Using Markov's Inequality with Power-Of-k Function for Probabilistic WCET Estimation

Sergi Vilardell
Polytechnic University of Catalonia, Barcelona, Spain
Barcelona Supercomputing Center (BSC), Spain

Isabel Serra
Barcelona Supercomputing Center (BSC), Spain
Centre de Recerca Matemàtica, Barcelona, Spain

Enrico Mezzetti
Barcelona Supercomputing Center (BSC), Spain
Maspatechnologies S.L, Barcelona, Spain

Jaume Abella
Barcelona Supercomputing Center (BSC), Spain

Francisco J. Cazorla
Barcelona Supercomputing Center (BSC), Spain
Maspatechnologies S.L, Barcelona, Spain

Joan del Castillo
Autonomous University of Barcelona, Spain

Abstract

Deriving WCET estimates for software programs with probabilistic means (a.k.a. pWCET estimation) has received significant attention during last years as a way to deal with the increased complexity of the processors used in real-time systems. Many works build on Extreme Value Theory (EVT) that is fed with a sample of the collected data (execution times). In its application, EVT carries two sources of uncertainty: the first one that is intrinsic to the EVT model and relates to determining the subset of the sample that belongs to the (upper) tail, and hence, is actually used by EVT for prediction; and the second one that is induced by the sampling process and hence is inherent to all sample-based methods. In this work, we show that Markov's inequality can be used to obtain provable trustworthy probabilistic bounds to the tail of a distribution without incurring any model-intrinsic uncertainty. Yet, it produces pessimistic estimates that we shave substantially by proposing the use of a *power-of-k* function instead of the default identity function used by Markov's inequality. Lastly, we propose a method to deal with sampling uncertainty for Markov's inequality that consistently improves EVT estimates on synthetic and real data obtained from a railway application.

2012 ACM Subject Classification Computer systems organization → Real-time system architecture

Keywords and phrases Markov's inequality, probabilistic time estimates, probabilistic WCET, Extreme Value Theory

Digital Object Identifier 10.4230/LIPIcs.ECRTS.2022.20

Funding This work has been partially supported by the Spanish Ministry of Economy and Competitiveness (MINECO) under grant PID2019-110854RB-I00 / AEI / 10.13039/501100011033 and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 772773).

34th Euromicro Conference on Real-Time Systems (ECRTS 2022).
Editor: Martina Maggio; Article No. 20; pp. 20:1–20:24
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Introduction

Deriving Worst-Case Execution Time (WCET) estimates of programs is pivotal to show that a real-time system meets its timing requirements [55]. A strand of works tackles this challenge with probabilistic analysis [13] as a way to deal with the increasing complexity of the processors used in real-time systems. These works predict the values of exceedance probabilities of the uppermost tail (a.k.a. high quantiles) of the execution time distributions of programs, which is normally referred to as probabilistic WCET (pWCET) estimate.

Extreme Value Theory (EVT) [16] has been consolidated as a modeling approach for pWCET estimation [1, 7, 27, 46, 52]. Sound applications of EVT can deliver tight and trustworthy pWCET estimates with high confidence. However, the use of EVT for pWCET estimation suffers from several sources of uncertainty that can be categorized into statistical and model-intrinsic (or simply model for short) uncertainty [8]. Statistical uncertainty is, in fact, intrinsic to any sample-based process. It encompasses as the first aspect the testing conditions under which the experiments are performed in reference to those that can arise during system operation. Testing conditions that are representative or worse than operation conditions are the basis to attain representativeness of the sample data (execution time) [2, 4, 25, 38] so that the pWCET estimate holds during system operation. A second aspect of statistical uncertainty relates to the natural uncertainty of a sampling process that, in general, reduces as the sample size increases, and that is handled with confidence intervals. Sampling uncertainty impacts summary statistics (e.g. mean) and tail fitting methods, whose goodness – either of their hypotheses or outcome – is assessed with specific methods [4, 42]. Model uncertainty, instead, relates to uncertainties intrinsic to the mathematical model used for tail prediction. In the case of EVT, model uncertainty relates to determining the threshold from which the upper tail starts. This threshold plays a key role on the trustworthiness (safeness) of EVT results since only samples above it (i.e. the maxima data set) are fed into EVT for pWCET estimation. There is not an exact mathematical method to derive this threshold. Instead, current methods estimate the tail of a distribution [10] based on plot inference [12, 19, 28] and regression analysis [9].

As the first set of contributions, we show that Markov's inequality [36] provides pWCET estimates that are trustworthy by construction at the theoretical (analytical distribution) level and hence, free of any model uncertainty. We, then, illustrate that Markov's inequality is highly pessimistic for decreasing exceedance probabilities, which are specially relevant for pWCET estimation (Section 3). To cope with this limitation, we propose the use of Markov with power-of-k (MIK) functions, $f(X) = X^k$, instead of the identity function, $f(X) = X$, used by the default Markov's formulation. We develop the MEMIK (Minimum Envelope for Markov's Inequality) algorithm that exploits MIK, delivering trustworthy upper-bounds to the underlying distribution, while achieving much tighter estimates than Markov for arbitrarily low exceedance probabilities (Section 4).

As the second set of contributions, at the empirical level, we address the statistical uncertainty of the Markov model by proposing the RESTK (restricted k) method. In order to approximate the expected value of X^k, also defined as the k-th moment of a distribution, RESTK derives the *sample moment* of a distribution from a collected sample (Section 5).

We evaluate RESTK on a set of representative distributions and tail functions and show that it consistently outperforms EVT: EVT produces 21.8% over-estimation on average (up to 37%) and RESTK keeps average over-estimation at 9.4% (up to 20%). We also evaluate RESTK on sampled data from a real railway use case: the central safety processing unit of the European Train Control System (ETCS) reference architecture. On average, RESTK over-estimates less than 9% with respect to real data (distribution) and it outperforms EVT.

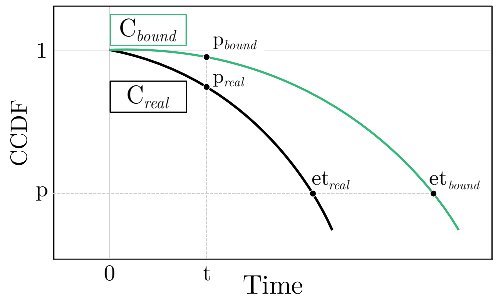


Figure 1 Generic representation of the tightness of pWCET estimations. CCDF stands for Complementary Cumulative Distribution Function.

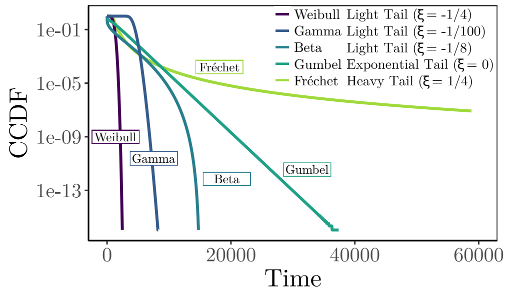


Figure 2 CCDF for GEV distributions with $\xi = -1/4$, $\xi = 0$, and $\xi = 1/4$; $\xi = -1/8$ for Beta; and $\xi = -1/100$ for Gamma.

The rest of this paper is structured as follows. Section 2 presents some relevant background, and introduces and exemplifies EVT model uncertainty stemming from tail selection. Section 3 introduces Markov's inequality and discusses its use for pWCET estimation. Section 4 introduces our proposal of using $f(X) = X^k$ to tighten pWCET estimates. Section 5 introduces the RESTK method as a way to deal with statistical uncertainty for Markov's inequality. Section 6 compares the pWCET projections obtained with EVT and Markov models. Section 7 shows analogous results for the railway use case. Section 8 surveys the main related works, and Section 9 presents the main conclusions of this work.

2 Background and Problem Statement

2.1 Probabilistic WCET estimation

Let us assume a random variable $X > 0$, corresponding to the execution time of a real-time program and whose maximum value is finite. We aim at bounding the probabilities of X in its uppermost tail (i.e. high execution times) by providing safe and tight timing bounds.

An execution time probability distribution C_{bound}, see Figure 1, is said to upper-bound another probability distribution C_{real} – so being a pWCET bound – when for any exceedance probability p the execution time of the former, $et_{bound}(p)$, is higher (or equal) than that of the latter, $et_{real}(p)$. It also holds that, for any given execution time et, the exceedance probability of the former $p_{bound}(et)$ is higher (or equal) than that of the latter $p_{real}(et)$. This can be expressed as $tightness(p) = \frac{et_{bound}(p)}{et_{real}(p)}$.

2.2 Representative Distributions

In the scope of timing analysis of real-time programs, it is reasonable to assume that the target program will always terminate. Hence, a WCET value upper-bounding all possible program's executions always exists. It therefore follows that, the tail of its execution time distribution is necessarily a light distribution, which has been shown to be safely and tightly upper-bounded with light and exponential tail distributions [1, 17, 46]. For this reason, we focus on light and exponential tails.

While heavy tails are, therefore, unnecessarily pessimistic and hence, left out of our discussion, an execution time sample could apparently correspond to a heavy tail distribution. This could be, for instance, the case when the sample size is not large enough to provide

Table 1 Reference distributions used to drive discussions. The parameters are μ (mean), σ (standard deviation), α (shape) β (shape), λ (scale), and σ (scale).

Acronym	Parameters
Gaussian	$\mu = 100$, $\sigma = 10$
Weibull	$\alpha = 4$, $\lambda = 80$
Beta	$\alpha = 1/4$, $\beta = 8$
Mixture	$\mu = \{5, 50, 100\}$, $\sigma = 10$ $w = \{0.6, 0.39, 0.1\}$

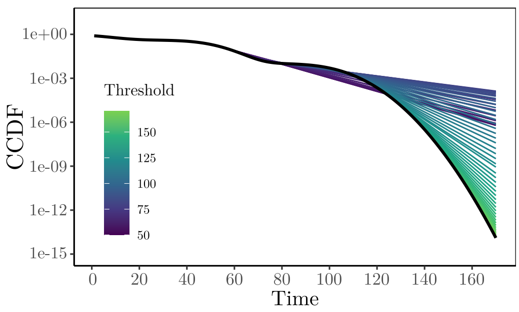


Figure 3 EVT estimation on the mixture distribution (three Gaussian with parameters $\mu = \{5, 50, 100\}$, $\sigma = \{10, 10, 10\}$ and weights $w = \{0.6, 0.39, 0.01\}$, respectively.

sufficient representativeness in a mixture distribution. In the general case, this concern relates to the sampling process (sample size in particular), shared across all applications of statistics, and therefore, beyond the scope of our methodology. Nevertheless, nothing precludes the use of our methodology for heavy tails.

The different types of tails are illustrated with the Complementary Cumulative Distribution Function (CCDF) of several example distributions in Figure 2. All three Generalized Extreme Value (GEV) distributions have location $\mu = 0$ and scale $\sigma = 1000$: the Weibull distribution (light tail) with shape $\xi = -1/4$ has a sharp slope and a maximum value (2500 in the example); Gumbel (exponential tail) with $\xi = 0$ has also a relatively sharp slope but it has no maximum; and the exceedance probability for the Fréchet distribution (heavy tail) with $\xi = 1/4$ decreases polynomially. The Beta (light tail) distribution has a similar profile to the Weibull distribution and it is commonly used to model high quantiles of random variables with a finite defined interval [3, 30, 37], which fits WCET modeling. And the Gamma (light tail) distribution is also typically used in EVT and WCET [4].

All former distributions are unimodal, which is a common way to represent execution time profiles. However, it has also been shown that the execution time of many programs presents "clusters". That is, the program's execution time varies around two or more central values rather than vary around a single mean. This results in *mixture distributions* that can arise both in sequential applications and parallel applications [1, 56, 57]. As an example of the former, let us assume a program whose execution time profile is influenced by the latency of its load/store operations which can hit or miss the data L1 cache and L2 cache across different runs depending on the program's inputs. This results in a mixture distribution with 3 clusters (peaks) around the data L1, L2 and memory latencies, respectively. An example of mixture distribution is represented by the black line in Figure 3.

Overall, we use a solid set of reference distributions that are in line with the state of the art [4, 11, 27, 42]. In particular, we use unimodal (Gaussian, Weibull, Beta, and Gamma) and challenging multi-modal distributions (Gaussians and Weibulls) with different tail profiles to increase representativeness. Besides, we use a real railway use case in Section 7. In order to drive the explanations, we use a specific set of reference parameters for several distributions (see Table 1), while in the evaluation section we analyze a wider set of parameters (see Table 3). We have not included the Gamma in the reference distributions as it yields very similar results to the Weibull. However, results are provided for two different Gamma distributions in the (evaluation) Section 6 as part of the complete result set.

2.3 EVT usage for pWCET estimation

The main theorems of EVT can be deemed as two different ways of thinking about the extremes [22]. The first, Block Maxima (BM), splits a sample of a distribution into fixed-size *blocks* and selects the maximum value in each block. The second one, Peaks over Threshold (PoT) defines a *threshold* such that the values above it are deemed as "rare enough", i.e. they are considered the tail of the distribution. EVT's main result is characterized on both theorems when the set of maxima and the threshold tend to infinity.

Both, PoT and BM, share the fundamental goal of identifying where the tail of the distribution starts to fit the proper distribution to the tail. However, for PoT such cut point is defined explicitly as a threshold, therefore allowing its use in the domain of distributions from a mathematical perspective. Hence, in this work we focus on PoT for our analysis, although mathematical uncertainties identified are generally shared by PoT and BM.

Peaks Over Threshold (PoT). Given a random variable X with a cumulative distribution function (CDF), F, and a threshold, $u > 0$, such that $y = x - u$, the excess random variable X_u defined as $(X - u \mid X > u)$ is given by the CDF F_u defined as

$$F_u(y) = P(X - u \leq y | X > u) = \frac{F(u+y) - F(u)}{1 - F(u)}, \quad y \geq 0 \tag{1}$$

A PoT model is a semi-parametric model where the law of X for $x < u$ is described by the empirical distribution, and for $x > u$ is defined by

$$P(X > x) = S(x) = S(u)S_u(x - u) \tag{2}$$

where $S(x) = 1 - F(x)$ and $S_u(x_u) = 1 - F_u(x_u)$ are the CCDFs, with $x_u = x - u$.

▶ **Theorem 1** (Pickands–Balkema–de Haan theorem [5])**.** *Let F be a distribution function such that $F(x) < 1$ with F_u being its conditional excess distribution function. Then, F_u converges in probability to the generalized Pareto distribution (GPD) for large u. That is, $F_u \xrightarrow{\mathcal{L}} G(y; \xi, \sigma)$ as $u \to \infty$, where*

$$G(y; \xi, \sigma) = \begin{cases} 1 - \left(1 + \frac{\xi y}{\sigma}\right)^{-\frac{1}{\xi}} & \text{if } \xi \neq 0 \\ 1 - \exp\left(-\frac{y}{\sigma}\right) & \text{if } \xi = 0 \end{cases} \tag{3}$$

With $\sigma > 0$, and $y \geq 0$ when $\xi \geq 0$ and $0 \leq y \leq -\sigma/\xi$ when $\xi < 0$. This result is crucial for tail estimation. If the conditions for the theorem are met, the GPD family of functions results in accurate estimates for the most extreme values on the tail. However, the conditions are met when the threshold $u \to \infty$, which brings uncertainty in the implementation of estimates with the GPD since finite values of u need to be used.

Overall, there is an unavoidable model uncertainty when selecting the threshold u by definition due to having to select a finite value. This is beyond specific aspects related to statistical uncertainty from the sampling process or the threshold u estimation process. Tail selection has been addressed by techniques like the Hill Plot based on Hill estimator [28], the Mean Excess Plot [19], or the CV Plot [12]. The selection of the threshold u changes the fit of the GPD model and requires careful analysis and selection to prevent GPD from over-approximating or under-approximating the analyzed distribution.

As an illustrative (visual) example, Figure 3 shows a mixture of three Gaussians with $\mu = \{5, 50, 100\}$, $\sigma = \{10, 10, 10\}$ and weights $w = \{0.6, 0.39, 0.01\}$. Each mode around Time $\{5, 50, 100\}$ represents a Gaussian in the mixture. For each different tail threshold u such that

$u > 60$, we fit an exponential tail, given that Gaussian distributions have exponential tails. In Figure 3 we see three scenarios that we exemplify with approximate ranges of u. For values of u around $[60, 80]$ (purple lines), we see how the exceedance probability of the reference distribution is underestimated in the Time range $[80, 120]$ (probabilities $10^{-3} - 10^{-5}$). For u in the range $[120, 140]$ (blue-green lines), GPD over estimates the reference distribution. For values of u above 140 (green lines), the estimate becomes increasingly tight.

3 Chebyshev and Markov Inequalities for pWCET Estimation

This section analyzes the applicability of Markov's inequality as an alternative model to EVT for the problem of trustworthy pWCET estimation and shows that it is not subject to any model uncertainty, thus resulting in provably safe bounds for the analyzed distribution. Yet, as we also show in this section, Markov's inequality produces very pessimistic estimates. For completeness, we start by introducing Chebyshev's inequality, as it is the generalization of Markov's inequality.

3.1 Chebyshev's Inequality

▶ **Definition 2.** *Let X be a discrete random variable with probability function $f_X(x)$, where x are the particular values that X can take. The expected value of X is:*

$$E(X) = \sum_{x \in X} x f_X(x) = \sum_{x \in X} x P(X = x) \tag{4}$$

▶ **Theorem 3** (Chebyshev's Inequality [49]). *Let X be a non-negative random variable, $b > 0$, and f a non-negative and increasing function. Chebyshev's inequality states that:*

$$P(X \geq b) \leq \frac{E(f(X))}{f(b)} \tag{5}$$

For WCET estimation, X corresponds to the execution time distribution to be bounded, and b to an execution time for which we want to find its upper-bound probability.

Regarding f, it needs to be defined to realize the general Chebyshev's inequality into a specific upper-bound function. The function f can be any non-negative function so that for a particular domain D, Property 6 holds.

$$\forall x \in D, \quad f(x) > 0 \tag{6}$$

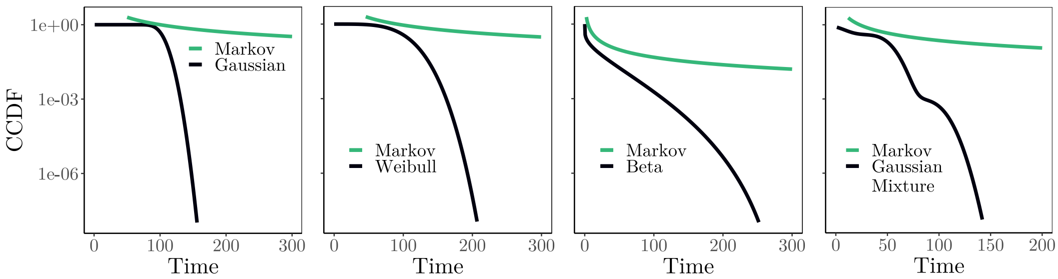
f must also be an increasing function so that for a given interval I, Property 7 holds.

$$a, b \in I \quad | \quad a < b, \quad \Rightarrow f(a) \leq f(b) \tag{7}$$

Interestingly, Chebyshev's inequality does not require determining where the actual tail distribution starts but instead works with the entire distribution, hence removing EVT's model uncertainty for tail selection. In fact, Chebyshev's inequality carries no model uncertainty.

▶ **Observation 4.** *Chebyshev's inequality is a model uncertainty free general model for pWCET estimation.*

Chebyshev's inequality also applies to continuous and discrete distributions regardless of their characteristics (e.g. shape, variance, kurtosis, etc.). Also, it is non-parametric, i.e. it makes no assumption on parameters for the studied distribution.



(a) Gaussian distribution. **(b)** Weibull distribution. **(c)** Beta distribution. **(d)** Gaussian Mixture.

Figure 4 Markov's Inequality bound for the reference distributions.

3.2 Markov's Inequality

Markov's inequality is a specific instantiation of Chebyshev's inequality.

▶ **Corollary 5** (Markov's Inequality [36]). *Let $X > 0$ and let the function f be the identity function $f(X) = X$. Hence, Markov's inequality yields:*

$$P(X \geq b) \leq \frac{E(X)}{b} \tag{8}$$

As for the baseline Chebyshev's inequality, Markov's inequality holds for any real-valued random variable with a finite expected value and positive value b. Also, it (i) is a trustworthy upper-bound, by construction, of the underlying distribution; (ii) has no model uncertainty; (iii) is non-parametric; and (iv) can be applied to discrete and continuous distributions.

3.3 Markov's inequality on low probabilities

Besides trustworthiness, pWCET estimates are also required to be reasonably tight, specially for the range of relevant probabilities usually considered for pWCET estimation, e.g. $[10^{-6}, 10^{-15}]$. In this line, our analysis shows that Markov's inequality tends to be hardly useful for pWCET estimation.

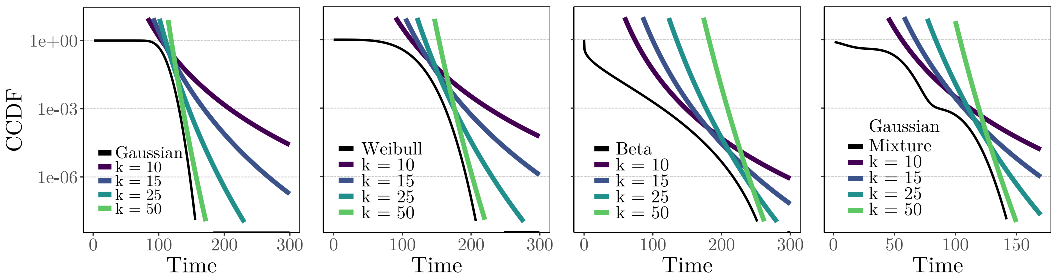
This is better illustrated in Figure 4 which shows for all considered distributions the probability bounds given by Markov's inequality. We can observe that estimates are very pessimistic, orders of magnitude higher than the real probability. This includes the range of probabilities of interest for pWCET estimation. In fact, we see that Markov's inequality never goes below 10^{-2} for all distributions for the execution time value range plotted.

▶ **Observation 6.** *Markov's inequality in its original form is too pessimistic to be usable in practice for pWCET estimation.*

4 Power-of-K functions for Markov's Inequality

One of the main insights of this work is that the key reason for Markov's inequality resulting in loose pWCET bounds lies on the fact that it builds on the identity function, $f(X) = X$, of a random variable. In this section, we show how a different function can lead to increased tightness on the produced pWCET estimates while preserving trustworthiness.

In particular, we contend that the power function for any $k \in \mathbb{R}_{>0} = \{k \in \mathbb{R} | k > 0\}$, i.e. $f(X) = X^k$, or power-of-k function for short, can be safely used instead of the identity function to obtain tighter and trustworthy pWCET bounds.



(a) Gaussian distribution. **(b)** Weibull distribution. **(c)** Beta distribution. **(d)** Gaussian Mixture.

Figure 5 MIK bounds for the reference distributions.

▶ **Definition 7.** *Let X be a discrete random variable and k a positive real value. The expected value of X^k, also defined as the* k-th *(theoretical) moment, is:*

$$E(X^k) = \sum_x x^k P(X = x) \tag{9}$$

▶ **Corollary 8** (Markov's Inequality to the power-of-k)**.** *Let $X > 0$ and let the function f be the power-of-k function $f(X) = X^k$. Markov's inequality to the power-of-k yields:*

$$P(X \geq b) \leq \frac{E(X^k)}{b^k} \tag{10}$$

Hence, the probability that X takes a value greater or equal to b is bounded by $E(X^k)/b^k$. This makes Markov's inequality with $f(x) = x^k$ (**MIK** for short) a safe pWCET estimate when X represents the execution time of a program.

Proof. Theorem 3 holds true when Property 6 and Property 7 are fulfilled. The power-of-k function *does not fulfill those properties in general.* However, when the x domain is restricted to the positive real numbers $\mathbb{R}_{>0} = \{x \in \mathbb{R} | x > 0\}$, which in fact includes the domain of execution time profiles, the power-of-k function does fulfill Properties 6 and 7 since x is positive, so x^k is also positive and an increasing function. ◀

Overall, for this application scenario ($x \in \mathbb{R}_{>0}$), Equation 10 is an upper-bound when using the power-of-k function onto the reference distribution for any value of $k \in \mathbb{R}_{>0}$. Hence, it can be leveraged for pWCET estimation.

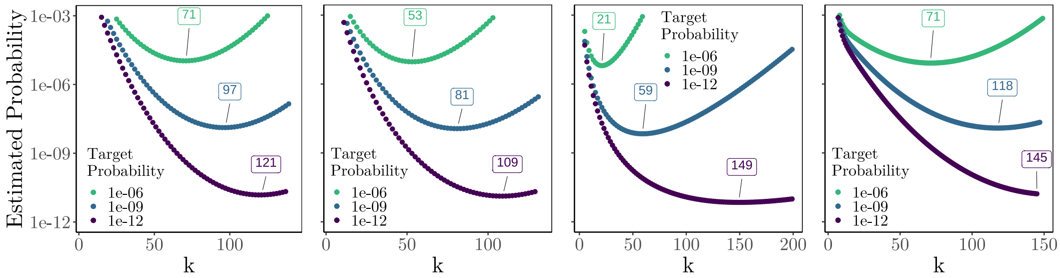
It is worth noting that other functions can exist that fulfill Properties 6 and 7. While exploring them is part of our future work, as shown in Section 4.1, MIK (i.e. $f(X) = X^k$) achieves very tight pWCET estimates which leaves small room for improvement.

▶ **Observation 9.** *For every value of $k \in \mathbb{R}_{>0}$, MIK (i.e. Markov's inequality with $f(X) = X^k$) is a safe pWCET estimate when X represents the execution time of a program.*

Note that there is no theoretical constraint on the maximum value of k, which can be any positive real number k.

4.1 Tightness of MIK for increasing values of *k*

Once we have established the safe use of MIK for pWCET estimation, we illustrate the impact of varying k on tightness. Figure 5 shows for several values of k, $\{10, 15, 25, 50\}$, and the reference distributions presented before, that MIK dramatically increases the tightness



(a) Gaussian distribution. **(b)** Weibull distribution. **(c)** Beta distribution. **(d)** Gaussian Mixture.

Figure 6 Evolution of MIK bounds with the value of k.

provided by Markov's inequality (Figure 4), while remaining a trustworthy upper-bound for every value of k. We also observe that MIK tightness remains for high exceedance probabilities, which hence makes it a promising model to provide pWCET estimates.

▶ **Observation 10.** *Markov's inequality with $f(X) = X^k$ heavily reduces the pessimism of Markov's inequality.*

Intuitively, from Figure 5, higher values of k result in tighter estimates, i.e. minimizing the distance between the reference distribution and the upper-bound distribution. However, this is not always the case. For instance, if we take a closer look at the Beta distribution (Figure 5(c)), we see that at cut-off probabilities 10^{-3} and 10^{-6} the tightest MIK estimates are not obtained for the highest value of k evaluated (50).

▶ **Observation 11.** *For a given threshold probability, higher values of k do not necessarily result in a tighter MIK bound.*

This is better illustrated with the examples in Figure 6 that shows quantitatively the evolution of the MIK bound obtained for varying values of k.

In this experiment, for the value of the target distribution at each probability, we evaluate MIK for different values of k. As it can be seen, for every threshold probability and distribution the value of k resulting in the tightest estimation is different. For instance, for the Gaussian distribution (Figure 6(a)) and target probability 10^{-6}, $k = 71$ produces the tightest estimate, while for 10^{-9} and 10^{-12} the best k is 97 and 121, respectively. As a general trend, we see that the lower the target exceedance probability, the higher the value of the best k is. Yet, the highest value of k evaluated for each target probability does not produce the tightest bound. Overall, for each probability there exists a value of k producing the tightest upper-bound, with the optimal value of k depending on the actual reference distribution.

▶ **Observation 12.** *Increasing the tightness of MIK for each probability is an optimization problem on k which only increases accuracy and does not affect trustworthiness.*

In order to address this optimization problem, we propose **MEMIK** (Minimum Envelope for MIK, i.e. Markov's Inequality to the power-of-K). MEMIK combines the results of MIK bounds obtained for any value of k by keeping, for each point in an interval, the value of k producing the tightest estimate. This set of points form an envelope that is a provable trustworthy and tight tail bound by construction for any exceedance probability. Therefore, MEMIK improves the pessimistic upper-bounds of Markov's inequality (see Figure 4), with a much tighter envelope that is usable for pWCET estimation. Formally, the MEMIK bound is defined as follows:

$$P(X \geq b) \leq \min_k \frac{E(X^k)}{b^k} \quad \text{for} \quad k > 0 \tag{11}$$

Algorithm 1 Compute an envelope using Power-of-k.

1: **function** MEMIK(t_{range}, t_{step}, p_{all}, $max_k(p_{all})$, k_{step})
2:     **for** $t \in t_{range}, t_{step}$ **do**
3:         **for** $p \in p_{all}$ **do**
4:             $mik_{best}(p) \leftarrow \infty$
5:             **for** $k \in ([1, max_k(p)], k_{step})$ **do**
6:                 $vpred \leftarrow EVAL(t, k, p)$
7:                 **if** $vpred < mik_{best}(p)$ **then**
8:                     $mik_{best}(p) = vpred$
9:                     $k_{best}(p) = k$
10:                 **end if**
11:             **end for**
12:             $envelope(t, p) \leftarrow < mik_{best}(p), k_{best}(p) >$
13:         **end for**
14:     **end for**
15:     **return** *envelope*
16: **end function**

MEMIK, see Algorithm 1, which uses point-wise power-of-k Markov's inequality values, performs a simple complete exploration of MIK values over a given time range t_{range} and over a configurable range of k, determined by the maximum value max_k to be explored for each probability p in the set of probabilities of interest p_{all}. The granularity of MEMIK exploration over t and k is determined by the t_{step} and k_{step} parameters respectively. For each probability p in the interval of interest p_{all} (line 3) and k in the range determined by $max_k(p)$ (line 5), the algorithm estimates the value of the target distribution, *EVAL(t, k, p)* (line 6). To that end, we evaluate Equation 10 with $E(X^k)$, which corresponds to the theoretical k-th moment of the target distribution from Equation 9, obtaining *vpred* that we compare to the best MIK value so far for all considered k (line 7).

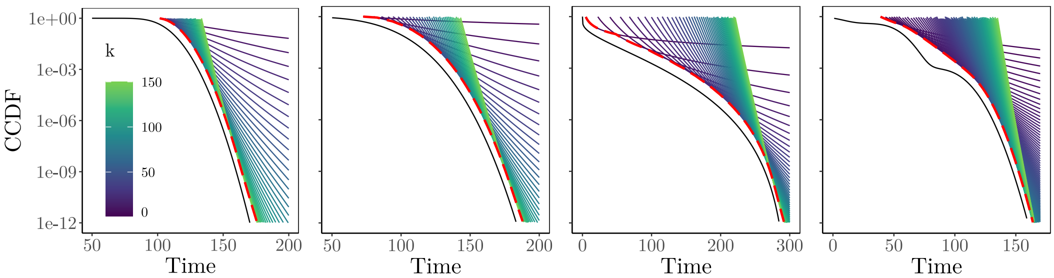
The minimum MIK value produced for a given t and across all k values ($mik_{best}(p)$) is stored, together with the corresponding $k_{best}(p)$, in the data structure *envelope* (line 12). Eventually, after iterating over the whole time interval, the algorithm returns the *envelope* data structure (line 15) which holds the point-wise definition of the approximation envelope. Note that, if for any value t, the value of $k_{best}(p)$ matches $max_k(p)$, then $max_k(p)$ can be increased to find tighter bounds.

We applied MEMIK to our reference distributions, for which we can derive the theoretical moments. For this experiment, we varied the value of k up to 150 with $k_{step} = 1$. We obtained the envelopes depicted in Figure 7 which provides evidence that MEMIK produces tight and trustworthy estimates for all distributions, with an observed error of around 5%.

Overall, this section provides the key result that our proposal, Markov's inequality to the power-of-k (MIK), unlike EVT, suffers no model uncertainty at the theoretical level and hence, provides correct-by-construction pWCET estimates that are much tighter than those provided by the default Markov's inequality. This leaves sampling uncertainty as the problem to address.

5 Handling Markov Sampling Uncertainty

So far we have been reasoning on examples for which we could compute the theoretical k-th moments for each distribution. This was possible since the distributions considered were known and, hence, we could compute exactly the value of each moment (i.e. $E(X^k)$ for each



(a) Gaussian distribution. **(b)** Weibull distribution. **(c)** Beta distribution. **(d)** Gaussian Mixture.

Figure 7 MEMIK bound (envelope) on the reference distributions.

value of k) using its analytical closed form. However, we need to consider the scenario in which only samples are available. Hence, as for any other sample-based method, we need to deal with the underlying sampling uncertainty.

A commonality of sample-related methods like EVT [13], and something that we also assume, is that, input samples are independent and identically distributed [16] (i.i.d.) or at least exhibit extremal independence [44]. The i.i.d. property can be pursued with platform randomization [31] or data (time measurements) sample randomization [33].

For the case of the Markov's inequality, this translates into deriving the **sample moments**, referred to as $\hat{E}(X^k)$ (for each value of k). In particular, we need an estimator for high-order moments that can produce good estimates for any distribution.

5.1 Sample moment estimation

The k-th moment of a random variable X can be estimated as (N is the sample size [23]):

$$\hat{E}(X^k) = \frac{1}{N}\sum_{i}^{N} X_i^k \tag{12}$$

In general, this estimator is the best one to deal with high-order sample moments [26], as it is asymptotically unbiased. Given that it asymptotically tends to a Gaussian distribution [6], the properties of the Central Limit Theorem's apply to it. However, the estimator is asymptotically unbiased [23] only when using large amounts of data. For instance, for a sample of a Gaussian distribution with $\mu = 100, \sigma = 10$ and $N = 10^3$, the difference between the 3^{rd} exact moment, and the sample moment using Equation 12 is about 0.02%, it is between 1% and 3% for the 50^{th} moment, and can be up to 160% for the 100^{th} moment.

When the sample moment, i.e. the estimate of the k-th moment, is higher than the theoretical moment, there is a risk of underestimating the upper tail of the distribution by assigning to a certain probability a smaller quantile than it has in reality. The approach we propose to limit that risk consists in setting a maximum value of k allowed for each different probability, which we refer to as max_k. In order to illustrate the effect of not controlling max_k, Figure 8 assesses the tightness of MEMIK over a particular set of nine probabilities (from 10^{-7} to 10^{-15}). The red triangles represent the theoretical bound of MEMIK using Equation 11 with $E(X^k)$ being the theoretical moments, while the orange dots (NO RESTK) represent the application over multiple simulations of Equation 11 using the sample moment from Equation 12. On both applications we set up a high value for max_k, 150. We can see how the loss of consistency of the sample moment estimator on Equation 12 results in bad estimates.

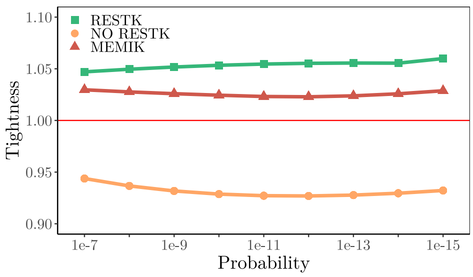


Figure 8 MEMIK with sample moments ($n = 1000$ and $n_{sims} = 100$) on the reference Gaussian distribution with (RESTK) and without (NO RESTK) restricting k. Also MEMIK evaluated with theoretical moments.

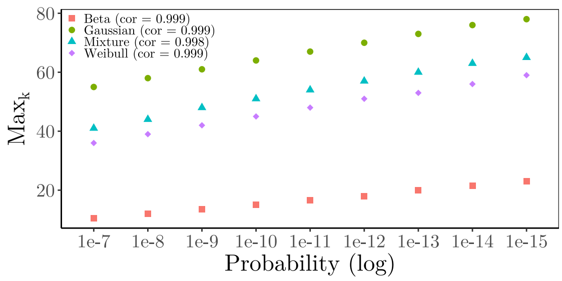


Figure 9 Minimum max_k values for the reference distributions used in this work.

An intuitive way to control the gap between the k-th sample and the theoretical moments is to vastly increase the sample size, which in our domain would require an unaffordable number of runs of the task under analysis. Alternatively, one can control the range of values of k explored in Equation 12. By doing so, we trade some tightness for trustworthiness. That is, if we explore values of k until a low max_k limit, we can see in Figure 6 that the theoretical bound is not optimal in terms of accuracy. On the other hand, small values of max_k also limit the inaccuracy of the sample moment estimator. Note that it is not possible to identify a general optimal value of max_k for any kind of data under analysis. The appropriate max_k value changes across distribution types, across the same distribution type with different parameters, and even across probabilities for a given distribution. For this reason, we propose the *restricted k method* (**RESTK**) that builds on the information gathered directly from the samples to derive max_k so as to produce trustworthy and tight results.

5.2 Understanding the behavior of max_k

We gain insight on the behavior of max_k along 3 axes. We analyze i) whether for a given distribution there exists a pattern for max_k that can provide tight and safe results using the sample moment estimator; ii) whether this pattern can be predicted using only the information from the sample; and iii) whether the pattern can be generalized for any distribution.

We focus on the same example distributions used in previous sections. We fix the interval $[1, 150]$ as exploration range for k. In order to account for sample uncertainty, we perform $n_{sims} = 10^3$ Monte-Carlo simulations, each one considering a random sample of size $n = 10^3$.

We first compute Markov to the power-of-k (MIK) using Equation 10 with the sample moment estimator in Equation 12, for all selected k. In each simulation, we increase values of k and find the **first** (smallest) value of k that produces underestimation. This is computed by comparing the estimation with the actual value of the distribution. That is, we take the value of k for which the estimated quantile is smaller than the theoretical quantile. Then, we set $k - 1$ as our max_k. For each Monte-Carlo simulation, we compute the max_k for all target probabilities. When all simulations are performed, we keep the smallest max_k for each probability. As a result, for each experiment of n_{sims} Monte-Carlo simulations, we obtain one value of max_k for each probability under study. We plot those values in Figure 9 from which we derive two main conclusions.

We observe that the values in Figure 9 follow a linear distribution. For each distribution we fit minima max_k values to a linear model and we derive the resulting correlation coefficient. The correlation coefficient quantifies the strength of the linear relation between two variables. It ranges between -1 and 1, with 1 or -1 indicating perfect correlation (all points would lie along a straight line).

The distributions we use in this work include 4 types of unimodal distributions, 2 multi-modal distributions and several parameters thereof (see Table 3 in Section 6). For all those distributions, Table 2 shows that the correlation coefficient is very high and steadily stays above 0.99. Even the empirical distributions derived from the case study analyzed in Section 7 result in a high coefficient of correlation (0.98 on average), despite they tend to produce more variability in the estimations. Hence, empirical evidence in support of linearity for the minimum observed value of max_k is strong for the distribution tails and range of probabilities representative for the WCET. Besides, the application of RESTK includes a method to assess whether the linearity property holds, building on the observed data. It is also noted that, similar empirical reasoning is used to support statistical arguments whether phenomena adheres to specific distributions builds on empirical tests. For instance, in the case of EVT, previous work uses QQ-plots to assess, based on observation, whether some tails can be considered exponential [34].

Table 2 Correlation Coefficient for all the distributions used in this work.

Gaus1	**Gaus2**	**Weib1**	**Weib2**	**Beta1**	**Beta2**	**Gam1**	**Gam2**	**Mix1**	**Mix2**	**Mix3**	**Mix4**
.999	.999	.999	.999	.999	.998	.999	.999	.998	.998	.997	.998

Overall, by restricting max_k, one can avoid under-estimating the upper tail of the modeled distribution. This is exemplified in Figure 8, where the green squares (RESTK) represent the estimates obtained for the Gaussian distribution when applying the max_k values in Figure 9. By restricting max_k, we address the lack of trustworthiness in Figure 8 (NO RESTK) and produce tight and trustworthy bounds. Analogous results are obtained for the other distributions.

5.3 Deriving max_k from unknown distributions

When deriving Figure 9, we built on the values of the theoretical quantiles so as to determine the value of k for which the sample moment starts underestimating. Given a sample of size 10^p, we can estimate confidently quantiles from exceedance probabilities bigger than $1/10^{(p-1)}$. In this case, based on the law of large numbers, it is very likely to see around 10 realizations whose probability is of the order of $10^{(p-1)}$ [47]. That is, on a sample of size $N = 1000$ we will see around 10 realizations whose probability is 0.01 (1%). Therefore, for a sample of size 10^4, quantiles corresponding to exceedance probabilities 10^{-3} and 10^{-2} can be estimated easily with the usual quantile estimation functions from statistical packages [29]. The computation of confidence intervals for quantile estimation can be done using distribution-free methods like Kaigh and Lagenbruch or bootstrap [48]; and in any case the accuracy of the estimation can be increased using a bootstrap technique to correct variability.

RESTK estimates at least three max_k points to construct its model and assess linearity. The latter is assessed by deriving the correlation coefficient for these three points. If such coefficient is above a threshold $th = 0.95$, we deem max_k boundary to be linear and vice-versa (in which case RESTK cannot be applied). For instance, the quantiles corresponding to exceedance probabilities 10^{-3}, 10^{-4}, and 10^{-5} can be estimated very accurately with a sample of size $n = 10^6$. These reference points allow us to assess when RESTK underestimates,

Algorithm 2 Computing the boundary necessary to apply RESTK approach.

```
1: function RESTKBoundary(sample, k_range, k_step, n_boot, n_sims, p_test, p_all, th)
2:     for p ∈ p_test do
3:         q_est ← estimateQuantiles(sample, p)
4:         for sim ∈ n_sims do
5:             sample_boot ← bootstrap(sample, n_boot)
6:             max_k(p) ← ∞
7:             tightness_best(p) ← ∞
8:             for k ∈ k_range, k_step do
9:                 vref ← q_est
10:                vpred ← MIK(k, sample_boot, p)
11:                tightness ← vpred/vref
12:                if tightness < 1 then
13:                    break
14:                end if
15:                if tightness < tightness_best(p) then
16:                    current_k(p) = k
17:                end if
18:            end for
19:            if current_k(p) < max_k(p) then
20:                max_k(p) = current_k(p)
21:            end if
22:        end for
23:    end for
24:    max_k(p_all) ← buildLinearModel(max_k(p_test), p_all)
25:    if corr(max_k(p_all)) < th then
26:        return no pWCET estimate
27:    end if
28:    return max_k(p_all)
29: end function
```

and hence generate three max_k points, one for each probability. With those points, we can generate the regression line that projects max_k for any probability of interest (e.g., those in Figure 9) and assess that the correlation coefficient is above the desired threshold.

Algorithm 2 generalizes the RESTK process starting from a main sample of the distribution under analysis (size 10^p), selecting the range of k to explore and the number of n_{sims} to run. First, RESTK estimates the quantiles at given probabilities p_{test} from the main sample (Line 3), e.g. $10^{-(p-1)}, 10^{-(p-2)}$, and $10^{-(p-3)}$. For each simulation, n_{boot} bootstrap samples of size 10^{p-3} are generated from the main sample (Line 5). For each of these samples, we compute the maximum k and maximum tightness for all the probabilities to test. The predicted value *vpred* is obtained by computing MIK from Equation 10 with the sample moment estimator, $\hat{E}(X^k)$ in Equation 12, (Line 10). The tightness of the predicted value is computed (Line 11) by using as reference value *vref* the estimated quantiles obtained before (Line 3). The algorithm finds the values of k in the considered range that produce the tightest estimate (Lines 15-16) and terminates its exploration as soon as a k that underestimates ($tightness < 1$) is found (Line 12). After exploring all selected probabilities for all k and all simulations, the algorithm returns the smallest max_k across all simulations (Lines 20-21).

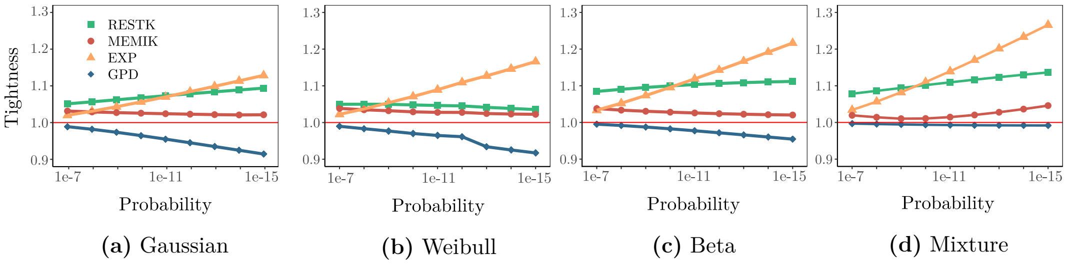


Figure 10 MEMIK with sample moments ($n = 1000$ and $n_{sims} = 100$) on the reference distributions, hence restricting k (RESTK). Also MEMIK evaluated with theoretical moments (MEMIK), and EVT evaluated with exponential tails (EXP) and with a GPD with light tails (GPD).

Once the max_k (e.g. for $10^{-(p-1)}$, $10^{-(p-2)}$, and $10^{-(p-3)}$) are obtained, RESTK builds a linear model for all possible probabilities (Line 24). The final check (Line 25) will ensure that the correlation of max_k is above $th = 0.95$, otherwise RESTK provides no pWCET estimate. As we show in Table 2, the correlation should always be close to 1 for max_k. The threshold $th = 0.95$ is a standard and stringent threshold for confidence in statistics, and used as a way to discard estimates that do not meet the safety criteria of finding a proper max_k.

RESTK enables the computation of a value for max_k that reduces the risk of underestimation for any probability of interest. This can be directly exploited by running MEMIK (Algorithm 1) on a predetermined range for k and for each probability p under study by using $max_k(p)$ as the upper bound for k, instead of considering an arbitrary range. Also, note that with RESTK, we use Equation 10 with sampled moments ($\hat{E}(X^k)$) as $EVAL(t, k, p)$ function in MEMIK. The rest of the MEMIK algorithm remains unchanged when using the RESTK method.

6 RESTK and EVT PWCET Estimates on Distributions

Our implementation of RESTK and MEMIK is programmed in R [40]. We run experiments on an Intel Core i5-7600K CPU clocked at 3.8GHz. The maximum execution time required per experiment was very short, 50 milliseconds or lower. We analyze values of k in the range $k \in [1, 150]$ with $k_{step} = 1$ to estimate max_k for all reference distributions, which, as shown in Figure 9 is a wide enough range to find the best max_k across distributions and probabilities. For all methods compared in this section, we use a sample size of $n = 10^6$. For RESTK, we set the number of bootstrap simulations to $n_{boot} = 2000$.

For EVT, we use the PoT methodology to fit tails and the CV Plot [12] to find the appropriate threshold for the PoT model. We use two EVT models fitted for pWCET estimation, namely exponential and light tails models. For each specific model, a different threshold using the CV plot will be found to ensure the best possible fit.

- *Exponential*: with an exponential model, the shape of the GPD is fixed to $\xi = 0$, which only leaves us to estimate the threshold u and the scale σ. The threshold is estimated with the CV Plot fixing $\xi = 0$, which finds where the exponential tail begins. Once we find the tail, we separate it from the rest of the sample and estimate the scale σ with it.
- *GPD light tails*: for the light tails model, we need the value of ξ, with $\xi < 0$, best fitting the data. Using the CV Plot we find the threshold u where the light tail begins. Then, we separate the tail from the sample and estimate the shape ξ and the scale σ.

Table 3 Distributions used for the analysis and their respective parameters.

Acronym	Type	Parameters
Gaussian1	Gaussian	$\mu = 100,\ \sigma = 10$
Gaussian2	Gaussian	$\mu = 100,\ \sigma = 50$
Weibull1	Weibull	$\alpha = 4,\ \lambda = 80$
Weibull2	Weibull	$\alpha = 8,\ \lambda = 80$
Beta1	Beta	$\alpha = 8,\ \beta = 1/4$
Beta2	Beta	$\alpha = 8,\ \beta = 1/8$
Gamma1	Gamma	$\alpha = 100,\ \lambda = 1$
Gamma2	Gamma	$\alpha = 150,\ \lambda = 1$
Mixture1	Mixture of Gaussians	$\mu = \{5, 50, 100\},\quad \sigma = 10,\ w = \{0.6, 0.39, 0.1\}$
Mixture2	Mixture of Gaussians	$\mu = \{50, 100, 400\},\ \sigma = 50,\ w = \{0.6, 0.39, 0.1\}$
Mixture3	Mixture of Weibulls	$\lambda = \{5, 50, 100\},\ \alpha = 4,\ w = \{0.6, 0.39, 0.1\}$
Mixture4	Mixture of Weibulls	$\lambda = \{5, 50, 100\},\ \alpha = 8,\ w = \{0.6, 0.39, 0.1\}$

Reference Distributions. We start the comparison with Figure 10 that depicts for the reference distributions the results of PoT with exponential and light tail models (EXP and GPD). It also shows the results obtained with MEMIK, which we can obtain as we have the actual distributions, and RESTK. Note that MEMIK provides the theoretical bound achievable with RESTK – it produces a safe bound and the tightest estimates. RESTK, EXP, and GPD build on a sample (the same one for a fair comparison) of the distribution. Following common practice, we show in Figure 10 the bias of our estimator, which is the expected value (mean) of RESTK output. It is noted that in RESTK application process all the distributions of this section fulfilled the linearity assessment (line 25 in Algorithm 2).

As we can see in this initial set of results, GPD tends to underestimate while EXP increases overestimation for high exceedance probabilities. RESTK produces values that are more consistent across all probabilities, improving EXP specially for higher exceedance probabilities. For 10^{-12} overestimates are 8%, 13%, 24% and 11% for the four reference distributions, respectively. The values increase to 13%, 20%, 37% and 17% for 10^{-15}. It can also be observed that the overestimation introduced by RESTK w.r.t. MEMIK to handle sampling uncertainty is limited: at 10^{-12} the difference is 4.75 percentage points (p.p) on average with a maximum of 8 p.p across all four reference distributions, and at 10^{-15} the overestimation difference is 5.25 p.p on average with a maximum of 10 p.p.

Extended set of Distributions. We consider a wider set of parameters for each distribution as listed in Table 3, resulting in 12 different distributions. The first distribution of each type (but the Gamma) is the reference distribution with the parameters used in previous sections. The rest of the distributions of each type encompass a different set of parameters to increase representativeness. The set of values explored for each parameter aims at showing the capabilities of RESTK under different scenarios. To that end we modify the following parameters.

1. the variance of the distribution for Gaussian1 and Gaussian2; and Mixtures1 and Mixture2
2. the shape of the tail of the distribution for Beta1 and Beta2, Weibull1 and Weibull2, Gamma1 and Gamma2, and Mixture3 and Mixture4.

Table 4 Tightness of the different models (MEK stands for MEMIK and RES for RESTK).

	probability 10^{-12}				probability 10^{-15}			
	GPD	EXP	MEK	RES	GPD	EXP	MEK	RES
Gaussian1	0.93	1.08	1.02	1.06	0.90	1.13	1.02	1.06
Gaussian2	0.90	1.20	1.05	1.14	0.86	1.28	1.03	1.11
Weibull1	0.91	1.13	1.02	1.09	0.87	1.20	1.02	1.09
Weibull2	0.96	1.09	1.03	1.04	0.94	1.14	1.05	1.04
Beta1	0.98	1.24	1.03	1.18	0.98	1.37	1.03	1.20
Beta2	0.98	1.17	1.03	1.11	0.98	1.26	1.02	1.13
Gamma1	0.93	1.09	1.03	1.07	0.89	1.11	1.03	1.07
Gamma2	0.95	1.09	1.02	1.06	0.92	1.13	1.02	1.07
Mixture1	0.92	1.11	1.03	1.03	0.88	1.17	1.02	1.02
Mixture2	0.94	1.16	1.03	1.07	0.90	1.23	1.03	1.05
Mixture3	0.88	1.25	1.03	1.15	0.83	1.37	1.02	1.13
Mixture4	0.94	1.15	1.02	1.15	0.91	1.23	1.03	1.16

This covers **all** possible variability scenarios as the scale and location do not affect the results for Markov's Inequality. As shown in [36], a change of location does not affect the inequality if the shift keeps the random variable positive, $P(X-a \geq b-a) \leq \frac{E(X-a)}{b-a} < \frac{E(X)}{b}$. Also, scaling the random variable X as λX where λ is real-valued, does not affect Markov's Inequality as $P(\lambda X \geq \lambda b) \leq \frac{E(\lambda X)}{\lambda b} = \frac{\lambda E(X)}{\lambda b} = \frac{E(X)}{b}$.

Looking at the results for the broader set of experiments in Table 4, we observe the following:

- GPD is always close to the true quantile, but in all cases it produces optimistic results. Furthermore, the higher exceedance probability, the more optimistic the estimate is. For instance, GPD on the Gaussian1 has a tightness of $\{0.93, 0.90\}$ at $p = \{10^{-12}, 10^{-15}\}$ respectively. This behavior is observed for all reference distributions.
- EXP follows the opposite pattern. In general, we can see in Figure 10 that, for an exceedance probability $p = 10^{-7}$, the estimates across distributions are always safe and quite tight. However, for higher exceedance probabilities, EXP tends to give more pessimistic estimates. For instance, EXP on the Gaussian1 has a tightness of 1.01 at $p = 10^{-7}$ that increases to 1.13 at $p = 10^{-15}$. At this probability and across all distributions EXP overestimation is 21.8% on average, 37% in the worst case.
- The estimates with MEMIK, i.e. with theoretical moments, are very tight, below 6% for all distributions.
- RESTK achieves results similar to MEMIK, preserving tightness and trustworthiness. Even for very high exceedance probabilities, RESTK is able to produce consistent estimates. Building on the Gaussian1 distribution, we see that while EXP can achieve a tighter estimate for low exceedance probabilities (e.g. $p = 10^{-7}$), EXP suffers from increased pessimism for higher exceedance probabilities whereas RESTK stays stable. This behavior is more striking in distributions harder to analyze like mixtures. For instance, for Mixture1 RESTK not only maintains tightness stable across probabilities, $\{1.03, 1.02\}$, but it gets also a tighter bound than EXP for probabilities $p = 10^{-9}$ and beyond as seen in Figure 10. At 10^{-15} RESTK estimates across all distribution overestimate by 9.4%, far below the 21.8% of EXP. In the worst case it is 20% (w.r.t. to 37% of EXP).

Overall, experimental results show the ability of RESTK to produce bounds suitable for pWCET estimation, being those trustworthy, tight and stable across probabilities and distributions, as opposed to existing models, which fail to meet all three goals simultaneously. The benefits of RESTK increase as the cutoff probability decreases to $10^{-12} - 10^{-15}$, which are the main range of interest for pWCET estimation considering maximum failure rates of 10^{-9} per hour and tasks running thousands of times per hour.

7 Railway Use Case

In order to evaluate the effectiveness of RESTK, we use an industrial critical real-time use case. In particular, we focus on the central safety processing unit of the European Train Control System (ETCS) reference architecture. The ETCS is a safety-critical application (SIL 4) responsible of signaling and control in the European Rail Traffic Management System (ERTMS) framework.

ETCS protects the train motion by constantly monitoring traveled distance and speed, and is programmed to activate the emergency break system whenever unauthorized speed values are detected. The ETCS subsystem comprises three main tasks that are executed sequentially to provide the required safety function: the Odometry module, estimating a set of parameters based on the inputs collected from the train environment (e.g., estimated position); the Service module, managing the Service braking system; and the Emergency module, actually controlling the Emergency braking system. While all three tasks do exhibit strict real-time requirements, we focus our evaluation on the Emergency module, the core of the ETCS safety-critical module.

The ETCS validation suite, which is made available with the application, includes 10 different input vectors (TEST0 to TEST9) corresponding to the operating conditions for functional and timing validation regarded as relevant by the application owner. We collected execution times for each input vector, which stimulates a different Emergency module operational scenario with its own timing distribution. Experiments were conducted on a Cobham Gaisler LEON3 platform.

Ground truth. For real programs, for which the actual distribution is not known, common practice consists in using as 'ground truth' the observed quantiles for samples as large as reasonably possible (e.g. 10^4 [34] and 10^8 [46]). We follow the same approach and consider the quantile observed for the 10^7 sample as the reference value. More than two weeks of execution were needed to complete the execution of all the 10^7 runs per input vector (TEST).

Setup. The setup and parameters used to run RESTK are the same used for the reference distributions. As the number of runs we have is $n = 10^7$ per input set, we make projections for $p = 10^{-6}$, for which the observed frequencies closely match the actual probability (i.e. 95% confidence intervals are within 0.1% of the mean). In this case, the estimated quantile at probability $p = 10^{-6}$ is our ground truth. Only TEST6 is an exception to this and, due to the variability observed for that quantile, we use a 95% confidence interval, which is 1% off the mean. In this section, we use a sample size of $n = 10^4$ for GPD, EXP and RESTK. Note that MEMIK with theoretical moments cannot be used since the actual distribution is unknown and we only have sampled data.

Results. As part of the RESTK application process, all the 10 distributions fulfilled the linearity assessment (line 25 in Algorithm 2). As shown in Table 5, pWCET estimates are similar to those presented in the previous section. In particular:

Table 5 Tightness of the estimates for the ETCS case study.

	probability 10^{-6}						
	GPD	EXP	RESTK		GPD	EXP	RESTK
TEST0	1.01	1.21	1.10	TEST5	0.98	1.12	1.07
TEST1	0.98	1.12	1.06	TEST6	0.94	1.06	1.01
TEST2	1.00	1.13	1.11	TEST7	1.01	1.16	1.09
TEST3	0.98	1.20	1.10	TEST8	1.01	1.23	1.12
TEST4	0.99	1.17	1.13	TEST9	0.98	1.11	1.10

- GPD achieves extremely tight estimates for 4 tests, with tightness up to 1.01, but on the other 6 tests it produces optimistic results. In general, while potentially very tight, GPD can easily underestimate the bounds.
- On the other hand, EXP never underestimates although it produces pessimistic results, as high as 23% for such a relatively low quantile.
- RESTK consistently produces tighter estimates than EXP in all tests for the use case. On average, EXP exhibits 15.1% pessimism, whereas RESTK reduces it to 8.9%.

Discussion. Overall, with the combined results over a wide set of distributions, shown in the previous section, and the results for the ETCS case study presented in this section, we conclude that RESTK consistently provides tighter estimates than EXP and improves for lower exceedance probabilities.

8 Related Works

Probabilistic and statistical approaches [13, 18] have been increasingly considered as a promising solution to cope with the rise in complexity of hardware and software systems, as determined by unprecedented increasing computing performance requirements [20].

Extreme Value Theory (EVT), a consolidated approach for modeling and predicting the occurrence of rare events, has emerged as the preferred option for modeling the worst-case execution time (WCET) of software programs. EVT has been considered particularly fit for probabilistic modeling of WCET as the latter is normally considered a rare event in the program's timing behavior. EVT is at the foundation of several Measurement-Based Probabilistic Timing Analysis (MBPTA) approaches [1, 7, 13, 27, 46, 52], which have been already positively assessed in some industrial use cases [54]. Probabilistic approaches building on sample and population sizes have also been built for overlapping concerns across task scheduling and timing analysis [39], hence being orthogonal to WCET estimation. Several works assess the necessary conditions for a correct application of EVT to the timing problem since an inattentive application of the EVT statistical tools can severely affect both trustworthiness and quality of the derived probabilistic WCET (pWCET) bounds [24, 34, 38].

Several studies focus on EVT applicability preconditions on the (timing) observations being independent and identically distributed (i.i.d.) random variables [13, 17]. EVT has also been shown to be applicable also to stationary data preserving extremal independence [44]. Different statistical tests have been assessed for that purpose in the real-time literature [1, 13, 17, 42]. Also, platform randomization [31] or data sample randomization [33] have been used to meet the i.i.d. requirement. However, even in case such statistical preconditions are met, the reliability of the obtained pWCET bounds is still affected by the choice in the EVT inputs (i.e. selection of samples belonging to the tail) and parameters of the fit distribution. Several methods have been proposed for selecting and assessing the quality of EVT parameters and sample selection and, in turn, their impact on the trustworthiness of the computed bounds [1, 4, 43].

Statistical tests have been also proposed, for example, to assess the reliability of pWCET bounds in [4]. Unfortunately, model uncertainty in EVT application cannot be removed or quantified since no existing approach provides an exact, optimal method for tail selection. A different challenge to pWCET reliability relates to *representativeness* of the observations [2, 4, 25, 38] which corresponds, in the timing domain, to guaranteeing the observations are actually representative or upperbound the execution time distribution, which in fact, is an inherent trait for any sampled-based approach. Statistical and model uncertainties in the scope of EVT are discussed in [8] where robustness in tail estimation is achieved by considering, together with GEV, a family of *plausible* probability models (i.e., close to GEV) and selecting the most conservative estimated probability value among all models. Markov's and Chebyshev's inequalities have been historically applied in a wide variety of fields such as Engineering [50], Big Data sampling [45], and radiative transfer [51]. In the context of real-time computing, moment-based bounds on tail probabilities have also been considered in the scope of probabilistic schedulability analysis [14, 15, 21, 53]. Chernoff bounds [14, 15] and generalizations thereof [53] are exploited to compute the cumulative distribution of the interference caused by higher-priority tasks on a task response time, ultimately delivering a probability for deadline misses. While building on application of Markov's inequality to the timing dimension, these approaches do not address the problem of computing probabilistic bounds to tasks' execution time, which are instead assumed to be available, but offer a scalable alternative to the computational complexity of convolutions. Some works [35, 41] consider the use of Chebyshev's inequality for WCET and/or cache hit and miss rates estimation. However, those works consider Chebyshev's inequality (without power-of-k functions) only to estimate the impact of the variance on those metrics, and focus on the analysis of statistical uncertainties. Other works consider using higher moments to improve concentration inequalities similar to Markov [32]. Although their approach is similar, the work focuses purely on the theoretical tightness of the probability bounding. The challenge we overcame in our work was translating this theoretical advantage into a practical reality for unknown distributions, which can easily lead to optimistic bounds without proper care as we shown. That is, neither model uncertainties nor tightness aspects for pWCET estimation for high quantiles are considered. Authors in [58] consider a similar approach to those works for WCET estimation, but discard Chebyshev's inequality altogether given the pessimism expected for high quantiles.

9 Conclusions

In this work we presented for the first time a method based on Markov's Inequality for pWCET estimation that represents a solid alternative approach to EVT. In particular, we showed that MIK (Markov Inequality to the power-of-k) has no model uncertainty and proposed a method to handle sampling uncertainty (RESTK) that consistently provides more trustworthy, tighter and stable results than EVT in different scenarios, including a railway case study. These promising results suggest that RESTK can be effectively used as a standalone method for pWCET estimation, or even as an alternative approach to validate EVT results in those cases where EVT is already consolidated. In this line, the fact that MIK (RESTK) and EVT build on completely different mathematical foundations provides stronger evidence on the trustworthiness of the obtained pWCET estimates.

References

1 Jaume Abella, Maria Padilla, Joan Del Castillo, and Francisco J. Cazorla. Measurement-based worst-case execution time estimation using the coefficient of variation. *ACM Trans. Des. Autom. Electron. Syst.*, 22(4), June 2017. doi:10.1145/3065924.

2 Jaume Abella, Eduardo Quiñones, Franck Wartel, Tullio Vardanega, and Francisco J. Cazorla. Heart of gold: Making the improbable happen to increase confidence in MBPTA. In *26th Euromicro Conference on Real-Time Systems, ECRTS 2014, Madrid, Spain, July 8-11, 2014*, pages 255–265. IEEE Computer Society, 2014. doi:10.1109/ECRTS.2014.33.

3 Charalampos Antoniadis, Dimitrios Garyfallou, Nestor Evmorfopoulos, and Georgios Stamoulis. Evt-based worst case delay estimation under process variation. In *2018 Design, Automation Test in Europe Conference Exhibition (DATE)*, pages 1333–1338, 2018. doi:10.23919/DATE.2018.8342220.

4 Luis Fernando Arcaro, Karila Palma Silva, Rômulo Silva de Oliveira, and Luís Almeida. Reliability test based on a binomial experiment for probabilistic worst-case execution times. In *41st IEEE Real-Time Systems Symposium, RTSS 2020, Houston, TX, USA, December 1-4, 2020*, pages 51–62. IEEE, 2020.

5 August Aimé Balkema and Laurens de Haan. Residual Life Time at Great Age. *The Annals of Probability*, 2(5):792–804, 1974. doi:10.1214/aop/1176996548.

6 Francesco Bartolucci and Luca Scrucca. Point estimation methods with applications to item response theory models. In Penelope Peterson, Eva Baker, and Barry McGaw, editors, *International Encyclopedia of Education (Third Edition)*, pages 366–373. Elsevier, Oxford, third edition edition, 2010. doi:10.1016/B978-0-08-044894-7.01376-2.

7 Kostiantyn Berezovskyi, Luca Santinelli, Konstantinos Bletsas, and Eduardo Tovar. WCET Measurement-Based and Extreme Value Theory Characterisation of CUDA Kernels. In *Proceedings of the 22nd International Conference on Real-Time Networks and Systems*, RTNS '14, pages 279–288, New York, NY, USA, 2014. Association for Computing Machinery. doi:10.1145/2659787.2659827.

8 Jose Blanchet, Fei He, and Karthyek Murthy. On distributionally robust extreme value analysis. *Extremes*, 23:317–347, 2020. doi:10.1007/s10687-019-00371-1.

9 Frederico Caeiro and Maria Gomes. Semi-parametric tail inference through probability-weighted moments. *Journal of Statistical Planning and Inference*, 141(2):937–950, 2011. doi:10.1016/j.jspi.2010.08.015.

10 Frederico Caeiro and Maria Gomes. Threshold selection in extreme value analysis. *Extremes*, September 2014. doi:10.1007/s10687-021-00405-7.

11 Enrique Castillo, Ali Hadi, Narayanaswamy Balakrishnan, and José Sarabia. *Extreme Value and Related Models With Applications in Engineering and Science*. Wiley series in probability and statistics. Whiley, January 2005.

12 Joan Del Castillo, Jalila Daoudi, and Richard Lockhart. Methods to distinguish between polynomial and exponential tails. *Scandinavian Journal of Statistics*, 41(2):382–393, 2014. doi:10.1111/sjos.12037.

13 Francisco J. Cazorla, Leonidas Kosmidis, Enrico Mezzetti, Carles Hernandez, Jaume Abella, and Tullio Vardanega. Probabilistic worst-case timing analysis: Taxonomy and comprehensive survey. *ACM Comput. Surv.*, 52(1), February 2019. doi:10.1145/3301283.

14 Kuan-Hsun Chen and Jian-Jia Chen. Probabilistic schedulability tests for uniprocessor fixed-priority scheduling under soft errors. In *2017 12th IEEE International Symposium on Industrial Embedded Systems (SIES)*, pages 1–8, 2017. doi:10.1109/SIES.2017.7993392.

15 Kuan-Hsun Chen, Niklas Ueter, Georg von der Brüggen, and Jian-Jia Chen. Efficient computation of deadline-miss probability and potential pitfalls. In *2019 Design, Automation Test in Europe Conference Exhibition (DATE)*, pages 896–901, 2019. doi:10.23919/DATE.2019.8714908.

16 Stuart Coles. *An Introduction to Statistical Modeling of Extreme Values*. Springer, 2001. doi:10.1007/978-1-4471-3675-0.

17 Liliana Cucu-Grosjean, Luca Santinelli, Michael Houston, Codé Lo, Tullio Vardanega, Leonidas Kosmidis, Jaume Abella, Enrico Mezzetti, Eduardo Quiñones, and Francisco J. Cazorla. Measurement-based probabilistic timing analysis for multi-path programs. In *2012 24th Euromicro Conference on Real-Time Systems*, pages 91–101, July 2012. doi:10.1109/ECRTS.2012.31.

18 Robert Davis and Liliana Cucu-Grosjean. A survey of probabilistic schedulability analysis techniques for real-time systems. *Leibniz Transactions on Embedded Systems*, 6(1):04–1–04:53, 2019. doi:10.4230/LITES-v006-i001-a004.

19 Anthony C. Davison and Richard L. Smith. Models for exceedances over high thresholds. *Journal of the Royal Statistical Society. Series B (Methodological)*, 52(3):393–442, 1990. doi:10.1111/j.2517-6161.1990.tb01796.x.

20 Deloitte. *Semiconductors – the Next Wave Opportunities and winning strategies for semiconductor companies*, 2019. URL: https://www2.deloitte.com/content/dam/Deloitte/cn/Documents/technology-media-telecommunications/deloitte-cn-tmt-semiconductors-the-next-wave-en-190422.pdf.

21 Jose L. Diaz, Daniel F. Garcia, Kanghee Kim, Chang-Gun Lee, Lucia Lo Bello, José M. Lopez, Sang Lyul Min, and Orazio Mirabella. Stochastic analysis of periodic real-time systems. In *23rd IEEE Real-Time Systems Symposium, 2002. RTSS 2002.*, pages 289–300, 2002. doi:10.1109/REAL.2002.1181583.

22 Paul Embrechts, Thomas Mikosch, and Claudia Klüppelberg. *Modelling Extremal Events: For Insurance and Finance*. Springer-Verlag, Berlin, Heidelberg, 1997.

23 Ronald A. Fisher. Moments and product moments of sampling distributions. *Proceedings of the London Mathematical Society*, s2-30(1):199–238, 1930. doi:10.1112/plms/s2-30.1.199.

24 Samuel Jimenez Gil, Iain Bate, George Lima, Luca Santinelli, Adriana Gogonel, and Liliana Cucu-Grosjean. Open challenges for probabilistic measurement-based worst-case execution time. *IEEE Embedded Systems Letters*, 2017. doi:10.1109/LES.2017.2712858.

25 Fabrice Guet, Luca Santinelli, and Jérôme Morio. On the representativity of execution time measurements: Studying dependence and multi-mode tasks. In Jan Reineke, editor, *17th International Workshop on Worst-Case Execution Time Analysis, WCET 2017, June 27, 2017, Dubrovnik, Croatia*, volume 57 of *OASICS*, pages 3:1–3:13. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2017. doi:10.4230/OASIcs.WCET.2017.3.

26 Paul R. Halmos. The Theory of Unbiased Estimation. *The Annals of Mathematical Statistics*, 17(1):34–43, 1946. doi:10.1214/aoms/1177731020.

27 Jeffery Hansen, Scott Hissam, and Gabriel A. Moreno. Statistical-Based WCET Estimation and Validation. In Niklas Holsti, editor, *9th International Workshop on Worst-Case Execution Time Analysis (WCET'09)*, volume 10 of *OpenAccess Series in Informatics (OASIcs)*, pages 1–11, Dagstuhl, Germany, 2009. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. also published in print by Austrian Computer Society (OCG) with ISBN 978-3-85403-252-6. doi:10.4230/OASIcs.WCET.2009.2291.

28 Bruce M. Hill. A Simple General Approach to Inference About the Tail of a Distribution. *The Annals of Statistics*, 3(5):1163–1174, 1975. doi:10.1214/aos/1176343247.

29 Rob Hyndman and Yanan Fan. Sample quantiles in statistical packages. *The American Statistician*, 50:361–365, November 1996. doi:10.1080/00031305.1996.10473566.

30 Norman Lloyd Johnson. *Continuous univariate distributions. Vol. 2.* Wiley and Sons, New York, 2nd ed. / norman l. johnson, samuel kotz, n. balakrishnan. edition, 1994.

31 Leonidas Kosmidis, Jaume Abella, Eduardo Quiñones, and Francisco J. Cazorla. A cache design for probabilistically analysable real-time systems. In Enrico Macii, editor, *Design, Automation and Test in Europe, DATE 13, Grenoble, France, March 18-22, 2013*, pages 513–518. EDA Consortium San Jose, CA, USA / ACM DL, 2013. doi:10.7873/DATE.2013.116.

32 Bar Light. Concentration inequalities using higher moments information. *arXiv*, 2020. doi:10.48550/ARXIV.2006.05130.

33 George Lima and Iain Bate. Valid Application of EVT in Timing Analysis by Randomising Execution Time Measurements. In *2017 IEEE Real-Time and Embedded Technology and Applications Symposium (RTAS)*, pages 187–198, 2017. doi:10.1109/RTAS.2017.17.

34 George Lima, Dário Dias, and Edna Barros. Extreme Value Theory for Estimating Task Execution Time Bounds: A Careful Look. In *Euromicro Conference on Real-Time Systems, ECRTS*, 2016. doi:10.1109/ECRTS.2016.20.

35 Jyh-Charn S. Liu and Sharif M. Shahrier. On predictability of caches for real-time applications. In *Proceedings of International Workshop on Modeling, Analysis and Simulation of Computer and Telecommunication Systems*, pages 52–56, 1994. doi:10.1109/MASCOT.1994.284448.

36 Andrey Markov. On certain applications of algebraic continued fractions. *Ph.D. thesis, St. Petersburg*, 1884.

37 Thomas Mikosch. Regular variation, subexponentiality and their applications in probability theory. *International Journal of Production Economics - INT J PROD ECON*, January 1999.

38 Suzana Milutinovic, Enrico Mezzetti, Jaume Abella, Tullio Vardanega, and Francisco J. Cazorla. On uses of Extreme Value Theory fit for industrial-quality WCET analysis. In *12th IEEE International Symposium on Industrial Embedded Systems, SIES 2017, Toulouse, France, June 14-16, 2017*, pages 1–6. IEEE, 2017. doi:10.1109/SIES.2017.7993402.

39 Sims Osborne and James H. Anderson. Simultaneous multithreading and hard real time: Can it be safe? In Marcus Völp, editor, *32nd Euromicro Conference on Real-Time Systems, ECRTS 2020, July 7-10, 2020, Virtual Conference*, volume 165 of *LIPIcs*, pages 14:1–14:25. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020. doi:10.4230/LIPIcs.ECRTS.2020.14.

40 R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2021. URL: https://www.R-project.org/.

41 Archana Ravindar and Y. N. Srikant. Estimation of probabilistic bounds on phase CPI and relevance in WCET analysis. In *Proceedings of the Tenth ACM International Conference on Embedded Software*, EMSOFT '12, pages 165–174, New York, NY, USA, 2012. Association for Computing Machinery. doi:10.1145/2380356.2380388.

42 Federico Reghenzani, Giuseppe Massari, and William Fornaciari. Probabilistic-WCET reliability: Statistical testing of EVT hypotheses. *Microprocess. Microsystems*, 77:103–135, 2020. doi:10.1016/j.micpro.2020.103135.

43 Federico Reghenzani, Luca Santinelli, and William Fornaciari. Dealing with uncertainty in pWCET estimations. *ACM Trans. Embed. Comput. Syst.*, 19(5):33:1–33:23, 2020. doi:10.1145/3396234.

44 Luca Santinelli, Jérôme Morio, Guillaume Dufour, and Damien Jacquemart. On the Sustainability of the Extreme Value Theory for WCET Estimation. In Heiko Falk, editor, *14th International Workshop on Worst-Case Execution Time Analysis*, volume 39 of *OpenAccess Series in Informatics (OASIcs)*, pages 21–30, Dagstuhl, Germany, 2014. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. doi:10.4230/OASIcs.WCET.2014.21.

45 Ashwin Satyanarayana. Intelligent sampling for big data using bootstrap sampling and Chebyshev inequality. In *2014 IEEE 27th Canadian Conference on Electrical and Computer Engineering (CCECE)*, pages 1–6, 2014. doi:10.1109/CCECE.2014.6901029.

46 Karila Palma Silva, Luis Fernando Arcaro, and Romulo Silva De Oliveira. On Using GEV or Gumbel Models When Applying EVT for Probabilistic WCET Estimation. In *2017 IEEE Real-Time Systems Symposium (RTSS)*, pages 220–230, 2017. doi:10.1109/RTSS.2017.00028.

47 Didier Sornette. *Critical Phenomena in Natural Sciences: Chaos, Fractals, Selforganization and Disorder: Concepts and Tools.* Springer, January 2006. doi:10.1007/3-540-33182-4.

48 Seth M. Steinberg and Clarence E. Davis. Distribution-free confidence intervals for quantiles in small samples. *Communications in Statistics - Theory and Methods*, 14(4):979–990, 1985. doi:10.1080/03610928508805144.

49 Pafnuti Tchebichef. Des valeurs moyennes. *Journal de mathématiques pures et appliquées*, 12(2):177–184, 1867.

50 Vladimir Utkin. Calculating the reliability of machine parts on the basis of the Chebyshev inequality. *Russian Engineering Research*, 32, January 2012. doi:10.3103/S1068798X11120264.

51 Gladimir V.G. Baranoski, Jon G. Rokne, and Guangwu Xu. Applying the exponential Chebyshev inequality to the nondeterministic computation of form factors. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 69(4):447–467, 2001. doi:10.1016/S0022-4073(00)00095-9.

52 Sergi Vilardell, Isabel Serra, Jaume Abella, Joan Del Castillo, and Francisco J. Cazorla. Software timing analysis for complex hardware with survivability and risk analysis. In *2019 IEEE 37th International Conference on Computer Design (ICCD)*, pages 227–236, 2019. doi:10.1109/ICCD46524.2019.00036.

53 Georg von der Brüggen, Nico Piatkowski, Kuan-Hsun Chen, Jian-Jia Chen, and Katharina Morik. Efficiently approximating the probability of deadline misses in real-time systems. In *ECRTS*, 2018. doi:10.4230/LIPIcs.ECRTS.2018.6.

54 Franck Wartel, Leonidas Kosmidis, Adriana Gogonel, Andrea Baldovin, Zoë R. Stephenson, Benoit Triquet, Eduardo Quiñones, Code Lo, Enrico Mezzetti, Ian Broster, Jaume Abella, Liliana Cucu-Grosjean, Tullio Vardanega, and Francisco J. Cazorla. Timing analysis of an avionics case study on complex hardware/software platforms. In Wolfgang Nebel and David Atienza, editors, *Proceedings of the 2015 Design, Automation & Test in Europe Conference & Exhibition, DATE 2015, Grenoble, France, March 9-13, 2015*, pages 397–402. ACM, 2015. doi:10.7873/DATE.2015.0189.

55 Reinhard Wilhelm, Jakob Engblom, Andreas Ermedahl, Niklas Holsti, Stephan Thesing, David B. Whalley, Guillem Bernat, Christian Ferdinand, Reinhold Heckmann, Tulika Mitra, Frank Mueller, Isabelle Puaut, Peter P. Puschner, Jan Staschulat, and Per Stenström. The worst-case execution-time problem - overview of methods and survey of tools. *ACM Trans. Embed. Comput. Syst.*, 7(3):36:1–36:53, 2008. doi:10.1145/1347375.1347389.

56 Julien Worms and Sid Touati. Parametric and Non-Parametric Statistics for Program Performance Analysis and Comparison. *[Research Report] RR-8875, INRIA Sophia Antipolis - I3S; Université NiceSophia Antipolis; Université Versailles Saint Quentin en Yvelines; Laboratoire de mathématiques deVersailles*, 2017. URL: https://hal.inria.fr/hal-01286112.

57 Julien Worms and Sid Touati. Modelling program's performance with gaussian mixtures for parametric statistics. *IEEE Transactions on Multi-Scale Computing Systems*, 4(3):383–395, 2018. doi:10.1109/TMSCS.2017.2754251.

58 Pavel G. Zaykov and Jan Kubalčik. Worst-case measurement-based statistical tool. In *2019 IEEE Aerospace Conference*, pages 1–10, 2019. doi:10.1109/AERO.2019.8741824.