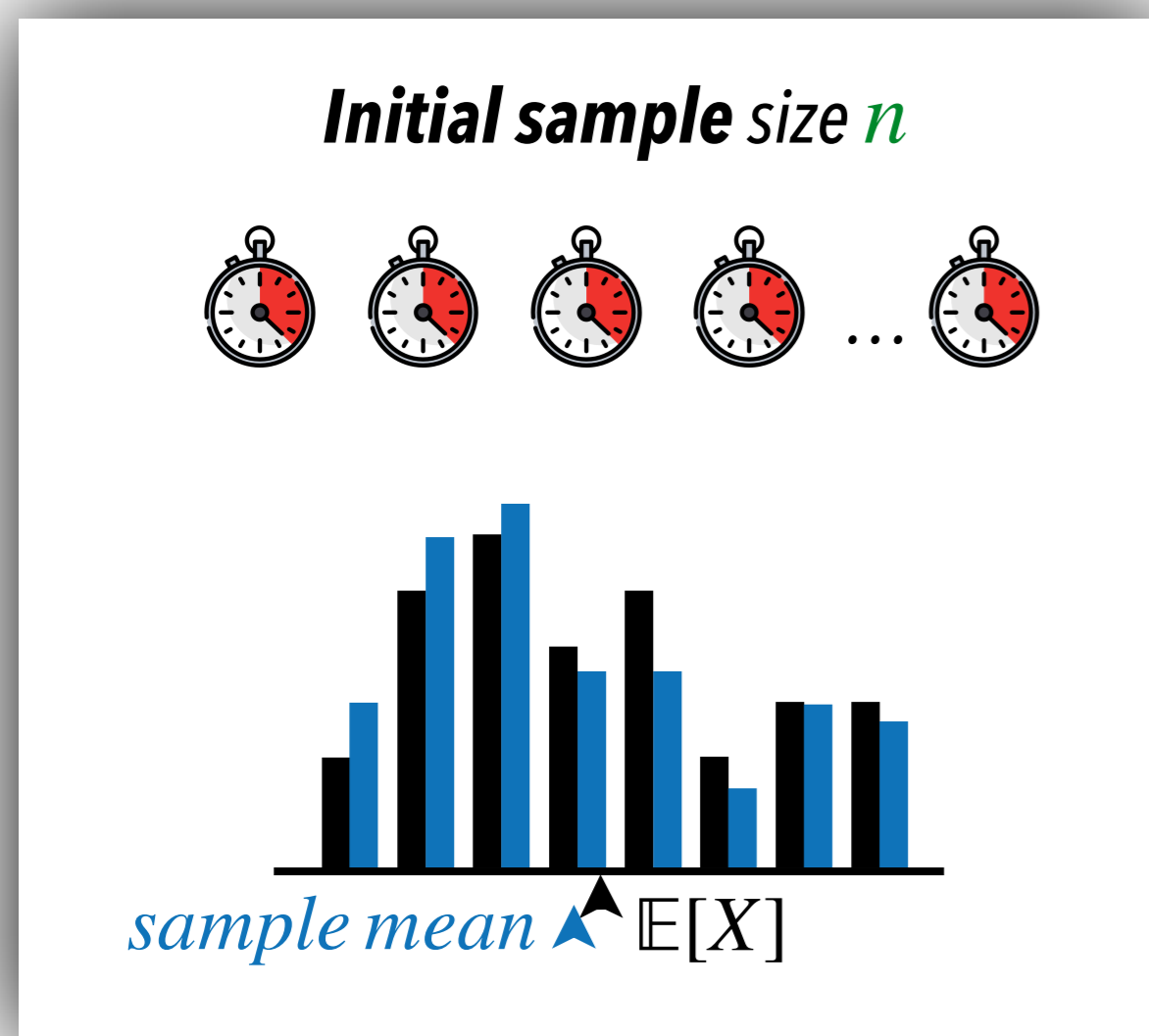
[5] “The correct interpretation of confidence intervals” Tan and Tan. *Proceedings of Singapore Healthcare* (2010)

CONFIDENCE-BASED TASK PARAMETERS

All CAA and CTA inputs are obtained with bootstrapping from bounds on confidence intervals

- ▶ \hat{e}_i : upper bound on the **mean** execution time (ET)
- ▶ \hat{s}_i : upper bound on the **standard deviation** of ET
- ▶ $\hat{v}_{i,i}$: upper bound on the ET **intra-task covariance**
- ▶ $\hat{v}_{i,k}$: upper bound on the ET **inter-task covariance**

Important tuning knobs



CASE STUDY

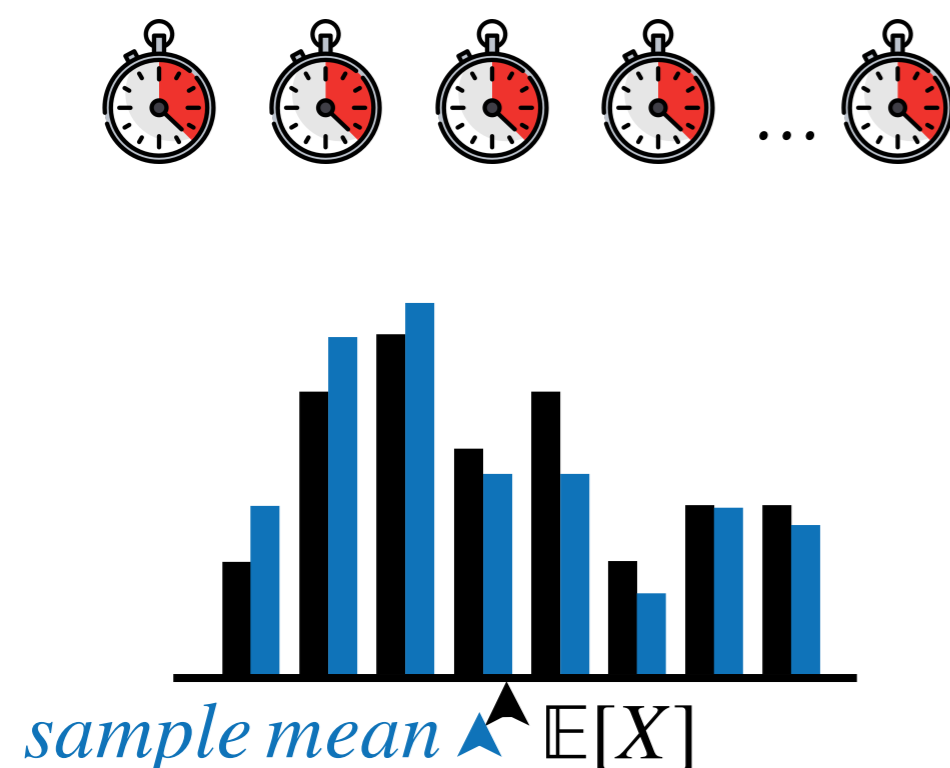
*How do CAA results compare to CTA **when using statistical inference**?*

Three experiments on a **proof-of-concept** case study from WATERS 2017 workloads were conducted to investigate:

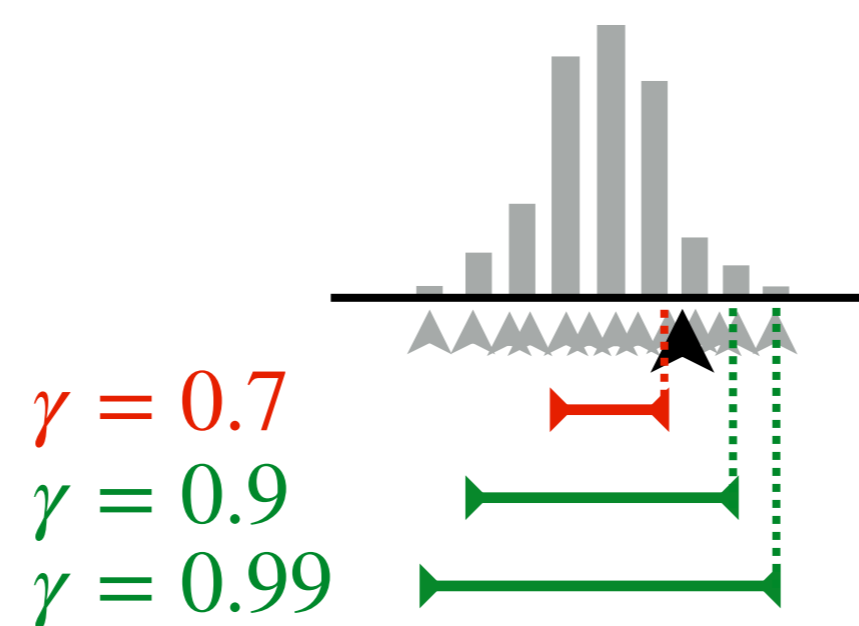
1. Influence of the **initial sample size** on DFP,
2. Influence of the **level of confidence** on DFP,
3. Influence of the **number of bootstrap samples** on DFP.

Important tuning knobs

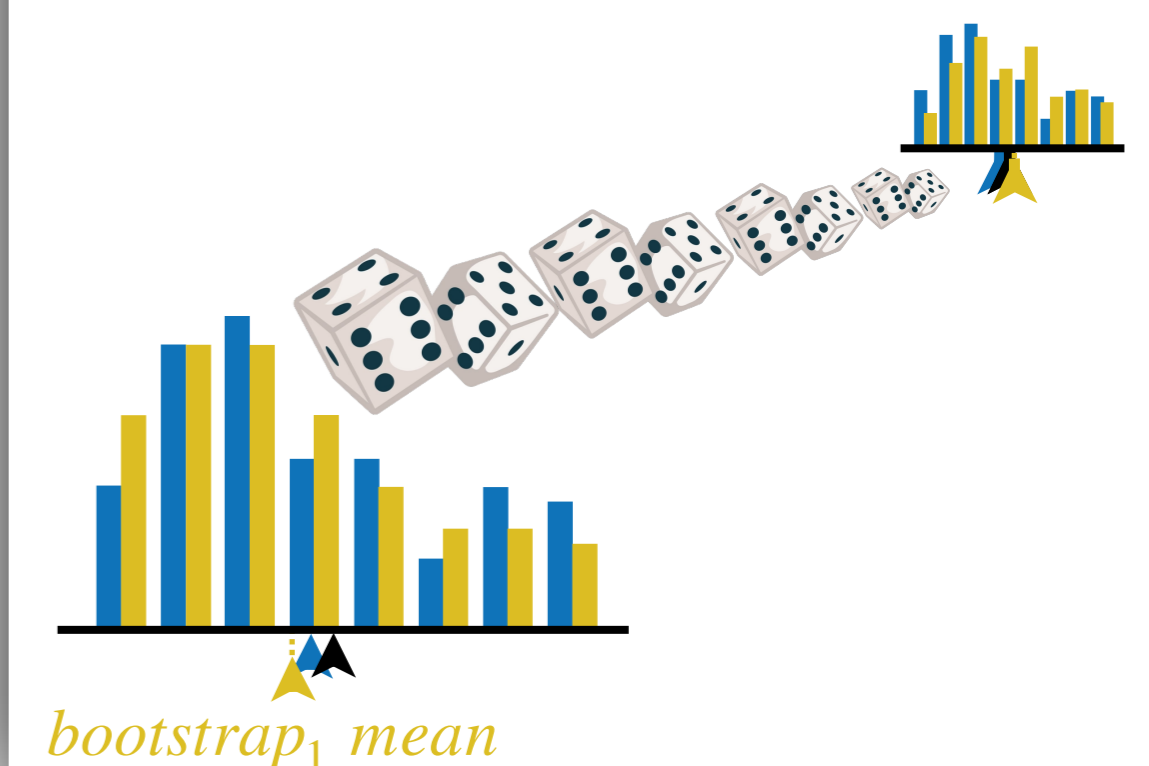
Initial sample size n



Level of confidence γ

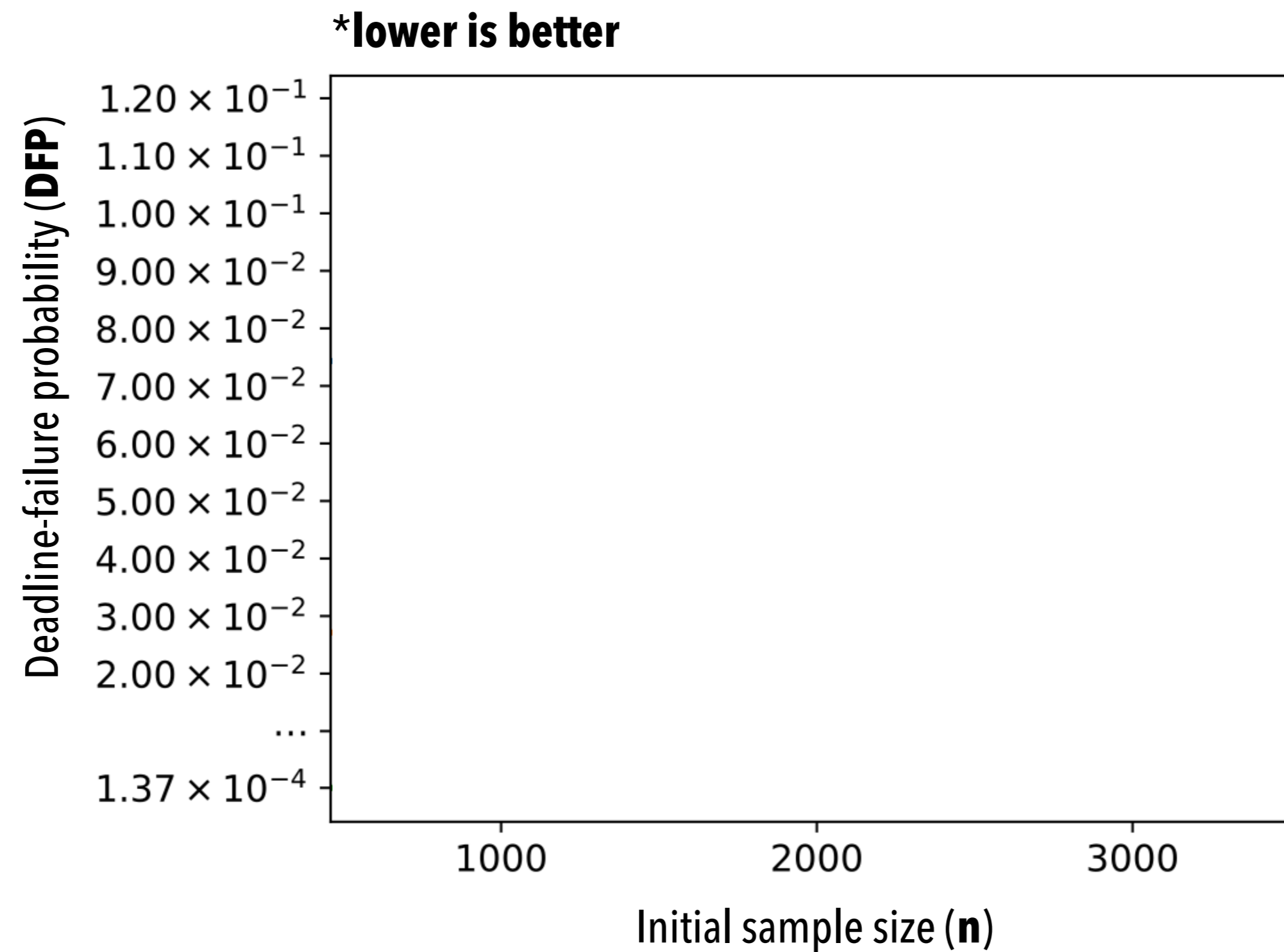


Number of bootstrap samples b



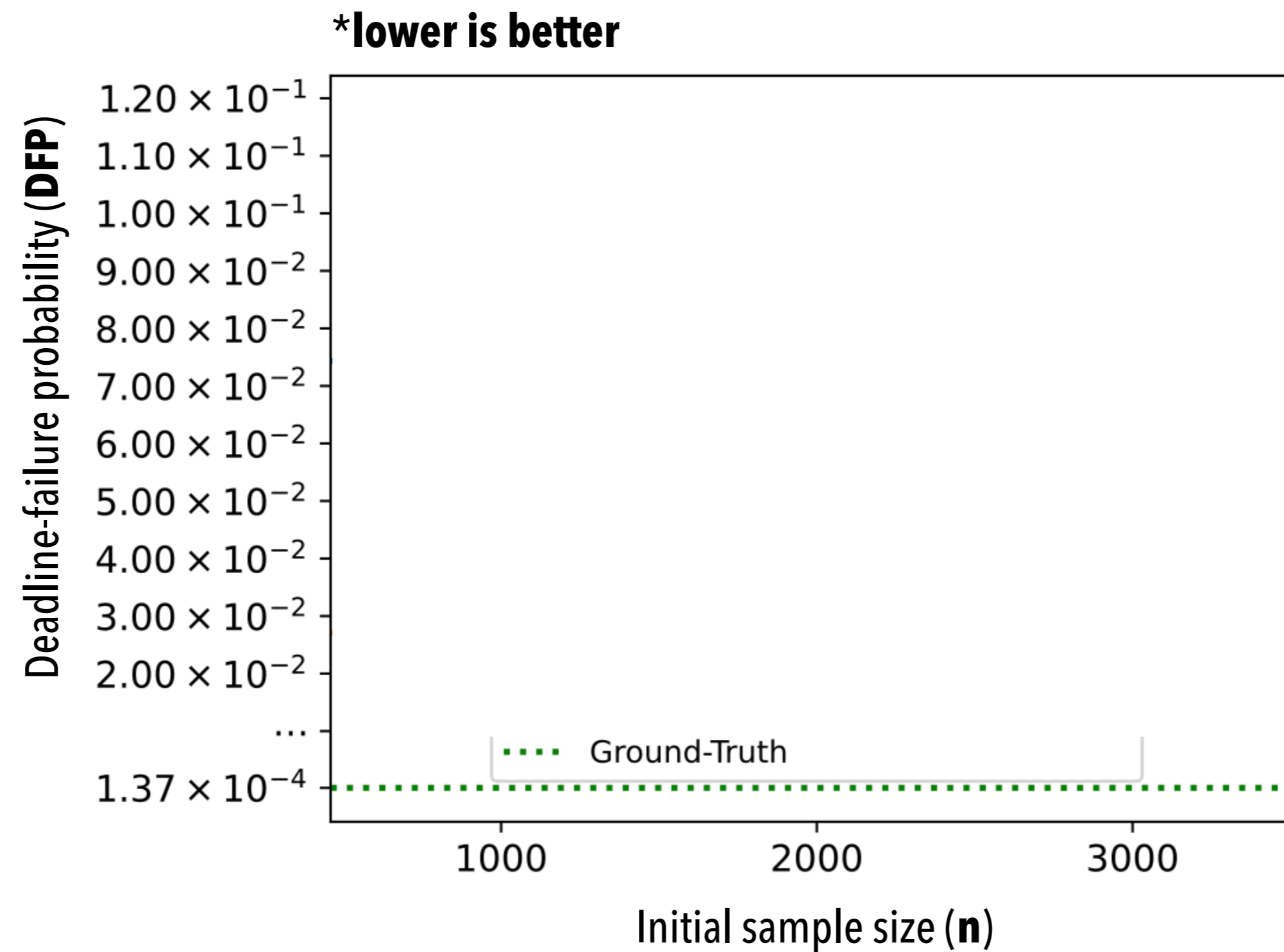
CASE STUDY

*Investigating the influence of the **initial sample size***



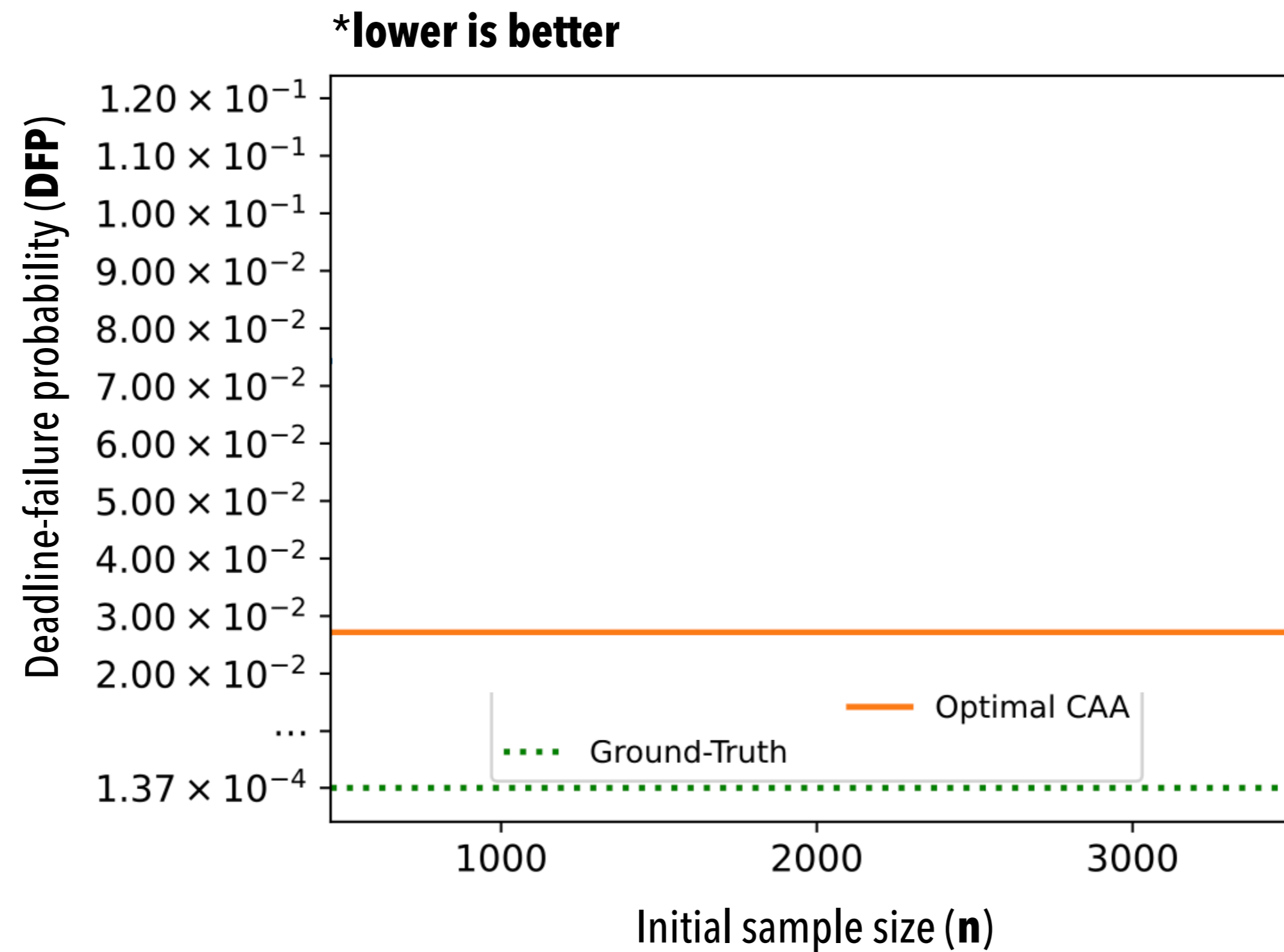
CASE STUDY

*Investigating the influence of the **initial sample size***



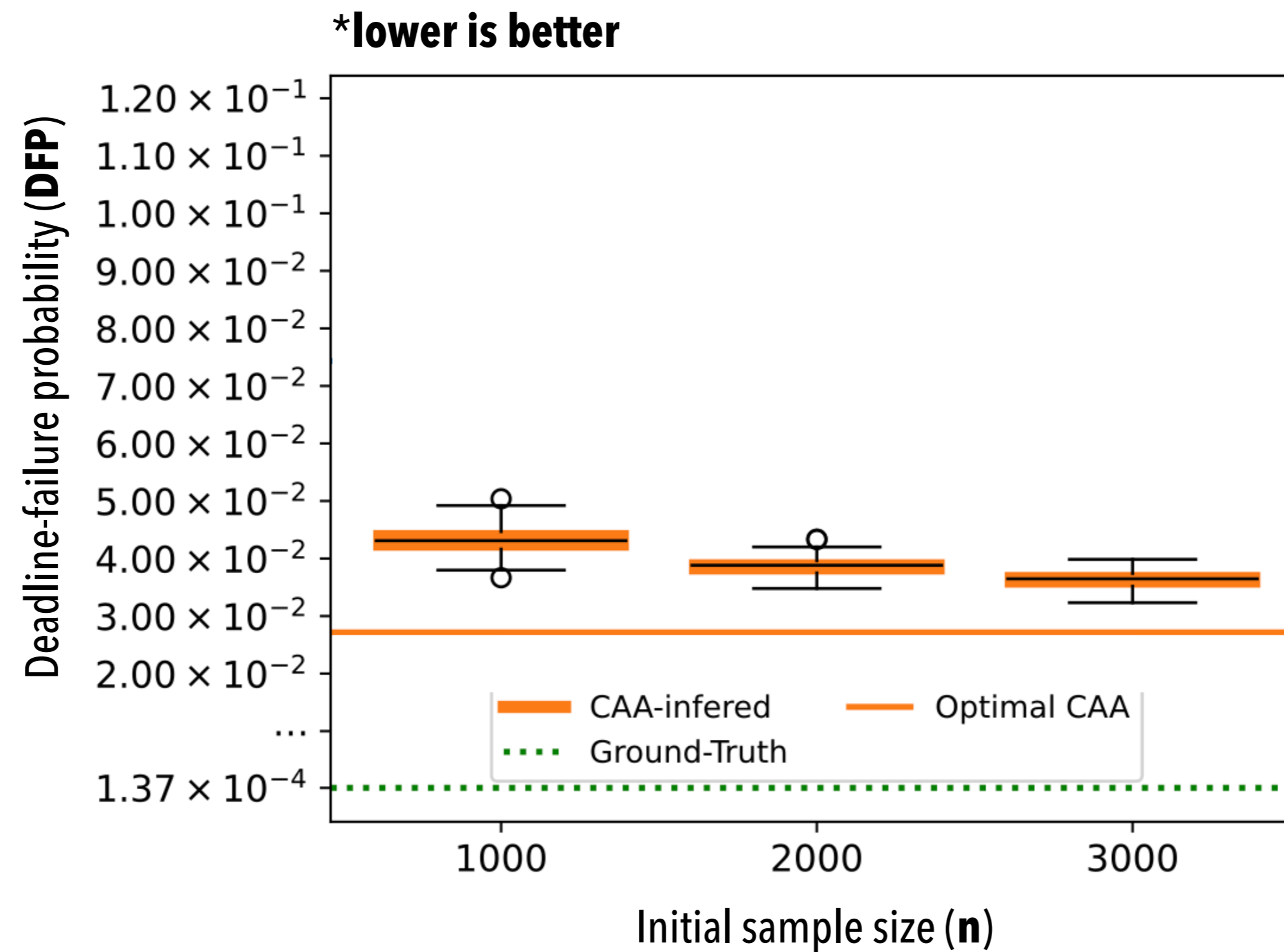
CASE STUDY

*Investigating the influence of the **initial sample size***



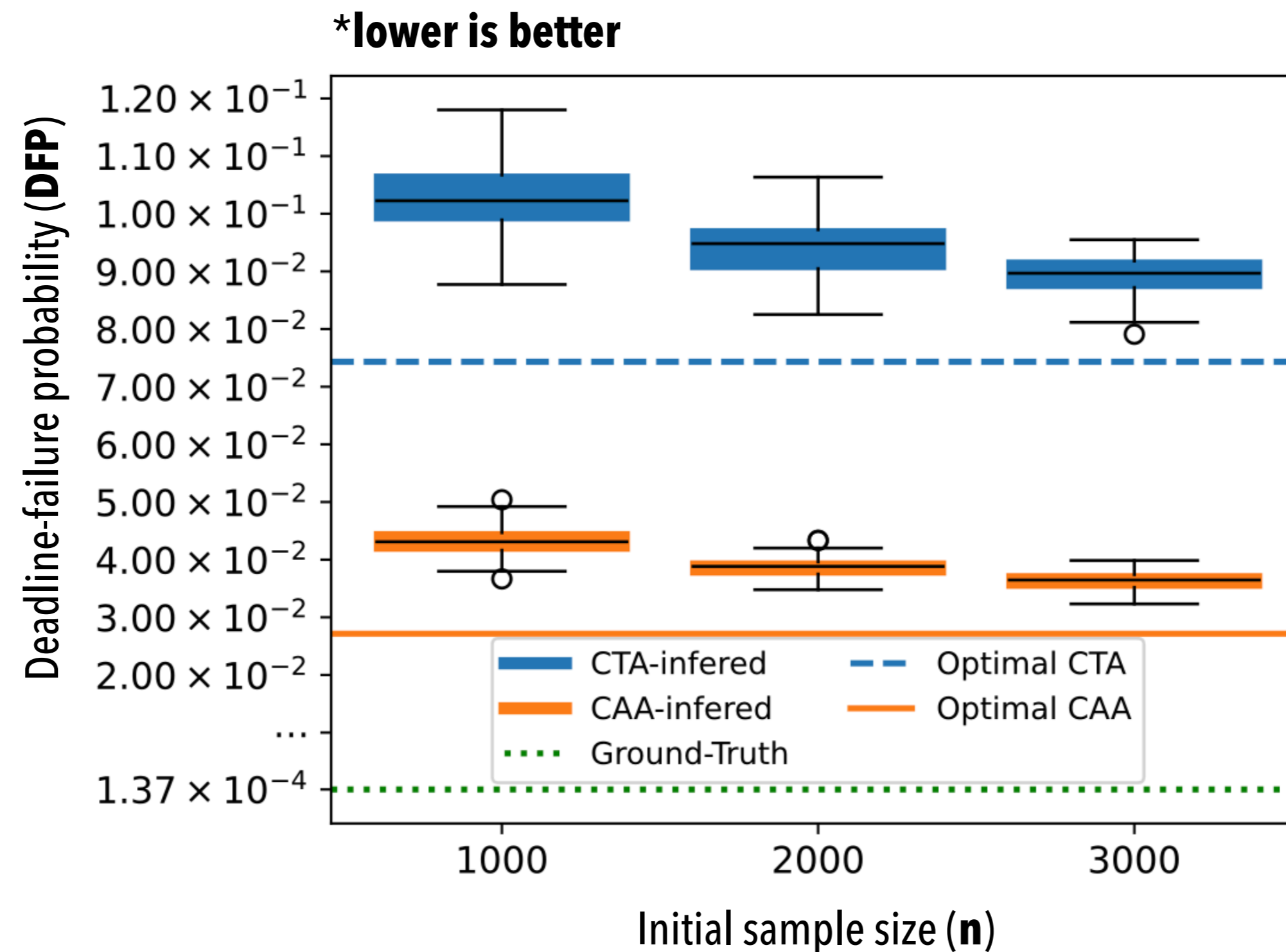
CASE STUDY

*Investigating the influence of the **initial sample size***



CASE STUDY

*Investigating the influence of the **initial sample size***

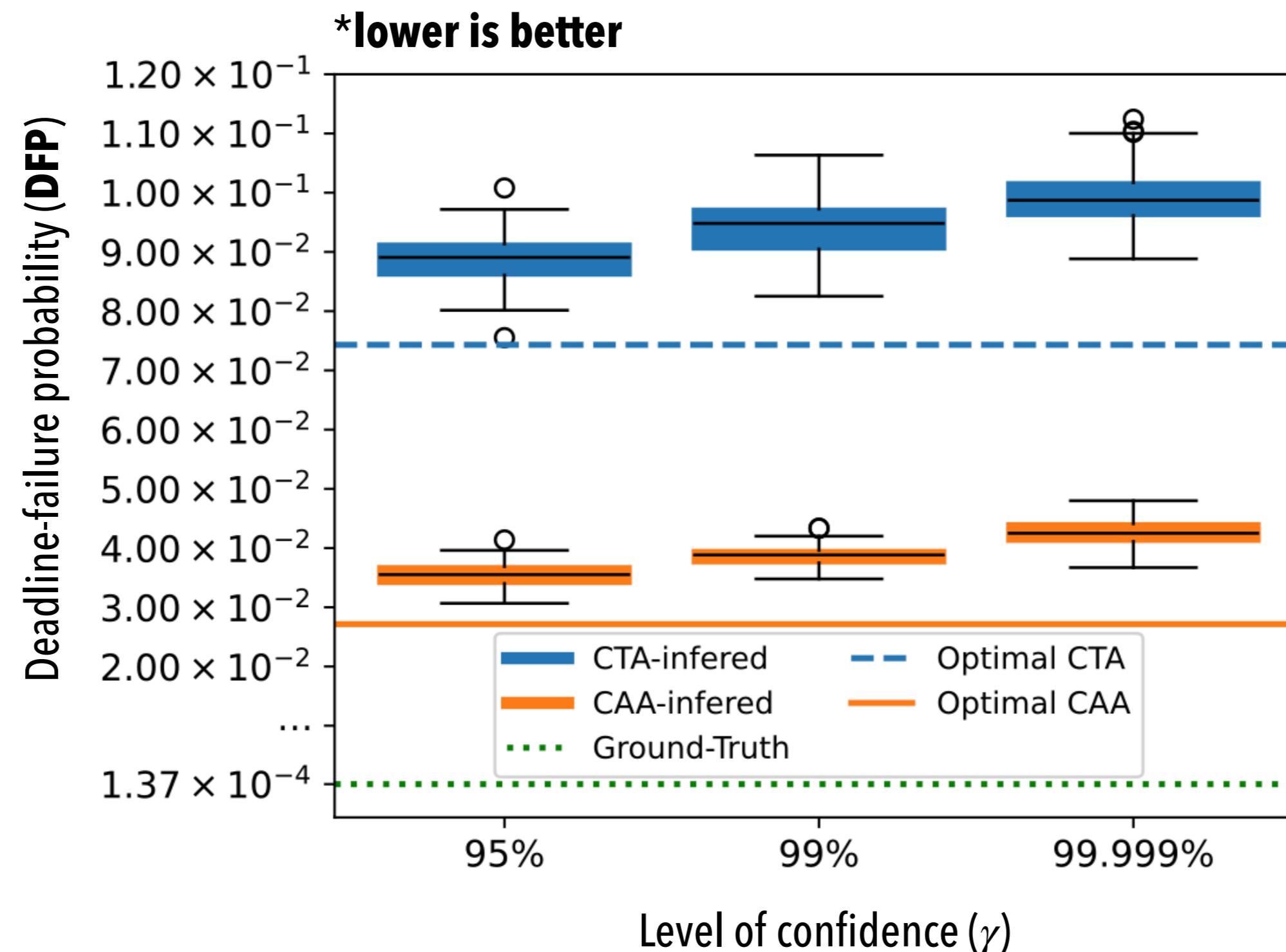


All DFP estimates are sound.

Both **CAA** and **CTA** tend towards their respective optimal DFP estimates as the initial sample size increases.

CASE STUDY

*Investigating the influence of the **level of confidence***



All DFP estimates are sound.

Increasing the level of **confidence** leads to **more conservative** results for both analyses.

SUMMARY

CONTRIBUTIONS

HOW TO MODEL DEPENDENCE?

“How to handle ... dependences between the execution times of (i) jobs of the same task, and (ii) jobs of different tasks?”

HOW TO QUANTIFY DEPENDENCE?

“The impact of these dependences may vary based on how strong they are.”

HOW TO USE THESE IN RT ANALYSIS?

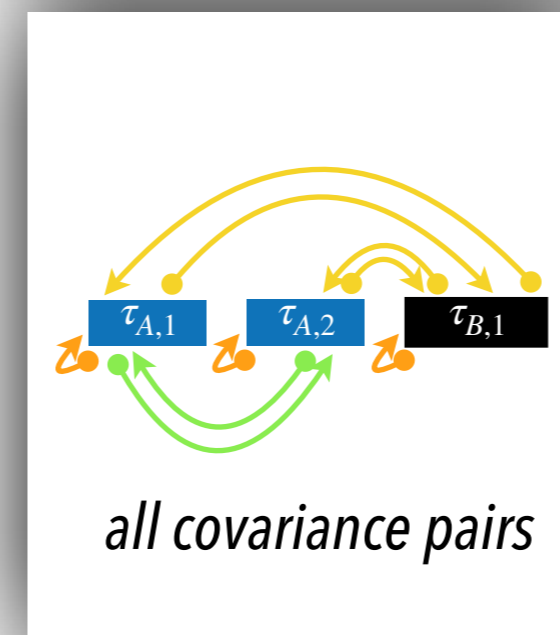
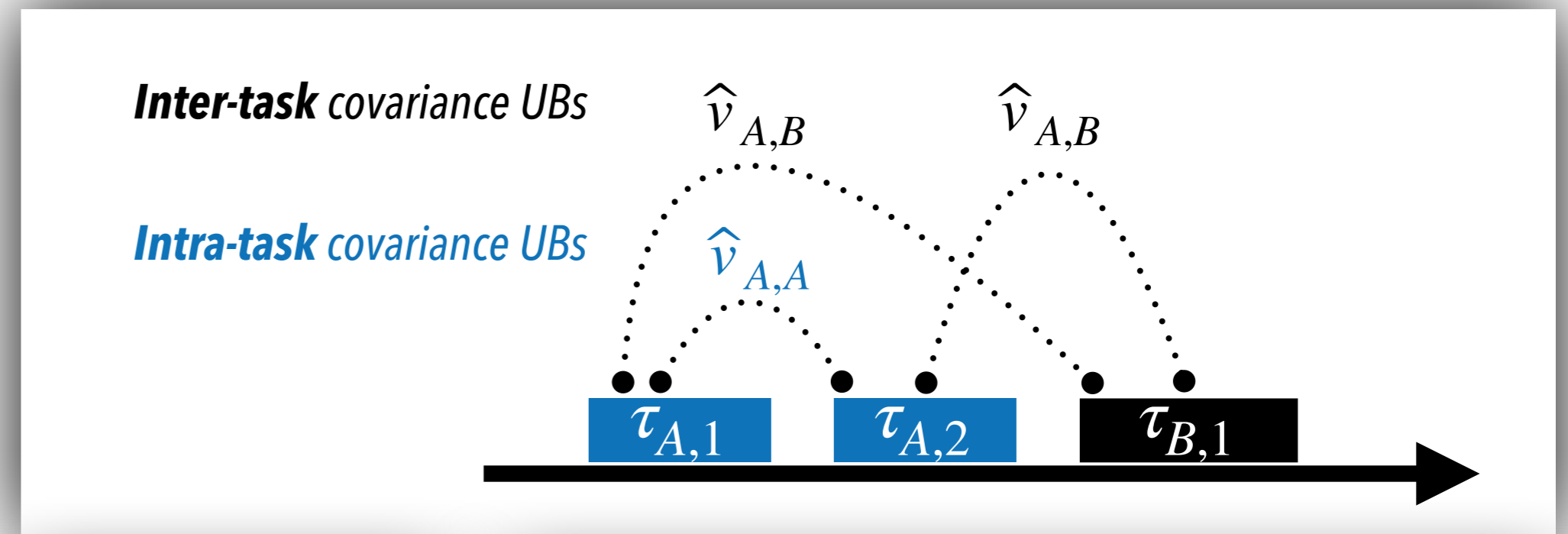
“Analyses are needed that can address dependencies.”

HOW TO STATISTICALLY INFER DEPENDENCE?

“Appropriate statistical studies are needed to investigate the types of dependences and their impact on probabilistic schedulability analysis”

BONUS: COMPUTATION EFFICIENCY!

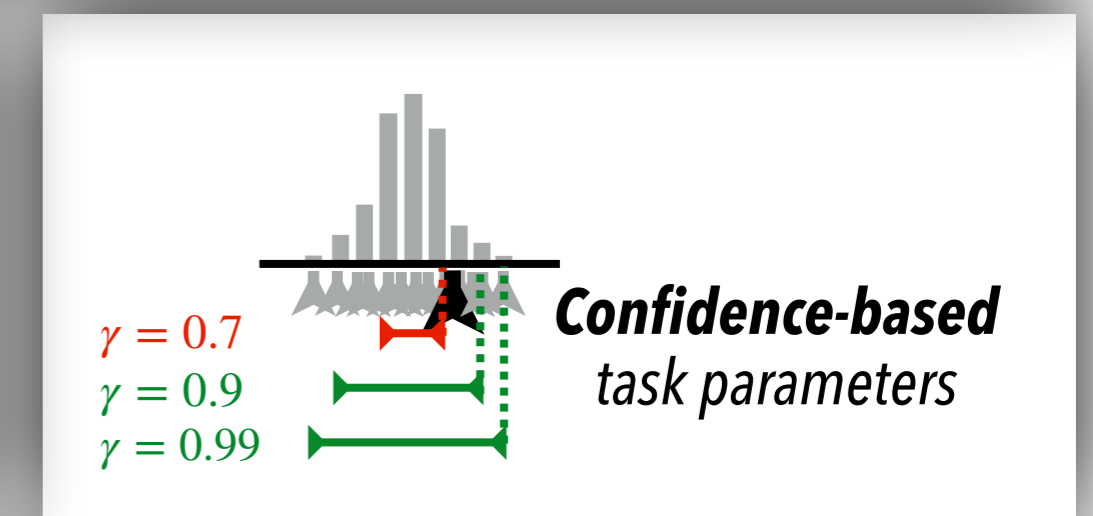
“ensuring that they [analyses] can be applied to problems of a practical size.”



deadline-failure probability

$$\mathbb{P}[R_2 \geq d] \leq \frac{2 \cdot \hat{s}_A^2 + \hat{s}_B^2 + 2 \cdot \hat{v}_{A,A} + 4 \cdot \hat{v}_{A,B}}{\hat{v}[R_B] + (d - \hat{E}[R_B])^2}$$

CAA (dominates CTA)



Closed-form solution

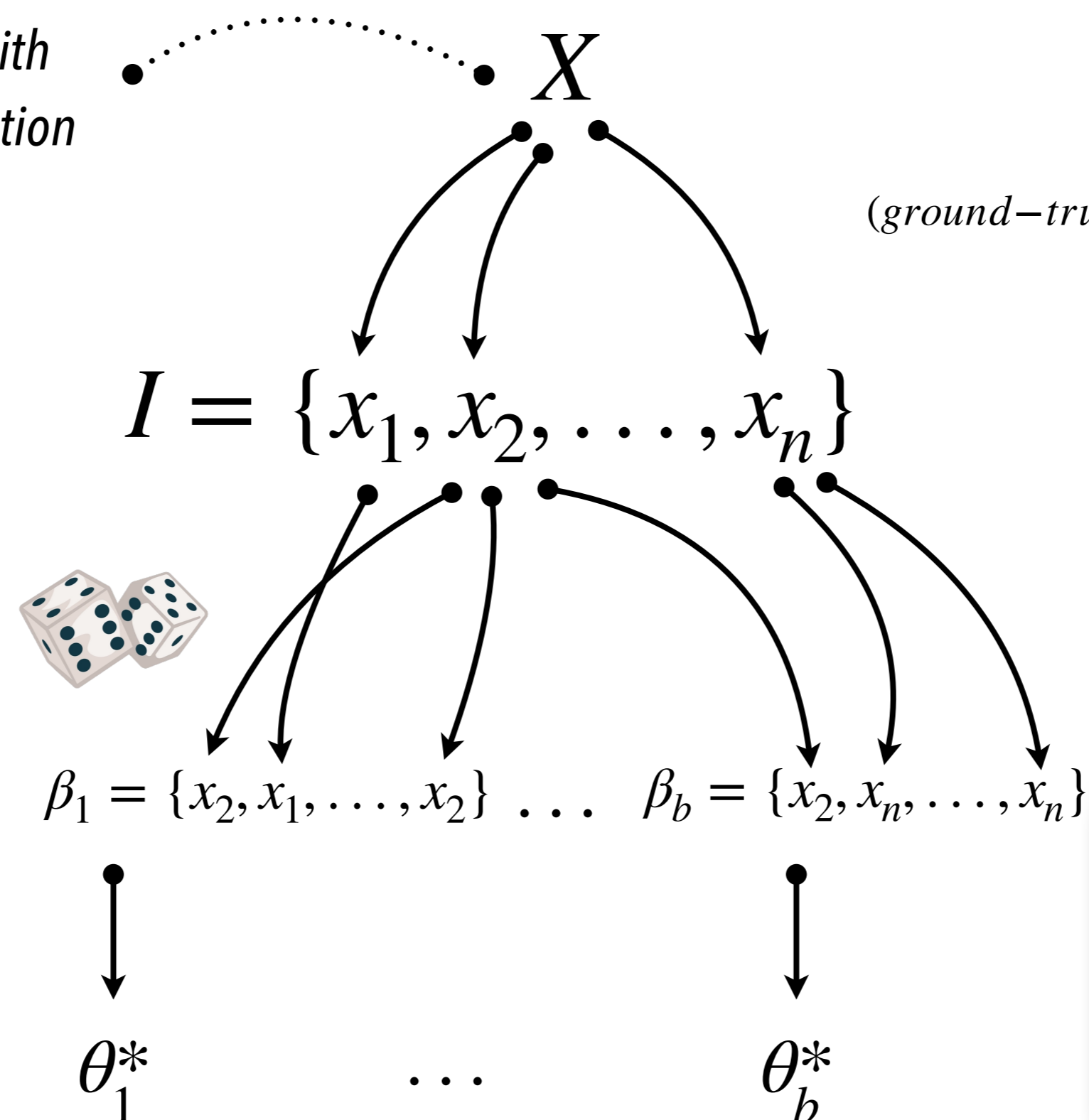
SPEED

[2] “A survey of probabilistic schedulability analysis techniques for real-time systems” Davis and Cucu-Grosjean LITES (2019)

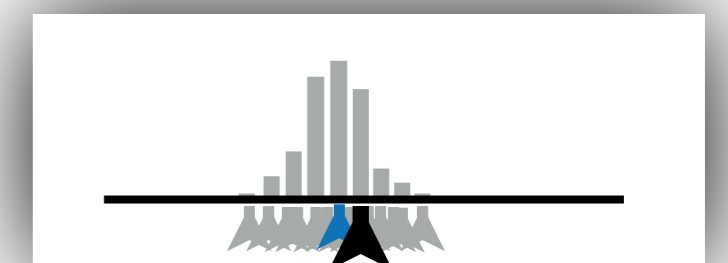
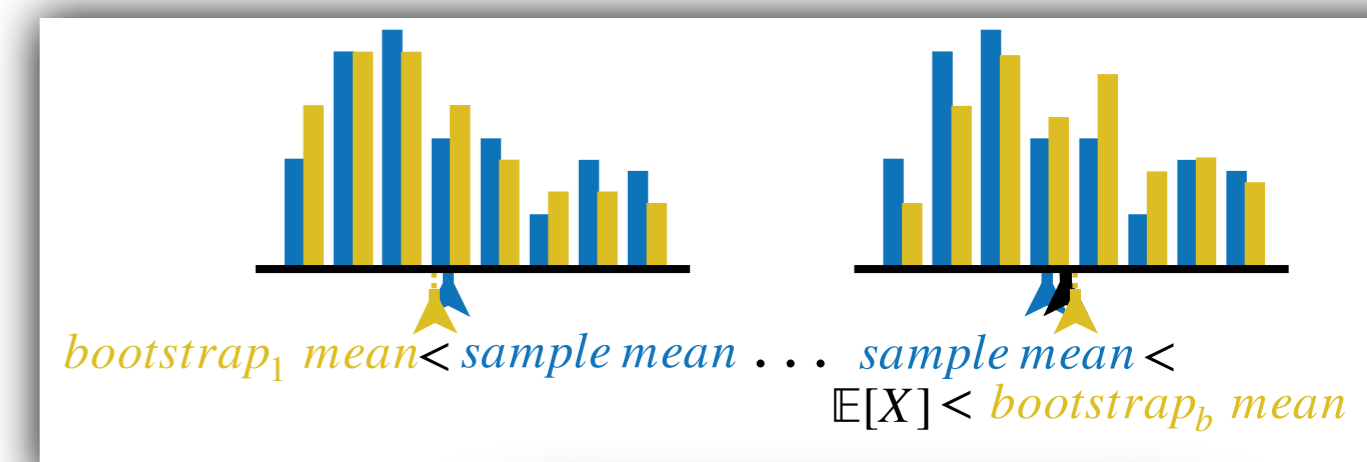
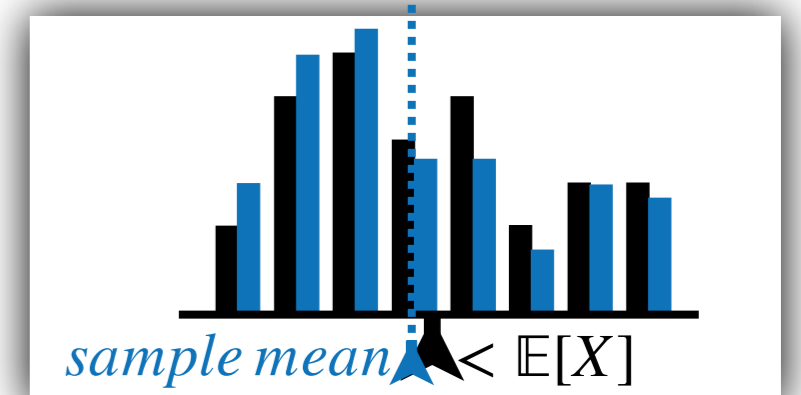
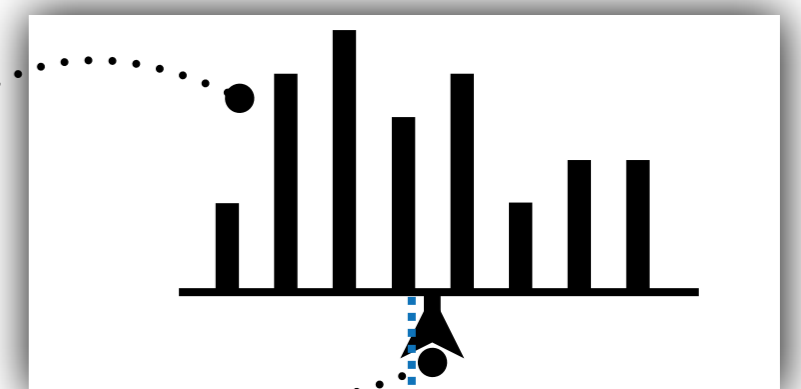
GOAL: DISTRIBUTION-AGNOSTIC STATISTICAL INFERENCE

We use *Nonparametric Bootstrap*

A random variable with an **unknown** distribution



$\mathbb{E}[X]$
(ground-truth expectation)



1. Draw an initial sample I of n independent observations of X

2. Generate b bootstrap samples (by randomly resampling with replacement from the initial sample I)

3. Compute the bootstrap statistic on each bootstrap sample B

Output: Bootstrap distribution of θ