Execution time distributions in embedded safety-critical systems using extreme value theory

Joan del Castillo and Maria Padilla*

Departament de Matemàtiques,
Universitat Autònoma de Barcelona,
08193 Bellaterra (Cerdanyola del Vallès), Spain
E-mail: castillo@mat.uab.cat
E-mail: mpadilla@mat.uab.cat
*Corresponding author

Jaume Abella

Barcelona Supercomputing Center (BSC),
08034 Barcelona, Spain
E-mail: jaume.abella@bsc.es

Francisco J. Cazorla

Barcelona Supercomputing Center (BSC) and
Spanish National Research Council (IIIA-CSIC),
08034 Barcelona, Spain
E-mail: francisco.cazorla@bsc.es

Abstract: Several techniques have been proposed to upper-bound the worst-case execution time behaviour of programs in the domain of critical real-time embedded systems. These computing systems have strong requirements regarding the guarantees that the longest execution time a program can take is bounded. Some of those techniques use extreme value theory (EVT) as their main prediction method.

In this paper EVT is used to estimate a high quantile for different types of execution time distributions observed for a set of representative programs for the analysis of automotive applications. A major challenge appears when the data set seems to be heavy tailed, because this contradicts the previous assumption of embedded safety-critical systems.

A methodology based on the coefficient of variation is introduced for a threshold selection algorithm to determine the point above which the distribution can be considered generalised Pareto distribution. This methodology also provides an estimation of the extreme value index and high quantile estimates. We have applied these methods to execution time observations collected from the execution of 16 representative automotive benchmarks to predict an upper-bound to the maximum execution time of this programs. Several comparisons with alternative approaches are discussed.

Keywords: worst-case execution times; extreme value theory; generalised Pareto distribution; threshold exceedances; high quantiles.

1   Introduction

The number of functionalities in the automotive, space, aerospace and railway domains involving computer systems is rapidly increasing. For instance, while in the past, car steering wheel was mechanically connected to the wheels, this connection is now electromechanical: a sensor detects the steering angle and this is transmitted to a computer, which processes the signal and uses electric motors to move the wheels. This whole process is subject to stringent safety-related timing constraints, since it affects the safety of the whole car.

Computing systems involved in critical functions, must provide evidence that the computation required to process new inputs executes correctly and at the right time. In the particular case of timing, the worst-case execution time (WCET) of the programs, that perform the required computations, needs to be estimated. From a probabilistic point of view this resembles a classic problem of extreme value theory (EVT), but from computer science terms it is not quite so.

Wilhelm *et al.* (2008), analyse ten commercial tools and research prototypes to estimate an upper-bound to the WCET of programs. They present two different classes of methods: static methods and measurement-based methods. The latter perform measurements from parts of a task in a particular hardware or simulator for a set of inputs. The former, static methods, use the executable code itself, analising the set of possible control flows to obtain upper bounds.

Static timing analysis has been the state-of-the-art practice, but requires a large and costly amount of detailed knowledge about the hardware and software components of the system. Probabilistic timing analysis has potential for reducing the weight of that demand. Further current research on measurement-based methods has focused on probabilistic tools, see also Bernat *et al.* (2002), Bernat and Newby (2006), and Hansen *et al.* (2009).

Cucu-Grosjean *et al.* (2012) present for the first time a measurement-based timing analysis technique based on EVT proven compatible with time-randomized platforms. This article is the closest probabilistic timing analysis variant to industrial practice and has been positively assessed in several industrial case studies, such as Wartel *et al.* (2013, 2015) and Kosmidis *et al.* (2014). Research on WCET and probabilistic WCET has been developed to date in the field of computer science. EVT has been added to this science along the lines of the fundamental theory, through such books as Beirlant *et al.* (2004), to cite just one classic and representative example. However, EVT has more tools that can enlighten the problems and this is what we aim to emphasize.

The collaboration between specialists in statistics and probability and engineers should be to the benefit of both parties. The execution time of programs running on a computing system has strong requirements regarding the guarantee that the longest execution time a program can take is bounded. The distribution of the execution time where we apply EVT should be light tailed, but data sets often seem heavy tailed. Hence, the application of EVT methods to WCET is a new and challenging problem.

In this paper we focus on the analysis of single-path automotive applications through the use of the EEMBC Autobench suite, see Poovey (2007). We consider a hardware architecture that implements such high-performance features as multicores and cache hierarchies analogous to that in Slijepcevic *et al.* (2014), and which has been modeled using a cycle-accurate simulator based on SoCLib (http://www.soclib.fr/trac/dev). The analised data sets are an example of the situation described in the preceding paragraph.

Cucu-Grosjean *et al.* (2012) also used several benchmarks from the EEMBC Autobench suite as a reference to reflect the current real-world demand of some

embedded systems. The EVT approach used in their analysis is the block maxima method. However, it is wasteful to only model block maxima for extreme value analysis if other data on extremes are available and basic assumptions can be assumed. Our EVT approach uses threshold exceedance models with generalised Pareto distribution (GPD), see Castillo and Serra (2015). We will follow a methodology, based on the coefficient of variation, developed by Castillo *et al.* (2014), and Castillo and Padilla (2016).

2   Exploratory data analysis

For this analysis we collect samples with N = 1,000 observations for each one of the 16 benchmarks in the EEMBC Autobench suite (Poovey, 2007), which is a well-known suite for real-time systems that includes a number of programs used in automotive embedded systems.

These datasets, which represent execution time in cycles, are, in alphabetic order: *a2time, aifftr, aifirf, aiifft, basefp, bitmnp, cacheb, canrdr, idctrn, iirflt, matrix, pntrch, puwmod, rspeed, tblook*, and *ttsprk*. Hereinafter, these sets will be called A (Autobench) plus their position in the list above. For example, A5 refers to *basefp*.

We start by drawing histograms of the data, and observe unimodal distributions with extreme values on the right tail in all cases. Figure 1 shows the histograms of the execution time for A2 and A11, where some extreme values can be observed for A2 that are far from the median, whereas a small new group of data can be observed on the far right for A11. Cucu-Grosjean *et al.* (2012) used A1, A3, A7, A8, A13, A14, A15, A16.

Box-plots have been made in order to show the existence of extreme values visually. Figure 2 shows the box-plots of six representative sets that represent the different behaviours observed in the tails, including the A2 and A11 sets shown above. The A5 dataset is fairly homogeneous with clustered outliers. The outliers in A2 and A9 are more dispersed than in the previous case, with two very extreme values. Outliers are dispersed homogeneously in A8, and finally we can observe two clusters of extreme values in A11 and A12. This difference in the type of tail behaviour can be achieved due to the combination of the characteristics of the program and those of the microarchitecture on which it is run.

Figure 1 Histograms of the execution time for A2 and A11.
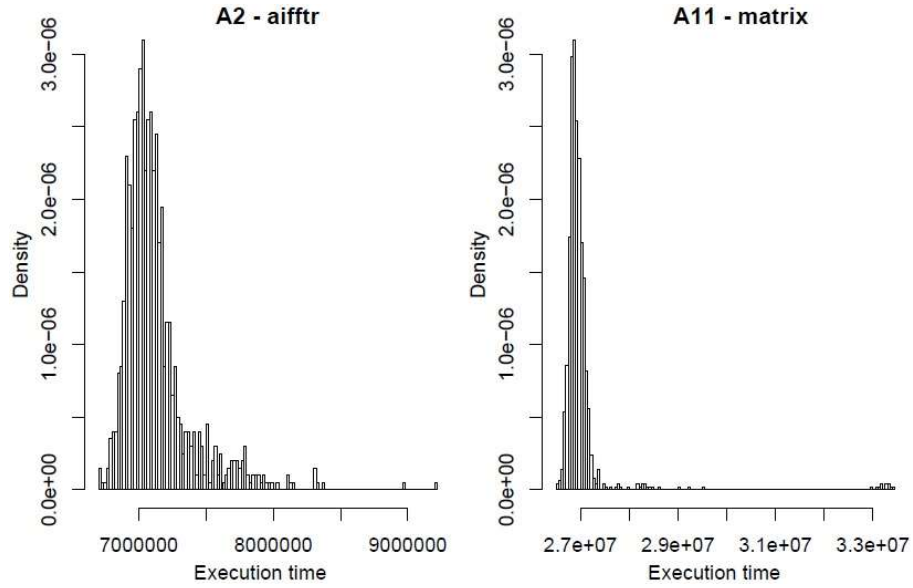
Figure 2 Box-plots of the execution time for six representative datasets.
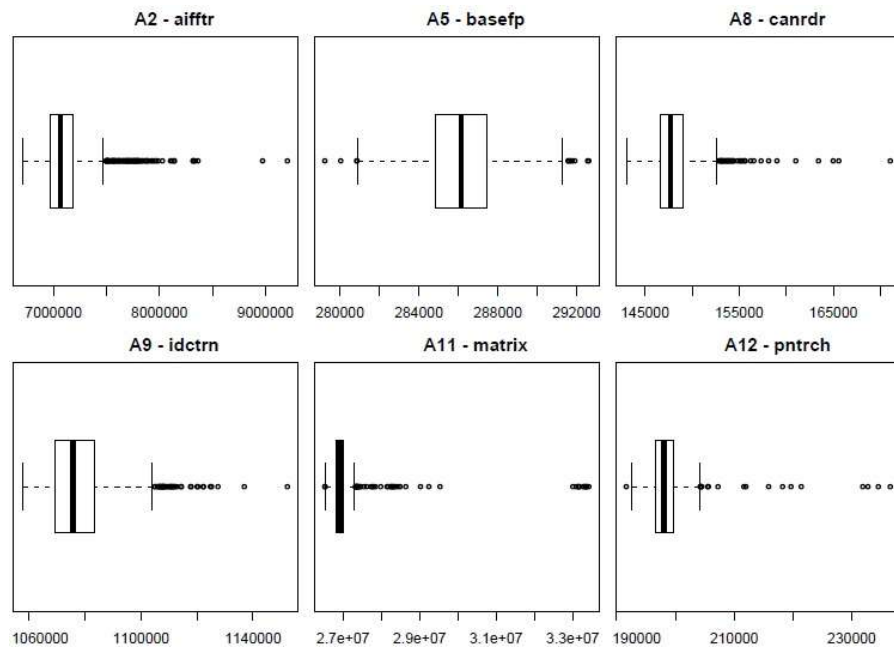
3 Detection of tail heaviness and dependence analysis

Before performing a more in-depth analysis of the data, it is necessary to check the number of finite moments and to test the independence of the data. With respect to the number of finite moments, this will be done with the group estimator proposed in Davydov *et al.* (2000) and with the ratio of the maximum to the sum described in the book by Markovich (2007, sections 1.2 and 1.3). In this section we use these methods as a rough test of tail heaviness. In the following section we will argue that the datasets probably have lighter tails, but now assume a most unfavorable situation to validate the techniques that need iid samples and fourth moment finite.

Let us consider an iid sample $X_1, \dots, X_n$ from a heavy-tailed distribution function $F(x)$ satisfying $1 - F(x) \approx x^{-\alpha}$ when $x \rightarrow \infty$. The sample is divided into l groups $V_1, \dots, V_l$, each group containing m random variables, that is, $n = l \cdot m$, where $l = [n/m]$ and $[a]$ denotes the integer part of a number $a > 0$.

Let

$$M_{ij}^{(1)} = \max\{X_j : X_j \in V_i\}$$

and let $M_{ij}^{(2)}$ deonte the second largest element in the sample group V_i. Let us denote

$$k_{li} = M_{ij}^{(2)}/M_{ij}^{(1)}, \quad z_l = (1/l) \sum_{i=1}^{l} k_{li}, \quad \alpha = z_l/(1 - z_l). \tag{1}$$

An estimator for the existence of moments is α that is called the group estimator. Table 1 shows finite fourth moment, with $m = 25$ and $l = 40$, for each dataset. We also tried with m from 20 to 100 with similar results.

Let $X_1, \dots, X_n$ be an iid sample. The ratio of the maximum to the sum is defined as

$$R_n(p) = M_n(p)/S_n(p), \quad n \geq 1, \quad p > 0, \tag{2}$$

where

$$S_n(p) = |X_1|^p + \dots + |X_n|^p, \quad M_n(p) = \max(|X_1|^p, \dots, |X_n|^p), \quad n \geq 1. \tag{3}$$

The maximum of the datasets ratios without A7 is 0.001, 0.002 and 0.004 for $p = 1$, $p = 2$, and $p = 4$, respectively, so there is evidence of the existence of a fourth moment. For A7, the ratios are 0.002, 0.005, and 0.019 for $p = 1$, $p = 2$, and $p = 4$, respectively, which suggests that the fourth moment might not be finite.

A popular and simple test of independence uses the first autocorrelation coefficient rejecting at the 5% significance level $\sqrt{n}|\hat{\rho}_1| > 1.96\sqrt{n}|\hat{\rho}_1| >$. The test applies the result $\sqrt{n}\rho_k \sim N(0, 1)$ for the autocorrelation coefficient at lag k, approximately, when H_0 is

true. When many coefficients are considered some will probably be significant even if H_0 is true. A more natural way to combine h coefficients into a single test statistic is the Ljung–Box test given by

$$LB = n(n+2) \sum_{k=1}^{h} \hat{\rho}_k^2 / (n-k)$$

where n is the sample size, and $\hat{\rho}_k$ is the sample autocorrelation at lag k. Under the hypothesis of independence, on the assumption that the data at least has a finite fourth moment, the statistic follows a χ_k^2 distribution.

The Ljung-Box has been applied to test independence for the 16 datasets of the Autobench suite. Table 1 shows the most significant p-value for each dataset, in the "p-val." column, using 10 and 20 lags. There is no evidence to reject independence in any case.

Although the estimates used guarantees finite fourth moment, the Box-Pierce statistic for random variables with infinite variance has been also applied to A7 (the less clear case), see Runde (1997). With 5 lags and $\alpha = 1.9$ the statistic is 1.51 and de critic value at 95% is 24.87. Then independence is not rejected.

4   Specifying the model

The purpose of this study is to determine at high probability the maximum value that can be achieved in a benchmark of the Autobench suite. The method that we will apply in Section 7, called "peak over threshold" (PoT), is based on determining a high enough threshold from which the distribution of the observations above this value, adjusted to zero, approaches a GPD distribution with cumulative distribution function.

$$G_{\xi,\psi}(x) = \begin{cases} 1 - (1 + \xi x/\psi)^{-1/\xi}, & \xi \neq 0, \\ 1 - \exp(-x/\psi), & \xi = 0, \end{cases} \qquad (4)$$

where $\xi \in \mathbb{R}$, and $\psi > 0$ are shape and scale parameters, respectively, $0 \leq x \leq -\frac{\psi}{\xi}$ if $\xi < 0$, and $x \geq 0$ if $\xi \geq 0$.

The value of the ξ parameter of the GPD used to model the threshold exceedances determines the tail type and is also called the extreme value index. If $\xi < 0$, we have a light tail, for $\xi = 0$ an exponential tail, if $0 < \xi < 1/4$ there is a heavy tail with four finite moments, and for $\xi > 1/4$ we will consider the tail too heavy. The mean of a GPD is $\psi/(1 - \xi)$ and the variance is $\psi^2/[(1 - \xi)^2(1 - 2\xi)]$ given $\xi < 1$ and $\xi < 1/2$, respectively. There is a limitation of the finite moments if $\xi > 0$, more precisely, a GPD has finite moments of order n if $\xi < 1/n$.

If it is obtained first the threshold from which the approach by the GPD is accurate, and the parameters of the GPD can easily be estimated, for instance by maximum likelihood. The problem is that there is generally a need to do both things at once. A standard procedure in EVT is to use the method proposed by Hill, or a modified Hill estimator based on second order parameters to avoid bias problems, see Gomes and

Pestana (2007). However, the Hill method always provides positive estimates of the extreme value index, because it is designed specifically for heavy tails. In the current situation we expect zero or negative values for the extreme value index, due to the software design and the static timing analysis. Cucu-Grosjean *et al.* (2012) always obtain Gumbel distributions for Autobench suite, using the block maxima method, hence, the extreme value index should be zero.

In this article we will follow the methodology based on the coefficient of variation (CV), defined for distributions with finite variance. The asymptotic theory of the empirical CV is developed by Castillo *et al.* (2014) and Castillo and Padilla (2016), assuming finite fourth moment. This theory can be used for exploratory data analysis, for testing the hypothesis of GDP, for estimating the extreme value index and for threshold selection. Let X be a continuous non-negative r.v. with distribution function $F(x)$. For any threshold, $t > 0$, the r.v. of the conditional distribution of threshold excesses $X - t$ given $X > t$, denoted $X_t = (X - t \mid X > t)$, is called the residual distribution of X over t. The cumulative distribution function of X_t, $F_t(x)$, is given by

$$1 - F_t(x) = \left(1 - F(x + t)\right) / (1 - F(t)). \qquad (5)$$

The quantity $M(t) = E(X_t)$ is called the residual mean and $V(t) = var(X_t)$ the residual variance. The residual coefficient of variation is given by

$$CV(t) \equiv CV(X_t) = \sqrt{V(t)}/M(t); \qquad (6)$$

like the usual CV, the function $CV(t)$ is independent of scale, and the function $t \rightarrow CV(t)$ is constant if and only if the distribution of X is GPD, see Castillo and Padilla (2016) and Castillo *et al.* (2014). The conditional distribution of threshold excesses of a GPD is again GPD and for any threshold, t > 0, the residual CV is given by

$$CV(t) = c_\xi = (1 - 2\xi)^{-1/2} \qquad (7)$$

The CV-plot has some advantages over ME-plot (Davison and Smith, 1990): first, it does not depend on the scale parameter; second, it is easier to detect constant functions than linear functions, since linear functions are defined by two parameters and the constants by only one. The uncertainty is essentially reduced from three to one single parameter.

5   Identifying tail behaviour

Given a sample $\{X_k\}$ of positive numbers, let $\{X_{(k)}\}$ be the ordered sample, so that $X_{(1)} \leq \cdots \leq X_{(n)}$. Defining a CV-plot as the representation of the empirical CV (cv) of the conditional exceedances, given by

$$k \rightarrow cv\left(X_{(j)} - X_{(k)} : j \geq k\right)$$

Given a sample of a GPD distribution with $\xi < 1/4$ and a threshold, $t > 0$, let $n(t)$ and $cv(t)$ the sample size and the CV of the residual sample. The asymptotic distribution of $cv(t)$ is given by

$$\sqrt{n(t)}\big(cv(t) - c_\xi\big) \to N\big(0, \sigma_\xi^2\big) \qquad (8)$$

where $\sigma_\xi^2 = 1$ and $c_\xi = 1$ for the exponential distribution ($\xi = 0$), and $\sigma_\xi^2 = 8/45$ and $c_\xi = 1/\sqrt{3}$ for the uniform distribution ($\xi = -1$). These values provide confidence intervals for the CV-plot of the mentioned distributions, see Castillo and Padilla (2016).

In order to fit the upper tail of each empirical cumulative distribution function of the 16 dataset, from the Autobench suite, by the GPD we examine at the decreasing part of the histogram, near to the upper half of the sample. On the other hand, we need large samples to increase the precision of the estimates. So we begin our analysis in this section with the $n = 500$ upper observations (tail size).

Figure 3 shows the *CV-plots* for some representative datasets. The black and grey dotted lines correspond to the confidence interval at 95% for the exponential and uniform distributions, respectively, and the constant line is CV=1.4, approximately CV for a GPD with $\xi = 1/4$, the limit for the four finite moments. It can be observed how the detected behaviours with the box-plots move to the *CV-plots*. The CV of A9 is clearly in the confidence interval of the exponential. For A2, it is exponential at the tail. For A5 looks like a constant light tail, for A8 it is heavy tail, with four finite moments and decrease in the confidence interval of the exponential distribution at the end. Finally, for A11 and A12, too heavy tails with a decrease to the exponential distribution at the end are observed. This latter behaviour is also observed for A3 and A7 in the Autobench suite.

While the basic model for probability distributions is the normal distribution, for the threshold exceedance it is the exponential distribution, the particular case of the GPD with shape parameter $\xi = 0$. It will be interesting, therefore, to find out whether this model is the most appropriate to fit the data. In this regard, a method to contrast exponential versus polynomial tails was developed by Castillo *et al.* (2014). The methodology based on the coefficient of variation has been extended to GPD, with finite fourth moment, by Castillo and Padilla (2016).

Using the building blocks given by (8) a multiple threshold test for a number m of thresholds as large as necessary for practical applications can be constructed. In order to avoid subjectivity as much as possible, to the limit of the number of thresholds m, we proceed as follows. Given de sample size n and the number of thresholds m, we choose a probability p, which determines the distance between the quantiles, such that $np^m = 8$. For $k = 0, \ldots, m$, let $p_k = 1 - p^k$ and $\{q_k\}$ the empirical quantiles coresponding to $\{p_k\}$.

Let us denote

$$T_m(\xi) = n \sum_{k=0}^{m} p^k \big(cv(q_k) - c_\xi\big)^2 \qquad (9)$$

This statistic allows testing if a sample is distributed as a GPD with parameter ξ. Moreover, if the parameter ξ is unknown, c_ξ can be estimated as the value such that achieves the minimum of $Tm(\xi)$ and reversing (7) provides and estimator of ξ.

Figure 3 CV-plots for the execution times for six representative datasets. The black and grey dotted lines correspond to the confidence interval at 95% for the exponential and uniform distributions, respectively, and the constant line is CV=1.4, approximately CV for a GPD with $\xi = 1/4$, the limit for the four finite moments.
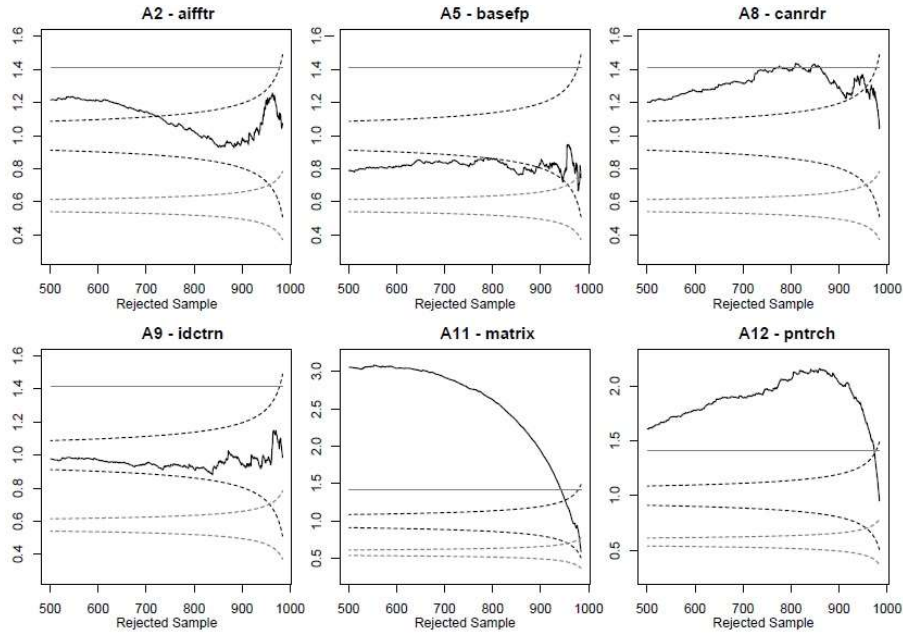
Applying $Tm(\xi)$ under exponentiality ($\xi = 0$) with $m = 20$ for all Autobench suite datasets ($n = 500$) the null hypothesis is not rejected in eight cases. The analysis is improved applying the contrast with $m = 20$ under the hypothesis of GPD, with ξ unknown, then GPD with light tail is not rejected for A1, A5, A6, A10, A13 and A15 and exponentiality is accepted for A4, A9, A14, A16, since ξ is close to 0 ($|\xi_i| \leq 0.04$). GPD with heavy tail is not rejected for A2 and A8. In short, for 12 datasets GPD is not rejected or, equivalently, is not rejected a constant residual CV. Table 1 shows the cases in which exponential or light tails are chosen for $n = 500$ values, see "Tail type" and "Tail size" columns.

Real-time systems only use a restricted form of programming, which guarantees that programs always finish; recursion is not allowed or explicitly bounded as are the iteration counts of loops. Considering this restriction, the appropriate models will be light tails, $\xi < 0$ for the GPD, or the exponential distribution, so it is advisable to add the boundary parameters in these cases, see Castillo and Serra (2015). If the basic model is the only option considered, we could be quite conservative in some cases when it really is not necessary.

For A2 and A8, a heavy tail is not rejected, but this restriction contradicts the bounded time for programs. The CV-plot for A2, in Figure 3, shows that for the second half of the sample the empirical CV is in the confidence interval of the exponential distribution. For A8 this can be accepted for the end part of the sample. This visual inspection can be automated with a threshold selection algorithm in the next section.

6   Threshold selection algorithm and extreme value index estimation

In the $T_m(\xi)$ calculation the number of thresholds m must be fixed by the researcher. This determines the thresholds where the CV is calculated, $\{0 = q_0 < q_1 < \cdots < q_m\}$, which are fixed throughout the procedure. We accept or reject the null hypothesis for the shape parameter using all the thresholds. If the hypothesis is rejected, the threshold excesses $\{x_j - q_1\}$ are calculated for the sub-sample $\{x_j \geq q_1\}$. The previous steps are repeated, but removing one threshold, to accept or reject the null hypothesis that the sample comes from a GPD with parameter ξ. This algorithm provides an estimation for the threshold as well as an estimation for the parameter ξ. If the parameter ψ is needed it can be estimated by maximum likelihood, see Castillo and Padilla (2016).

In 10 of the 16 datasets we obtained ≤ 0 ($n = 500$). With the *a priori* information that this is necessarily so from a threshold, we apply the algorithm above with $\xi = 0$ to datasets where GPD was rejected, and to A2 and A8, where GPD was not rejected but we obtained $\xi > 0$. Then, we not reject exponentiality ($\xi = 0$) for A2, A3, A7, A8, A11 and A12 with tail size 328, 40, 50, 40, 50 and 33, respectively, see Table 1. The tail size determines the standard error (se). Assuming $\xi = 0$, if $n = 500$ then $se = 0.04$, if $n = 20$ then $se = 0.22$.

Cucu-Grosjean *et al.* (2012) indicate that the Gumbel distribution fits well the distribution of maxima for 8 of the Autobench suite datasets. This means that the tails have to be exponential distributed in all cases, including A3, A7, A8, where we have achieved this only at the end of the distribution, with *a priori* information.

Applying the modified Hill estimator by Gomes and Pestana (2007) to Autobench suite, the parameter estimates show that the fourth moment is always infinite. For A3, A7, A11, even the variance is infinite. In order to summarize all views, from Section 3, the group estimator shows finite fourth moment for each dataset. This is confirmed qualitatively by the ratio of the maximum to the sum rough tests of tail heaviness.

Table 1 shows the parameter estimation, ξ column, and the number of values above the estimated threshold, "Tail size" column. A light tail (L) is chosen for six datasets, and exponential (E) for the rest, four cases with the full sample (n=500) and five at the tail of the distribution. The "Finite mom." column shows the parameter estimated by the group estimator.

7 PoT and VaR

The PoT method offers an approach to model the tail of a distribution based on the limit distribution from a threshold being a GPD. The first part consists of choosing a proper threshold, in the second part the data over this threshold is translated to the origin by $(X - t \mid x \geq t)$, and finally this new data is adjusted by a GPD, see Markovich (2007).

The estimated distribution function, $\widehat{F(x)}$, for values above the threshold t is

$$\widehat{F(x)} = \big(1 - F_n(t)\big)G_{\xi,t,\psi}(x) + F_n(t) \qquad (10)$$

where $F_n(t)$ is the empiric distribution function with $x = t$, and $G_{\xi,t,\psi}(x)$ is the distribution function of a GPD using an estimation of ξ, t, and ψ.

For threshold selection and the extreme value index estimation, t and ξ, we will use the threshold selection algorithm, and the maximum likelihood estimation for the scale parameter, ψ. Table 1 shows the t and ψ values, and ξ when the Tail type is L, if the Tail type is E then $\xi = 0$ is considered.

Some high quantiles for each sample are presented. In terms of risk, this concept is expressed as the value at risk (VaR). For a small p, $VaR_p = v$ if and only if $1 - F(v) = p$. Table 1 shows the different values for each proposed threshold and estimated parameters for the $VaR_{99.9\%}$. Figure 5 shows the fit of three examples to the data after which their use as predictors seems appropriate.

Finally, since we collect samples with $N = 1000$ observations, the $VaR_{99.9\%}$ should be about $x_{(999)}$, where $x_{(k)}$ is the kth order statistic of the sample. For the ten cases with tail size n = 500, and threshold $x_{(501)}$, the errors between the two quantities are small. For A4 and A16 are 2.17% and 1.06%, and less than 1% in the other eight cases. This is a validation of our methodology and allows extrapolate the VaR for much higher quantiles, although for small tail sizes the results may be less satisfactory.

Figure 4 Empirical CCDF (black line) and CCDF of the estimate GPD (dotted line) from quantile 95 for A2, A9, and A12.
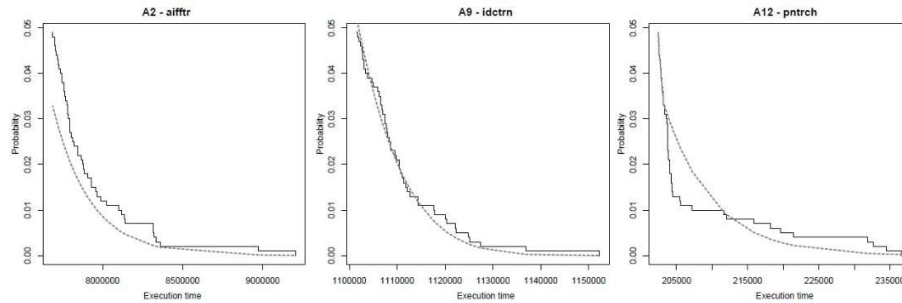
Table 1 Summary of the results of the 16 datasets. For each variable, the number of estimated finite moments by the group estimator, the most significant p-value for the Ljung-Box with 10 or 20 *lags*, the estimated ξ value with the threshold selection algorithm, the estimated ψ by MLE, the estimated threshold by the automatic algorithm, the tail size for the proposed threshold, the tail type (L for light tails and E for exponential tails), and $VaR_{99.9\%}$ for the PoT method are presented.

Var.	Name	Finite mom.	p-val.	ξ	ψ	Thresh.	Tail size	Tail type	VaR 99.9%
A1	a2time	44.8	0.336	-0.2	4.4×10^3	1.1×10^5	500	L	1.2×10^5
A2	aifftr	34.1	0.562	0	2.5×10^5	7.1×10^6	328	E	8.6×10^6
A3	aifirf	11.5	0.870	0	1.7×10^4	1.7×10^5	40	E	2.3×10^5
A4	aiifft	21.7	0.578	0	1.9×10^5	6.5×10^6	500	E	7.8×10^6
A5	basefp	302.8	0.380	-0.25	1.7×10^3	2.8×10^5	500	L	2.9×10^5
A6	bitmnp	61.1	0.117	-0.13	1.6×10^4	6.3×10^5	500	L	7.0×10^5
A7	cacheb	4.8	0.174	0	5.1×10^5	1.0×10^6	50	E	3.0×10^6
A8	canrdr	26.4	0.142	0	3.2×10^3	1.5×10^5	40	E	1.6×10^5
A9	idctrn	78.1	0.459	0	1.1×10^4	1.1×10^6	500	E	1.1×106
A10	iirflt	116.4	0.683	-0.15	2.3×10^3	1.7×10^5	500	L	1.8×10^5
A11	matrix	8.6	0.502	0	1.8×10^6	2.7×10^7	50	E	3.4×10^7
A12	pntrch	25.4	0.851	0	6.7×10^3	2.0×10^5	33	E	2.3×10^5
A13	puwmod	162.8	0.313	-0.17	2.2×10^3	2.2×10^5	500	L	2.3×10^5
A14	rspeed	43.7	0.700	0	1.1×10^3	6.5×10^4	500	E	7.3×10^4
A15	tblook	65.6	0.741	-0.11	1.8×10^3	8.0×10^4	500	L	8.8×10^4
A16	ttsprk	39.7	0.944	0	3.2×10^3	1.7×10^5	500	E	1.9×10^5

8   Concluding remarks

We have applied a new EVT methodology to the 16 datasets of the EEMBC Autobench suite described in this article, to predict their maximum execution time with certain confidence. The methodology is based on the following points:

1. Selecting an appropriate technique for the type of tail that we can find. For the EEMBC Autobench suite, methods based on the coefficient of variation, defined for distributions with finite variance, are appropriate. Otherwise, the methods that only work for heavy tails can give inappropriate estimates.

2. Deciding between the tail type (light, exponential, heavy, and too heavy) based on the graphics called CV plots and the exponentiality test T_m , for $\xi = 0$, using multiple thresholds.

3. Estimating the extreme value index, and the threshold at which the estimate is made, with a threshold selection algorithm based on contrasts of GPD for multiple thresholds.

4. Application of the PoT method, which divides the sample between the body and tail, from the previously selected threshold and the estimation of the extreme value index.

The methodology has worked successfully in most cases but presents doubts for the A7, A11 and A12 datasets. In these cases, other methods have been used that have not been conclusive. The proposed decision is to accept exponentiality for a very high threshold, which leads us to suspect a problem with contamination by a normal distribution in a lower proportion for an extremely high threshold.

Acknowledgements

References

Beirlant, J., Goeghebeur, Y., Teugels, J., and Segers, J. (2004) Statistics of Extremes: Theory and Applications. Wiley, Chichester.

Bernat, G., Colin, A., and Petters, S. M. (2002) 'WCET analysis of probabilistic hard real-time systems', Real-Time Systems Symposium, 2002. RTSS 2002. 23rd IEEE, pp. 279-288.

Bernat, G., and Newby, M. (2006) 'Probabilistic WCET analysis, an approach using copulas', Journal of Embedded Computing.

Castillo, J., Daoudi, J., and Lockhart, R. (2014) 'Methods to Distinguish Between Polynomial and Exponential Tails', IEEE Scandinavian Journal of Statistics, Vol. 41, No.2, pp.382–393.

Castillo, J., and Padilla, M. (2016) 'Modeling extreme values by the residual coefficient of variation', SORT Statist. Oper. Res. Trans. Vol. 40, num. 2.

Castillo, J., and Serra, I. (2015) 'Likelihood inference for Generalized Pareto Distribution', Computational Statistics and Data Analysis, Vol. 83, pp.116–128.

Cucu-Grosjean, L., Santinelli, L., Houston, M., Lo, C., Vardanega, T., Kosmidis, L., Abella,J., Mezzetti, E., Quinones, E., and Cazorla, F.J. (2012) 'Measurement-Based Probabilistic Timing Analysis for Multi-path Programs', Proceedings of the 2012 24th Euromicro Conference on Real-Time Systems, ECRTS '12, pp.91–101.

Davison, A. C., and Smith, R. L. (1990) 'Models for exceedances over high thresholds', J.Roy. Statist.Soc. Ser. B (Methodological), Vol. 52, pp.393–442.

Davydov, Y., Paulauskas, V. and Račkauskas, A. (2000) 'More on P-stable convex sets in Banach spaces', Journal of Theoretical Probability, Vol. 13, No.1, pp.39–64.

Gomes, M.I., and Pestana, D. (2007) 'A sturdy reduced-bias extreme quantile (VaR) estimator', Journal of the American Statistical Association, Vol. 102, No.477, pp. 280–292.

Hansen, J., Hissam, S. A., and Moreno, G. A. (2009) 'Statistical-based wcet estimation and validation', Proceedings of the 9th Intl. Workshop on Worst-Case Execution Time (WCET) Analysis.

Kosmidis, L., Quiñones, E., Abella, J., Farrall, G., Wartel, F., and Cazorla, F. J. (2014) 'Containing timing-related certification cost in automotive systems deploying complex hardware', 2014 51st ACM/EDAC/IEEE Design Automation Conference (DAC), pp.1-6.

Markovich, N. (2007) Nonparametric Analysis of Univariate Heavy-Tailed Data. Wiley Series in Probability and Statistics.

Poovey, J. (2007) 'Characterization of the EEMBC Benchmark Suite', North Carolina State University.

Runde, R. (1997). 'The asymptotic null distribution of the Box-Pierce Q-statistic for random variables with infinite variance an application to German stock returns', Journal of Econometrics, Vol. 78, No. 2, pp. 205-216.

Slijepcevic, M., Kosmidis, L., Abella, J., Quiñones, E., and Cazorla, F.J. (2014) 'Time-Analysable Non-Partitioned Shared Caches for Real-Time Multicore Systems', Proceedings of the 51st Annual Design Automation Conference, DAC '14, pp.1–6.

Wartel, F., Kosmidis, L., Lo, C., Triquet, B., Quinones, E., Abella, J., … and Vardanega, T. (2013). 'Measurement-based probabilistic timing analysis: Lessons from an integrated-modular avionics case study', 2013 8th IEEE International Symposium on Industrial Embedded Systems (SIES), pp. 241-248.

Wartel, F., Kosmidis, L., Gogonel, A., Baldovin, A., Stephenson, Z., Triquet, B., ... and Abella, J. (2015). 'Timing analysis of an avionics case study on complex hardware/software platforms', Proceedings of the 2015 Design, Automation & Test in Europe Conference & Exhibition, pp.397-402.

Wilhelm, R., Engblom, J., Ermedahl, A., Holsti, N., Thesing, S., Whalley, D., Bernat, G.,Ferdinand, C., Heckmann, R., Mitra, T., Mueller, F., Puaut, I., Puschner, P., Staschulat, J.,and Stenström, P. (2008) 'The Worst-case Execution-time Problem - Overview of Methods and Survey of Tools', ACM Trans. Embed. Comput.Syst., Vol. 7, No. 3, pp.36–53.

Websites

SoCLib. http://www.soclib.fr/trac/dev.