POLITECNICO
MILANO 1863

Tutorial

The Extreme Value Theory in Probabilistic Real-Time Computing

Federico Reghenzani
Giuseppe Massari
William Fornaciari

Politecnico di Milano

October 15, 2018

Abstract

In this article we introduce the statistical theory foundation of probabilistic real-time computing: the Extreme Value Theory. This tutorial aims at introducing the basis of the theory and provide some insights on how to apply it to the real-time computing problems. In fact, a method called Probabilistic Real-Time Computing has been recently proposed to overcome the issues in estimating the worst-case execution time with traditional approaches on modern processing architectures. Provided that certain hypotheses are satisfied, the Extreme Value Theory is able to produce a probabilistic estimation of the worst-case timing behaviour of a real-time task, i.e. the statistical distribution of its execution times.

This tutorial is regularly updated and improved. The current version is October 2018 (v.1).

1 Introduction

1.1 Why Probabilistic approaches

In recent years, it is possible to notice an extreme rising of the processors architecture complexity due to, on one hand, the increasing computational power demand, and on the other hand, the single-core power and performance barriers. Unfortunately, adding complexity makes harder to create reliable models of the processors. In particular, (hard) real-time computing requires a detailed knowledge of the timing profile of the operations on a specific processor, in order to be able to compute the **Worst-Case Execution Time (WCET)** of tasks to deploy. In modern architectures, mechanisms such as complex pipelines, hardware threading, multi-level and/or shared levels of caches, jump speculation, etc. make the computation of the WCET very hard or even impossible. The *unpredictability problem* is even amplified when we consider multi-/many-core platforms, with simultaneously running tasks that produce hard-to-analyze interferences.

For the previous reasons, traditional methodologies of static WCET analyses fail to obtain safe and tight WCET estimation for modern platforms, in a sustainable amount of computation time. Obviously, considering substantial limitations – e.g. one level of non-shared cache or dual-core processor without cache, etc. – is still possible to compute the WCET, but the performance obtained by such embedded system are still very far from the high-end general-purpose processor. One of the proposed approach to overcome these issues is rely on **probabilistic real-time computing**.

1.2 Probabilistic classes

In the last years the researchers have distinguished between two possible classes of probabilistic analyses: **Static Probabilistic Timing Analysis (SPTA)** and **Measurement-Based Probabilistic Timing Analysis (MBPTA)**. The first one, SPTA, is similar to the traditional analyses with the addition of probabilistic terms to the well known methods. This approach is still immature and not much considered in current ongoing research. The second method, MBPTA, is based on the direct observation of the execution time of a task. We will describe MBPTA in Section 3, after providing the statistical basis necessary to understand the measurement-based mechanism. In particular, this method is based on the *Extreme Value Theory* described in the next section. For completeness, it is worth to cite the existence of mixed approaches called **Hybrid Probabilistic Timing Analysis**, that we will not discuss in this tutorial.

2 Extreme Value Theory

The **Extreme Value Theory (EVT)** is a statistical theory developed in the 20th century to deal with the *extreme events*, i.e. events that have a probability to occur so low that have been never observed. An *extreme* is defined as the *maximum* or the *minimal* value that a random variable can assume. The result of this theory is opposed to the well known central theorem limit, that instead deals with the average values of the distribution.

The EVT has been traditionally used to predict natural disasters. For example, while measuring the water level in a river, it is possible to estimate the probability of an unseen water level that may possibly cause a flood. The EVT has been also applied to other fields, such as financial risk studies, drug industry, anthropological studies, etc.

2.1 Statistical findings

Let $X_1, X_2, ..., X_n$ a sequence of n iid[1] random variables representing the multiple observations of a phenomena. The Extreme Value Theory, in case we are interested in the maxima, looks for the following

[1] Independent and Identically Distributed

cumulative distribution function (cdf):

$$
\begin{aligned}
F^*(x) &= 1 - P(x \geq \max(X_1, X_2, ..., X_n)) \\
&= 1 - P(x \geq X_1, x \geq X_2, ..., x \geq X_n) \\
&= 1 - \prod_{i=1}^{n} P(x \geq X_i)
\end{aligned} \tag{1}
$$

The last expansion is guaranteed by the fact the the random variables are iid. In this tutorial we always refer to the maxima and not to the minima, however, the results can be easily changed to obtain the minima estimation.

From the Equation 1 it is possible to say that the cdf $F^*(x)$ is governed by the product of a sequence of cdf: $F^n(x)$. The most important result of the EVT is that there exists a sequence $a_n \in \mathbb{R}$ and a sequence b_n such that:

$$\lim_{n\to\infty} F^n(b_n + a_n x) = G(x) \tag{2}$$

where $G(x)$ is the cdf of a statistical distribution called *extreme value distribution*. The interesting thing of this formula is the **Fisher-Tippett-Gnedenko theorem**: if the previous limit is non-degenerating, $G(x)$ is a cdf of well known distributions, independently on the form of $F(x)$.

2.2 Extreme distributions

The original Fisher-Tippett-Gnedenko theorem provides three possible form for $G(x)$: Weibull, Fréchet and Gumbell distributions. However, it has been later proved that these distributions can be generalized in the **Generalized Extreme Value Distribution (GEVD)** that has the following form:

$$G(x) = \begin{cases} e^{-e^{\frac{x-\mu}{\sigma}}} & \xi = 0 \\ e^{-[1+\xi(\frac{x-\mu}{\sigma})]^{-1/\xi}} & \xi \neq 0 \end{cases} \tag{3}$$

The distribution has three parameters: the *location* μ, the *scale* σ and the *shape* ξ. The sign of ξ depends on the used convention; we used the following one: if $\xi > 0$ the GEVD converges to the Fréchet distribution, if $\xi < 0$ it converges to the Weibull distribution, otherwise $\xi \to 0$ and it converges to the Gumbel distribution. This distribution is consequently written as GEVD(μ, σ, ξ).

Another, but asymptotically equivalent, formulation for $G(x)$ is the **Generalized Pareto Distribution (GPD)** that has the following form:

$$G(x) = \begin{cases} 1 - e^{\frac{x-\mu}{\sigma}} & \xi = 0 \\ 1 - [1 + \xi(\frac{x-\mu}{\sigma})]^{-1/\xi} & \xi \neq 0 \end{cases} \tag{4}$$

The parameters are the same of GEVD, but in some books the GPD is expressed with its two parameters form, considering $\mu = 0$. If $\xi = 0$ the GPD converges to the exponential distribution, if $\xi > 0$ and $\mu = \frac{\sigma}{\xi}$ GPD converges to the Pareto distribution. The GPD distribution is asymptotically equivalent to the GEVD distribution when $n \to \infty$. This distribution is consequently written as GPD(μ, σ, ξ). We refer with GPD(σ, ξ) when the two parameters version is considered. Sometimes to distinguish these variants, the acronyms become GP2 and GP3.

These distributions estimate the tail (maxima or minima) of the original unknown distribution.

2.3 Estimation approaches

There exist two possible estimation classes in order to obtain valid data to fit GEVD or GPD. The first one is called **Block-Maxima**: to the sequence $X_1, X_2, ..., X_n$ the following *filter* is applied:

$$Y_i = max(X_{B\cdot(i-1)}, X_{B\cdot(i-1)+1}, ..., X_{B\cdot i}) \tag{5}$$

where B is a parameter called *block size*. In other words, the set of observations is divided in blocks of fixed size B and for each block the maximum value is taken. The sequence $Y_1, Y_2, ..., Y_B$ represents the maxima of the blocks. With the realization of these random variables it is possible to estimate the GEVD using any traditional algorithm, e.g. the Maximum Likelihood Estimator.

The second approach is called **Peak-over-Threshold** and it is a simple threshold filter:

$$Y = \{X_i > u\} \tag{6}$$

where u is a predefined threshold. The sequence $Y_1, Y_2, ..., Y_m$ represents the value of measurements that are greater than the threshold u. With the realization of these random variables it is possible to estimate the GPD using any traditional algorithm, e.g. the Maximum Likelihood Estimator.

3 Measurement-Based approaches

In the scenario of probabilistic real-time computing, the random variables represent the execution times of a given task running multiple times on a target system. Obtaining these execution times in a way that verifies the iid hypothesis is challenging and it is still an open research topic. The resulting time trace can be used as input of BM or PoT algorithm, as previously described, and then it is possible to estimate the GEV or GPD distribution. This distribution represents the so-called **probabilistic-WCET (pWCET)**, i.e. the statistical distribution of the extreme execution times of a given task. It is convenient to write the pWCET distribution with its cumulative distribution function:

$$p = P\left(X > \overline{\text{WCET}}\right) \tag{7}$$

where X represents the random variable of the execution time samples, $\overline{\text{WCET}}$ is a constant, and p is the violation probability. Given this cdf, it is possible to:

- Fix a value for $\overline{\text{WCET}}$ and compute the probability p of observing an execution time longer than this value
- Fix a value for the probability p and compute the value $\overline{\text{WCET}}$ for which the probability p represents the probability of observing a longer execution time.

However, in order to define some reliability comparisons, it is necessary to define the order of two distribution. We say that pWCET$_a$ is higher than pWCET$_b$ (pWCET$_a$ >pWCET$_b$) if:

- $\forall$ $\overline{\text{WCET}}$ it is true that p_a > p_b;
- $\forall$ p it is true that $\overline{\text{WCET}}_a$ > $\overline{\text{WCET}}_b$.

with this order relationship it is possible to write reliability consideration, since it is possible to say that a pWCET "upper-bounds" or not the real distribution. In fact, similar to the traditional WCET analyses, the pWCET estimation is safe iff its distribution upper-bounds the real distribution.

Bibliography

Enrique Castillo, Ali S. Hadi, N. Balakrishnan, José M. Sarabia, *Extreme Value and Related Models with Applications in Engineering and Science*, Wiley series in Probability and Statistics, ISBN 978-0-471-67172-X, 2005.

Rolf-Dieter Reiss, Thomas Michael *Statistical Analysis of Extreme Values with applications to insurance, finance, hydrology and other fields*, Birkhäuser Basel, Springer, ISBN 978-3-7643-7230-9, 2007.

Guillem Bernat, Antoine Colin, Stefan M. Petters *WCET analysis of probabilistic hard real-time systems*, Real-Time Systems Symposium, IEEE, 279–288, 2002

George Lima, DÃąrio Dias, Edna Barros *Extreme value theory for estimating task execution time bounds: A careful look* Real-Time Systems (ECRTS), 2016 28th Euromicro Conference on, IEEE, 2016, pages 200–211

Luca Santinelli, Fabrice Guet, JÃľrÃťme Morio, *Revising Measurement-Based Probabilistic Timing Analysis*, IEEE Real-Time and Embedded Technology and Applications Symposium (RTAS), 2017, 199–208.

Federico Reghenzani, Giuseppe Massari, William Fornaciari, *chronovise: Measurement-Based Probabilistic Timing Analysis framework*, Journal of Open Source Software, 2018, 3(28), 711, doi:10.21105/joss.00711.