Approximating heavy tailed behaviour with Phase type distributions

András Horváth, Miklós Telek[1]

Abstract: In this paper two main problems are investigated. The first one is the effect of the goal function of the applied fitting method on the goodness of Phase type fitting. We discuss a numerical method based on a simple numerical optimization procedure that allows us to fit any non-negative distribution with a Phase type (PH) distribution according to any arbitrary distance measure. By comparing the fitting results obtained by minimizing different distance measures, conclusions are drawn regarding the role of the optimization criteria.

The second considered problem is the tail behaviour of Phase type distributions obtained via different fitting methods. To limit the numerical complexity of fitting methods (basically the evaluation of distance measures) the computations (numerical integration) are truncated at some point. Hence the information on the tail behaviour of the distribution is not considered beyond this point.

To approximate distributions with heavy tail we propose a complex method that uses different techniques to fit the main part and the tail of the distribution. The proposed method combines the advantages of fitting techniques and this way it overcomes some of their limitations.

The goodness of the discussed fitting methods are compared in queuing behaviour as well. The behaviour of the M/G/1 queue is compared with the one of the approximating M/PH/1 queue.

1 Introduction

The well known Phase type fitting methods (an overview is provided in [9]) can be classified based on their optimization criteria. Some of them intend

[1]Department of Telecommunications, Technical University of Budapest, Sztoczek u. 2, 1521 Budapest, Hungary, {*horvath,telek*}*@pyxis.hit.bme.hu*

to fit only some parameters of the distribution (usually some moments), while others intend to minimize a distance measure (EMPHT [2], MLAPH [3]). In the second case the optimization criteria, i.e., the distance measure to minimize, was always the same, the cross entropy, that comes from the applied maximum likelihood principle. In [1] a section is devoted to the argument about the general advantages of this distance measure. Though there are cases (e.g., better fitting the tail of distributions) when the weakness of the method minimizing the cross entropy measure becomes dominant and other methods can outperform it.

The need to compare the properties of different fitting approaches was recognized a decade ago, and a set of tests was defined during the workshop on Fitting Phase type distributions, Aalborg, Denmark, organized by S. Asmussen in February 1991. In [4] the proposed set of tests was evaluated using the MLAPH method and some new measures were proposed to be considered as well. In [9] a wider set of fitting methods was compared and their fitting measures were evaluated. Some of the consequences are quite natural. The methods that intend to minimize the distance between the original and the approximating distribution regarding a given aspect surpass other methods regarding that aspect (even if it is not always the case). For example a moment matching method may be superior to the MLAPH method regarding the relative errors in the moments while MLAPH may top it concerning cross entropy. Based on this observation a numerical procedure is implemented that minimizes an arbitrary distance measure which was not possible with the available PH fitting methods.

In recent telecommunication systems the occurrence of heavy tail distributions is reported, which directed the attention to the tail behaviour of PH distributions. The tail of any Phase type distribution is known to be exponential, while recent research results indicate the importance of distributions with "heavy" tails. When distance measures that are more sensitive to the tail distribution than the cross entropy are used as the optimization criteria better "tail fitting" can be achieved.

The PH fitting methods can be classified also by their generality. The methods that minimize a distance measure (EMPHT [2], MLAPH [3]) intend to find a global minimum of the goal function over the valid subset of the parameter space, hence we refer to them as *general fitting methods*. Another set of methods uses special PH structures and fits their parameters according to some heuristic considerations, hence we refer to them as *heuristic fitting methods*. Feldmann and Whitt proposed a simple but very effective heuristic fitting method that is especially applicable for fitting the tail behaviour of heavy tail distributions [8]. Their method uses mixtures of exponentials and hence results in distributions with decreasing density function. The main advantage of their method is the effective heuristic way of fitting. The application of general fitting methods is computationally expensive when the tail behaviour has to be approximated due to the numerical integration up to a high upper limit. The method proposed by Feldmann and Whitt provides good approximation of the tail behaviour with negligible computational effort. In this paper we provide a complex method that overcome the limitation of this method.

The goodness of the studied Phase type fitting methods are compared, on the one hand, through several plots and parameters of the distributions, and on the other hand by the effect of PH representation of general service time distributions in queuing systems. We compare the queue length distribution of the M/G/1 queue with the one of the approximating M/PH/1 queue. The queuing behaviour of the M/G/1 queue with heavy tail service time distribution is evaluated by the method proposed by Roughan et al. [11]. The analytical results given by this method were verified by simulation for queue length probabilities greater than 10^{-5} and showed a perfect fit.

It should be noted that the general distributions considered in this paper are continuous and are available in an analytical form. Fitting of empirical distributions based on their samples is not considered here.

The rest of the paper is organized as follows. The next section introduces fitting parameters and different distance measures. Section 3 describes the applied fitting method with some implementation details. The effect of goal function on the goodness of fit is discussed in Section 4. The succeeding section shows the effect of the fitting parameters on the M/G/1 queue length distribution. Section 6 presents the combined fitting method and Section 7 discusses its features. The last section gives the conclusion. Several numerical results are given in the Appendix.

2 Fitting parameters and distance measures

Participants of the Aalborg workshop proposed a set of parameters to measure the goodness of Phase type fitting methods. The original set of parameters was extended in [4] and the weakness of some measures proposed at the workshop was reported as well. Later on the following set of (non-negative) parameters was commonly used (e.g., in [9]):

1. Relative error in the 1st moment: $\hat{e}_1 = |\, c_1(\hat{F}) - c_1(F)\,| \,/\, c_1(F)$

2. Relative error in the 2nd moment: $\hat{e}_2 = |\, c_2(\hat{F}) - c_2(F)\,| \,/\, c_2(F)$

3. Relative error in the 3rd moment: $\hat{e}_3 = |\, c_3(\hat{F}) - c_3(F)\,| \,/\, c_3(F)$

4. Density absolute area difference: $\hat{D} = \int_0^\infty |\, \hat{f}(t) - f(t)\,|\, dt$

5. Relative entropy: $H_r = \int_0^\infty f(t)\; log\left(\frac{f(t)}{\hat{f}(t)}\right)\; dt$

$c_i(F)$ denotes the ith centered moment of F. Note that the relative entropy and minus the cross entropy ($-\;\hat{H} = \int_0^\infty \log \hat{f}(t)\, dF(t)$) can be interchangeably used in PH fitting, since they differ only in a constant H. That is the intrinsic entropy of the original distribution, $\hat{H} - H_r = H = \int f(t)log(f(t))dt$. The advantage of using the relative entropy measure is that it is always non-negative and its minimum is 0 (while the lower bound of minus the cross entropy is H, which can be a negative value as well). Authors of papers dealing with Phase type fitting algorithms usually reported minus the cross entropy. In order to make the comparison easier with those papers, throughout the appendix minus the cross entropy and the intrinsic entropy are given.

All of the above parameters can be used as a goal function that should be minimized for fitting, but in the first three cases $f(t)$ and $\hat{f}(t)$ can differ significantly even if the (non-negative) measures equal to 0. We refer to these parameters as *parameters of goodness*. In contrast, the parameters that equal to 0 if and only if $f(t) \equiv \hat{f}(t)$ are referred to as *distance measures*. In this paper we consider the following three distance measures as the goal function of Phase type fitting:

Relative entropy: $H_r = \hat{H} - H = \int_0^\infty f(t)\; log\left(\frac{f(t)}{\hat{f}(t)}\right)\; dt$

Area difference (L_1 distance): $\hat{D} = \int_0^\infty |f(t) - \hat{f}(t)|\; dt$

Relative area difference: $D_r = \int_0^\infty \frac{|f(t) - \hat{f}(t)|}{f(t)}\; dt$

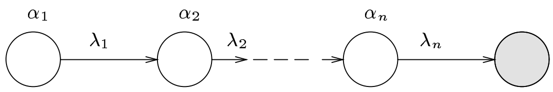


Figure 1: First Canonical Form (CF1)

The first two distance measures were already considered in [4]. The third distance measure was chosen to enlarge the effect of tail behaviour (at least compared to the area difference). In some cases the formula given for the relative area difference may be divergent, this problem will be relaxed in practice by defining a finite upper limit for the integral. Of course, there are several further reasonable distance measures that are not considered in this paper, e.g., the different parts of the distribution can be considered with different weights:

$$\sum_{j=0}^{n-1} a_j \int_{t_j}^{t_{j+1}} |f(t) - \hat{f}(t)| \, dt$$

where $t_0 = 0 < t_1 < \ldots < t_{n-1} < t_n \leq \infty$ provides a partitioning of the support set and $a_j > 0; j \in \{0, 1, \ldots, n-1\}$ are arbitrary weights. We consider only the above three distance measures because they are able to exhibit the effects that we would like to investigate.

3 A fitting method for arbitrary distance measure

The fitting problem may be formulated as an optimization problem the following way: find the parameters (the initial vector and the transition matrix) of the PH distribution such that the distance measure is minimal. Not having any restriction on the structure of the n stage PH distribution the number of free parameters is $n^2 + n$. In order to decrease the number of free parameters only Acyclic Phase (APH) distributions are considered. The procedure described in this section would be able to fit any Phase type structure by relaxing some constraints, but we believe that the flexibility of the APH class is practically equivalent to the flexibility of the whole PH class of the same order. Cumani [6] has shown that any acyclic APH distribution of order n may be transformed into the form represented in Figure 1 and referred to as First Canonical Form (CF1). The approximating PH distribution is in this form, described by the vectors $\underline{\alpha}$ and $\underline{\lambda}$.

The procedure starts from a random initial point of the parameter space that is the best of 200 random guesses, all of which has the proper mean

$\int tf(t)dt$. The best means the random guess with the least distance according to the applied measure. The distance measure is evaluated through numerical integration from 0 to T, where T is defined by $1 - F(T) = K$, where K is a small number that may have the value of $10^{-3}, 10^{-4}, 10^{-5}, \ldots$. The smaller K the higher the upper limit of the integration and the longer the approximation process.

Starting from the initial guess the non-linear optimization problem is solved by an iterative linearization method. In each step the following partial derivatives are numerically computed:

$$\frac{\partial D(f_{PH}(\underline{\alpha}, \underline{\lambda}, t), f(t))}{\partial \alpha_i}, \qquad \frac{\partial D(f_{PH}(\underline{\alpha}, \underline{\lambda}, t), f(t))}{\partial \lambda_i}, \qquad i = 1, \ldots, n,$$

where $D(f_{PH}(), f())$ stands for the distance between the pdf of the PH distribution ($f_{PH}()$) and the pdf of the original distribution ($f()$). Then, the simplex method is applied to determine the direction in which the distance measure decreases optimally. The constraints of the linear programming is given by probabilistic constraints (the initial probabilities have to sum up to one), by the restriction on the structure of the PH distribution (the λ_is are ordered [6]) and by confining the change of parameters (since the derivatives are valid only in a small area around $(\underline{\alpha}, \underline{\lambda})$). A search is performed in the direction indicated by the linear programming. The next point of the iteration is chosen to be the border of the linearized area (defined by the allowed maximum change in the parameters) in the optimal direction if the goal function is decreasing in that direction all the way to the border of the area. The next point is set to the (first) optimum if the goal function has an optimum in the optimal direction inside the linearized area. The iteration is stopped if the relative difference of the parameters in consecutive iteration steps are less than a predefined limit (10^{-5}), or if the number of iterations reaches the predefined limit (800). The allowed relative change of the parameters greater than 10^{-3} is less than Δ, where Δ starts from 0.1 and is multiplied by 0.995 in each step.

The necessary number of phases depends on the distribution to be fitted, and the required interval of fitting. After each fitting the measures of goodness may be examined and the probability density functions may be visually inspected. If the fitting is not satisfactory the number of phases may be increased. In general, increasing n results in better fitting but the higher n is the more time the estimation algorithm requires. Also, because of numerical problems, as the number of parameters is larger the goodness of fitting does not improve significantly (for instance, fitting the $[0; 1]$ uniform distribution by 32 phases do not show notable advances compared to the 24 phase fitting). Summing up, the necessary number of phases may be determined by

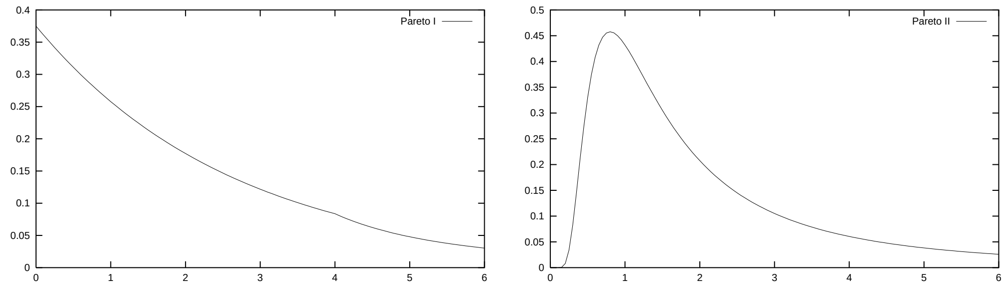


Figure 2: Pdf of Pareto I ($\alpha = 1.5$, $B = 4$) and II ($\alpha = 1.5$ and $b = 2$)

performing a series of fitting with different number of phases and choosing the most appropriate one taking into account the application, into which the PH distribution will be plugged in, as well.

Our numerical procedure is similar in some sense to the one proposed by Bobbio and Cumani [3], but our method is able to handle any goal function and it evaluates the derivatives via a simple numerical approximation, instead of the sophisticated calculation that is applicable only with the cross entropy measure.

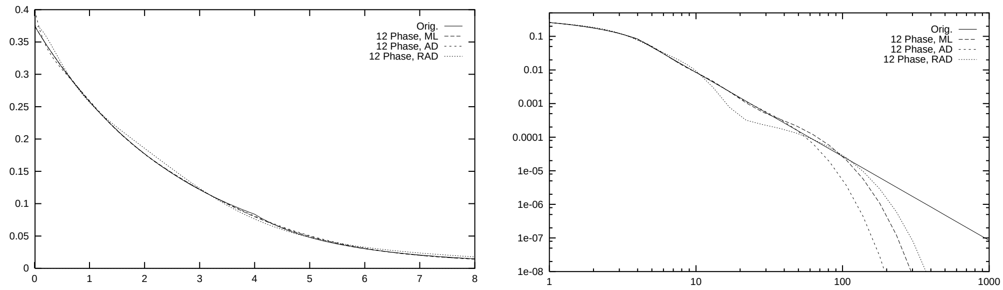
4 The effect of the goal function on the goodness of PH approximation

In this section we provide a representative set of Phase type fitting results obtained by our numerical method applying the mentioned goal functions. We have evaluated the complete benchmark as in [4, 9] with all distance measures, but here we provide only the results that we found meaningful. To investigate the goodness of fit heavy tail distributions we additionally consider the following Pareto-like distributions [11] (Figure 2):

$$\text{Pareto I:} \qquad f(t) = \begin{cases} \alpha B^{-1} e^{-\frac{\alpha}{B}\, t} & \text{for } t \leq B \\ \alpha B^{\alpha} e^{-\alpha} t^{-(\alpha+1)} & \text{for } t > B \end{cases}$$

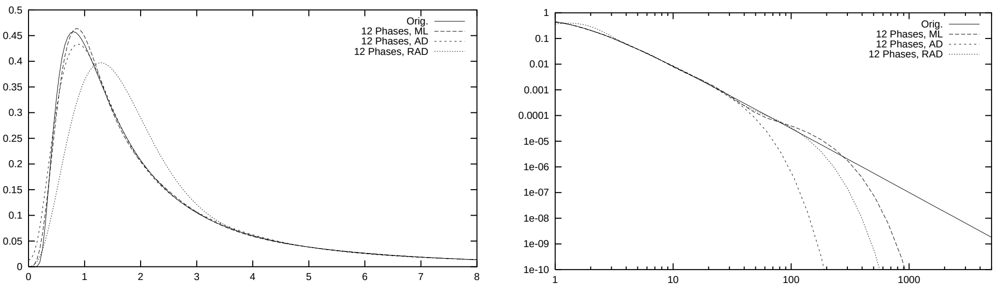
$$\text{Pareto II:} \qquad f(t) = \frac{b^{\alpha} e^{-b/t}}{\Gamma(\alpha)} x^{-(\alpha+1)}$$

There are significant differences between these distributions even if their tail behaviour is the same. Pareto I starts from a positive value and has monotone density while Pareto II starts from 0 (with 0 slope), hence it is not



Main part Tail behaviour

Figure 3: Pareto I distribution and its PH approximation



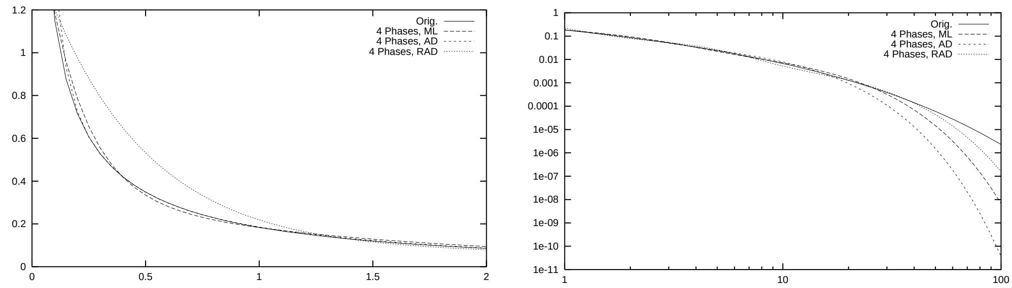
Main part Tail behaviour

Figure 4: Pareto II distribution and its PH approximation

monotone. The derivative of the density of Pareto I is not continuous, while it is continuous for Pareto II.

The results of the approximation of distributions with heavy tail are depicted in two parts. The main part of the distributions is shown in a linear – linear plot in the range of $\{0, \hat{t}\}$, where $\hat{t}$ is such that $F(\hat{t}) \sim 0.95$ and the tail of the distributions is shown in a logarithmic – logarithmic plot.

Figures 3 – 4 show the result of fitting with PH distributions of order 12 and with different distance measures. In the figures ML refers to fitting applying the relative entropy measure (that is related to the maximum likelihood phenomena), AD to the area difference measure and RAD to the relative area difference measure. The Pareto I distribution is used with $\alpha = 1.5$, $B = 4$ and Pareto II with $\alpha = 1.5$ and $b = 2$. In both cases the upper limit of the integration to evaluate the distance measure was defined by $1 - F(K) = 10^{-4}$, which results in $K = 683.0$ for Pareto I and $K = 767.0$ for Pareto II. As it can be observed in the figures for Pareto I the tail behavior is fitted better by using RAD as the distance measure while for Pareto II the approximation given by ML follows the tail further. Our general observation is that RAD fits the tail behaviour better for monotone distributions than ML does while



Main part Tail behaviour

Figure 5: PH approximations of distribution *W2*

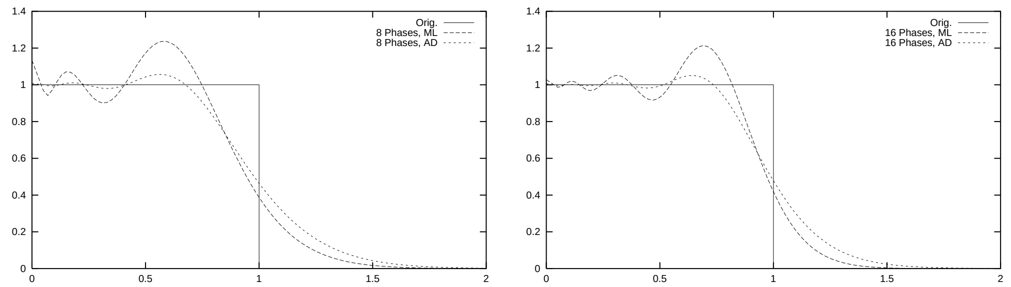


Figure 6: PH approximations of distribution *U1*

it may fail to give good approximation for non monotone distributions.

Figures 5 – 6 show other examples of the benchmark presented in [4] (see the appendix for the summary of the test cases used in the benchmark). *W2* is a long tail weibull distribution whose tail behaviour is followed better using RAD as the distance measure during the procedure. *U1* is a uniform distribution on [0 : 1]. Visual inspection gives the feeling that using AD results in better approximations than using ML. On the other hand the ML approximation gives significantly better results regarding the relative errors in the moments. The numerical parameters of the goodness of fit is provided in the Appendix.

Based on the fitting results of numerous different distributions applying different distance measures we draw the following conclusions:

1. Each distance measure has a “sensitivity structure”, meaning that they are not equally sensitive to the error of fitting at different “parts” of the distribution. The three considered measures can be classified as follows. The AD measure is sensitive to the main part of the distribution, the RAD measure to its tail (till the upper bound of the numerical integration), while the ML measure is sensitive to both, but it is less sensitive to the main part than the AD measure and in many cases less sensitive to the tail than the

RAD measure.

2. The "shape" of the distribution also has a significant role on the goodness of fit. Indeed the relationship between the shape of the distribution and the sensitivity structure of the applied distance measure affects the goodness of fit. Distributions with "non-Phase type" like behaviour in the main part can be better approximated using the AD measure. While distributions with "nice" behaviour at its main part and with "non-Phase type" behaviour in its tail can be better approximated using the RAD measure. The ML measure gives quite a robust method that works well in general without having a "strange" distribution to fit.

3. Of course, the goodness of fit is a general term. Parameters of fitting can be compared (next item), but the plots of the original and the approximating PH distributions provide an intuitive feel for the behaviour of fitting. The sensitivity structure of the applied distance measures can be recognized in the density plots as well. The fitting by the AD measure better approximate (when low order PH is used) the shape of the main part of the density in Figure 4 than the others do, while it is one whose tail "disappears" first in the case of heavy tail distributions. This trend of the tail behaviour was general in our experience. The tail of heavy tail distributions was best approximated by using RAD measure or ML measure and the worst tail fitting was achieved by using AD. (The relative error of the 3rd moment that is quite sensitive to the tail behaviour provides the same ranking.)

4. Usually, the best fit, according to a given fitting measure was reached by using that measure as the distance measure for the fitting, but there are several exceptions (the exceptions are highlighted by boldface characters in the Appendix). One potential reason of this phenomena is the numerical inaccuracy, but we think that the "shape" of the distribution plays role as well.

5. The fitting of distributions with low order Phase type (≤ 6) was usually terminated by reaching the required relative precision (i.e., the fitting method was not able to improve the approximation), while the fitting with higher order Phase type was terminated by reaching the maximum number of iterations.

5 The effect of PH approximation on the M/G/1 queue length distribution

One of the most important fields of application for Phase type distributions is in the area of traffic engineering of high speed communication systems. In this field the main question is not the goodness of fit of general service or interarrival time distributions, but the goodness of approximating the queuing behaviour of network elements with general service and/or interarrival time distributions.

In this section we compare the queuing behaviour of M/G/1 queues with the behaviour of their approximating M/PH/1 queue by considering the queue length distribution. The queue length distribution of the original M/G/1 queue with heavy tail service time distribution is evaluated using the method proposed by Roughan et al. [11].

The method of Roughan et al. [11] evaluates the queue length distribution by the following steps. First it calculates the asymptotic behaviour of the probability generating function (PGF) of the queue length distribution (given by the Pollaczek-Khintchine formula) via Tauberian theorems. Using the result it determines the asymptotic behaviour of the queue length distribution, then applies an Inverse Fast Fourier Transform (IFFT) on the PGF. It is shown by Daigle [7] that the result of the IFFT is contaminated by *alias* terms, but knowing the asymptotic behaviour of the queue length distribution they may be subtracted resulting in appropriate precision.

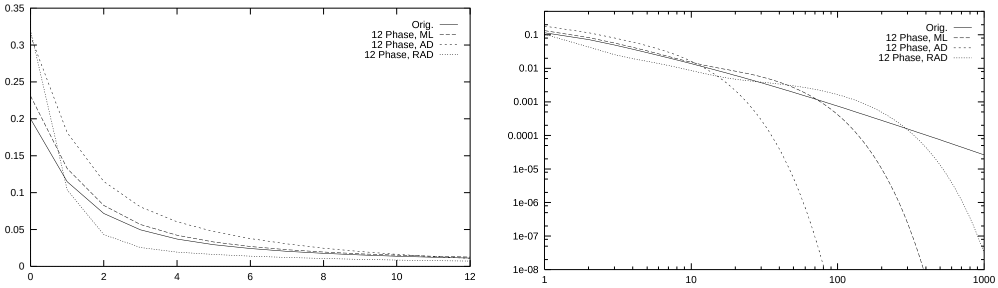
The method is applicable when the tail of the service time distribution has a power law tail which is the case for the two considered Pareto-like distributions.

Assuming that the Phase type distributed service time is given by the initial probability vector $\underline{a}$ (row vector) and transition matrix $\mathbf{B}$ the queue length distribution of the M/PH/1 queue can be evaluated using the matrix geometric method [10]:

$$p_0 = 1 - \lambda \; \underline{a} \; \mathbf{B}^{-1} \; \underline{e}; \quad p_i = p_0 \; \underline{a} \; \mathbf{R}^i \; \underline{e} \;\; \forall i \geq 1 \; .$$

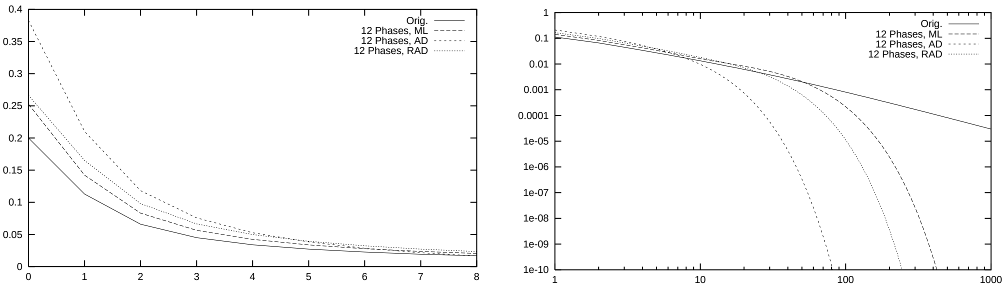
where λ is the arrival rate, $\underline{e}$ is a column vector of ones, $\underline{a} \; \mathbf{B}^{-1} \; \underline{e}$ is the mean of the Phase type distributed service time, and matrix $\mathbf{R}$ is defined as $\mathbf{R} = (\mathbf{I} - \underline{e} \; \underline{a} - 1/\lambda \; \mathbf{B})^{-1}$.

Figures 7 – 8 show the queue length distribution of an M/G/1 queue with Pareto I ($\alpha = 1.5$, $B = 4$) and Pareto II ($\alpha = 1.5$, $b = 2$) service time distribution and the approximating M/PH/1 queue, where the service time is PH distributed with order 12 and is obtained by minimizing different distance measures (ML, AD, RAD). The continuous curves in Figures 7 –



Main part Tail behaviour

Figure 7: Queue length distribution of an M/G/1 queue (with Pareto I) and its approximate M/PH/1 queue



Main part Tail behaviour

Figure 8: Queue length distribution of an M/G/1 queue (with Pareto II) and its approximate M/PH/1 queue

8 are obtained by joining the queue length probability values evaluated at integer points. Traffic intensity equals to 0.8.

6 A combined fitting method

As we have seen in the previous sections we can improve the tail fitting of *general fitting methods* by applying distance measures that are sensitive to the tail behaviour. But it is also mentioned that the applicability of this approach is limited by its computational complexity, which increases at least linearly with the considered subset of the support (the upper limit of the numerical integration)[2]. There is a trade off between the *general* and *heuristic* fitting methods. Generally, the computational complexity of the

[2]Numerical integration techniques with exponentially increasing step size is a way to avoid the linear increase of the computational complexity, but the complexity problem remains anyway.

general fitting methods is much higher than the complexity of *heuristic fitting methods*, but the general methods are much more flexible, i.e., they better approximate a wide range of distributions. *Heuristic fitting method*s that usually use special subclasses of the class of Phase type distributions are less flexible. They provide much poorer fitting for a wide range of distributions, but there might be a set of distributions that can be approximated by a *heuristic fitting method* as well as by using any *general fitting method.* When only this special set of distributions needs to be fitted it is worth applying the *heuristic fitting method.*

In practice, the main part of empirical distributions can have any general structure, while the tail of empirical distributions is assumed to be "nice" so that *heuristic fitting methods* can be used for tail fitting. Ugly tail behaviour, like the tail of Matrix Exponential distribution $f(t) = (1 + 1/(2\,\pi)^2\,)\,(1 - \cos(2\,\pi\,t)\,)\,e^{-t}$ [5, 4], or similar non-monotone functions with non-exponential decays, are not commonly used in practice.

Based on these considerations we propose one fitting method that uses a general approach to approximate the main part and a heuristic approach to approximate the tail of distributions. The heuristic method used for fitting tail behaviour is based on the method proposed by Feldmann and Whitt [8] and the general method to fit the main part is based on the numerical procedure introduced in the previous sections. Indeed, only a slight modification is needed to combine the two methods into a combined procedure.

The limitation of our combined method comes from the limitation of the heuristic method of Feldmann and Whitt. Their method is applicable only for fitting distributions with monotone decreasing density function. Hence the proposed combined method is applicable when the tail of the distribution is with monotone decreasing density. This restriction is quite loose since the border of the main part and the tail of the distribution is arbitrary, hence the restriction of applicability is to have a positive number C such that the density of the distribution is monotone decreasing above C.

The result of our fitting algorithm is a Phase type distribution of order $n + m$, where n is the number of phases used for fitting the main part and m is the number of phases used for fitting the tail. The structure of this Phase type distribution is depicted in Figure 9. The parameters $\beta_1, \ldots, \beta_m, \mu_1, \ldots, \mu_m$ are computed by considering the tail while the parameters $\alpha_1, \ldots, \alpha_m, \lambda_1, \ldots, \lambda_2$ are determined considering the main part of the distribution. The algorithm consists of the following steps.

First, we define the border of the main part and the tail, t_c, based on constant c by the equality $1 - F(t_c) = c$; c can depend on the distribution

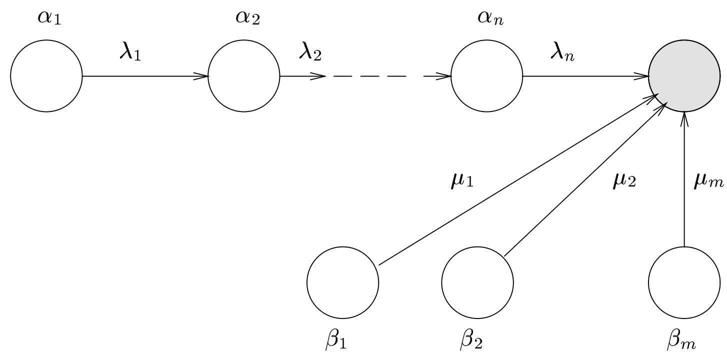


Figure 9: Structure of approximate Phase type distribution

and other consideration (e.g., the computational complexity[3]). Its typical value is between $0.0001 - 0.01$.

The upper bound of the tail approximation t_d is determined by an other constant d in a similar way by the equality $1 - F(t_d) = d$; d can vary in a wide range, e.g., $10^{-20} - 10^{-4}$, but it can be smaller than 10^{-20} as well. It does not affect the computational complexity, d is rather limited by the applied floating point arithmetic.

The method proposed by Feldmann and Whitt is a recursive fitting procedure that results in a hyperexponential distribution whose cumulative distribution function (cdf) at a given set of points is “very close” to the cdf of the original distribution. We use a slightly modified version of this algorithm to determine the parameters $\beta_1, \ldots, \beta_m, \mu_1, \ldots, \mu_m$.

Based on this limit we define $2m$ points ($0 < t_c = t_m < bt_m < t_{m-1} < bt_{m-1} < \ldots < t_1 = t_d < bt_1$) at which the approximate distribution is “close” to the original one.

$$t_i = t_d \delta^{-i+1}, \quad i \in \{1, 2, \ldots, m\}, \quad \text{where} \ \ \delta = \sqrt[m-1]{\frac{t_d}{t_c}} \ , \ \text{and} \ b < \delta.$$

We choose μ_1, β_1 to match the complementary cdf $F^c(t)$ at the arguments t_1 and bt_1. Arranging the two equations

$$\beta_1 e^{-\mu_1 t_1} = F^c(t_1), \quad \text{and} \quad \beta_1 e^{-\mu_1 b\ t_1} = F^c(b\ t_1),$$

we obtain

$$\mu_1 = \frac{1}{(b-1)\ t_1} \ln(F^c(t_1)/F^c(b\ t_1)), \quad \text{and} \quad \beta_1 = F^c(t_1)\ e^{\mu_1 t_1}.$$

[3]The complexity of the general method dominates the complexity of the combined method. The larger c the lower the border t_c, and the lower the computational complexity.

Throughout the procedure, we are assuming that $\mu_i, i = 2, \dots, m$ will be significantly larger than μ_1, so that

$$\sum_{i=1}^{m} \beta_i e^{-\mu_i t} \sim \beta_1 \ e^{-\mu_1 t}, \qquad \text{for } t \geq t_1. \tag{1}$$

As it is noted in [8] there is no guarantee that the above property holds, but it may be checked after the procedure is complete, and in general it is not a problem to define the set of points in such way that we have this property.

Using the notation

$$F_i^c(t) = F_{i-1}^c(t) - \sum_{j=1}^{i-1} \beta_j \ e^{-\mu_j t},$$

where $F_1^c(t) = F^c(t)$, our goal is to have

$$\beta_i \ e^{-\mu_i t_i} = F_i^c(t_i), \qquad \beta_i \ e^{-\mu_i b t_i} = F_i^c(b \ t_i),$$

rearranging we obtain

$$\mu_i = \frac{1}{(b-1) \ t_i} \ln(F_i^c(t_i)/F_i^c(b \ t_i)), \qquad \beta_i = F_i^c(t_i) \ e^{\mu_i t_i}, \tag{2}$$

for $2 \leq i \leq m$.

We have no guarantee that the sum of the initial probabilities associated with the hyperexponential part of the Phase type structure ($\sum \beta_i$) is lower than 1. If the sum is greater than 1 it may help to decrease d (which means to increase t_d) or m. It is discussed in [8] how the $2m$ points may be chosen efficiently.

Having $\beta_1, \dots, \beta_m, \mu_1, \dots, \mu_m$ we use the algorithm described in Section 3 to fit the main part of the distribution with two differences:

1. Not having the hyperexponential part fitting the tail, the initial probabilities of the acyclic structure sums up to 1. Having the hyperexponential part this constraint has to be modified as $\sum_{i=1}^{n} \alpha_i = 1 - \sum_{i=1}^{n} \beta_i$.

2. The structure of the approximate Phase type distribution differs from the one used before (Figure 1). The parameters associated with the additional m phases ($\beta_1, \dots, \beta_m, \ \mu_1, \dots, \mu_m$) are fixed during this stage of the fitting process.

The upper limit for the integral to evaluate the distance measure during fitting the main part is t_c.

Figure 10 pictures how different parts of the PH structure (CF1,hyper-exponential) contributes to the pdf of the ($\alpha = 1.5, \ B = 4$). The fitting procedure was run to approximate the Pareto I distribution with $n = 8, m = 10, c = 10^{-2}, d = 10^{-10}, t_c = 31.70, t_d = 6.83 \cdot 10^6$.

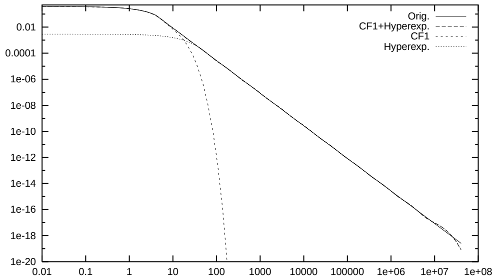


Figure 10: Different parts of the distribution are approximated by different parts of the PH structure

7 Goodness of the combined fitting method

The goodness of the fitting of the main part using the general method presented in Section 3 is not affected by the use of the heuristic procedure. Since the heuristic approach gives a good approximation for the tail, it is not rewarding to apply the relative area difference for fitting the main part. With a few exceptions using the relative entropy as the distance measure is the most promising choice.

The method presented by Feldman and Whitt results in a distribution whose pdf is oscillating around the pdf of the original distribution. The combined procedure has the same feature. The oscillation starts at t_c and ends at t_d. The number of "bumps" equals to m. After t_d the pdf of the PH distribution does not follow the pdf of the original distribution. The "amplitude" of the oscillation depends on the distance between t_c and t_d and on the number of phases used for fitting the tail (m). Increasing m decreases the amplitude of the oscillation. There is an upper limit for m as a result of assumption (1) and probabilistic considerations (the parameters given by (2) have to be proper for a PH distribution). The method of Feldman and Whitt is worth applying when the tail is "heavy" enough, otherwise applying the *general* method alone gives as good results as the combined one. As it is mentioned in [8] the method works well for distributions with decreasing hazard rate.

The pdf of the PH distribution drops slightly compared to the original one at t_c. If it is necessary this drop may be decreased by using $t_c^* > t_c$ as the upper limit for the integral to evaluate the distance measure. This feature is illustrated in Figure 11 with the Pareto I distribution and the following parameters: $n = 8, m = 4, c = 10^{-2}, d = 10^{-10}, t_c = 34.83, t_d = 7.67 \cdot 10^6, t_c^* = 164.6, F^c(t_c^*) = 10^{-3}$. Overlapping refers to the case when $t_c^* > t_c$.

The choice for the distance measure used during fitting the main part has

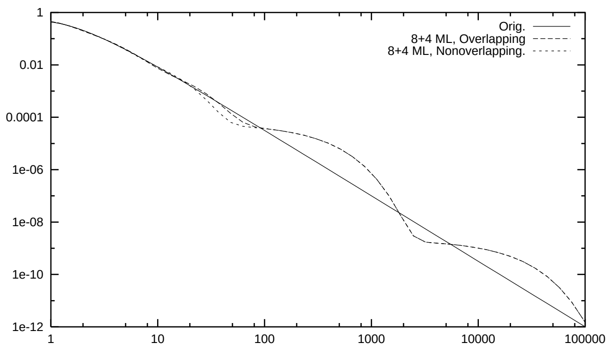


Figure 11: The effect of using $t_c^* > t_c$ as the upper limit for the integral to evaluate the distance measure

no effect on the procedure for fitting the tail.

The improvement achieved by fitting the tail using the *heuristic method* is indicated not only by visual inspection of the tail of the pdf but by the relative moment errors as well. Some examples are given in the Appendix. Since the 2nd and 3rd moments do not exist for the two Pareto-like distributions, for these cases the following is given to indicate the improvement instead of the original relative moment errors:

$$|\, c_2^*(\hat{F}) \;-\; c_2^*(F)\,|\,/\,c_2^*(F), \quad |\, c_3^*(\hat{F}) \;-\; c_3^*(F)\,|\,/\,c_3^*(F), \tag{3}$$

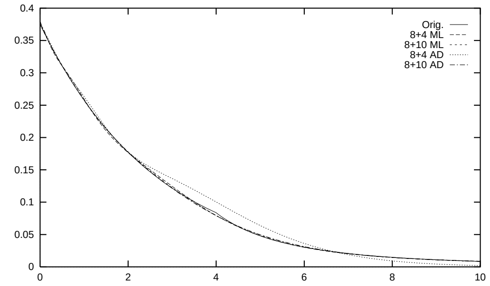
where

$$c_i^*(F) = \int_{x=0}^{C} (x - c_1(F))^i dF(x), \quad i = 2, 3,$$

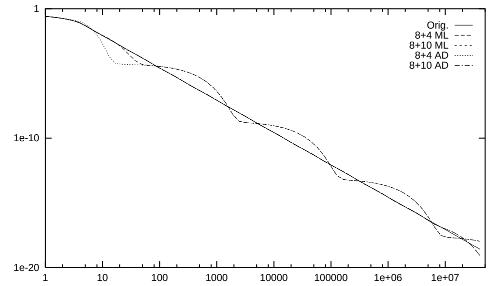
with C defined by $F^c(C) = 10^{-8}$.

Figures 12 (14) compares the behaviour of the Pareto I (Pareto II) distribution and its fitting Phase type distribution. The notation `n+m XX` defines the parameters of Phase type fitting. `n` is the number of phases used for fitting the main part and `m` is the number of phases used for fitting the tail of the distribution. `XX` (= AD or ML or RAD) gives the distance measure used for fitting the main part. The relative error of the pdf is defined as $(f(t) - \hat{f}(t))/f(t)$, and the hazard rate is $f(t)/(1 - F(t))$.

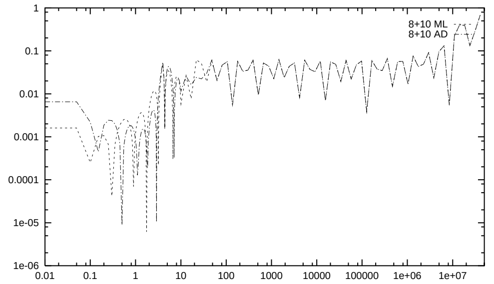
Figures 13 (15) shows the effect of Phase type fitting on the M/G/1 queue behaviour with Pareto I (Pareto II) service time. The oscillation that appeared on the tail of the pdf of the PH distribution can be seen on the queue length distribution as well.



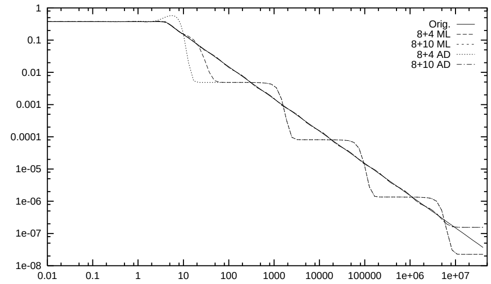
Pdf of the main part



Pdf of the tail

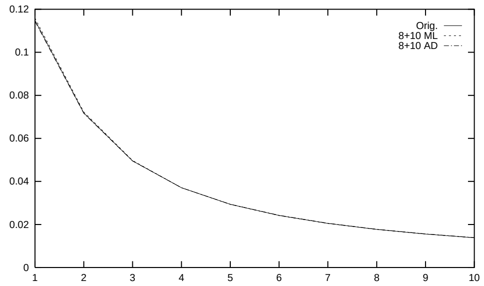


Relative error of the pdf

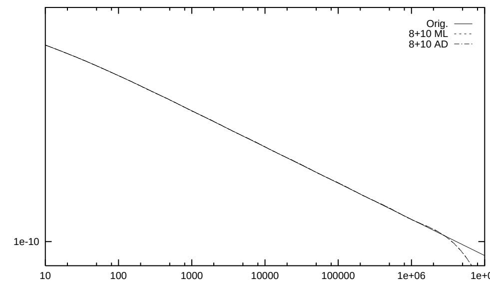


Hazard rate function

Figure 12: Pareto I distribution and its PH approximation with the combined method



Main part



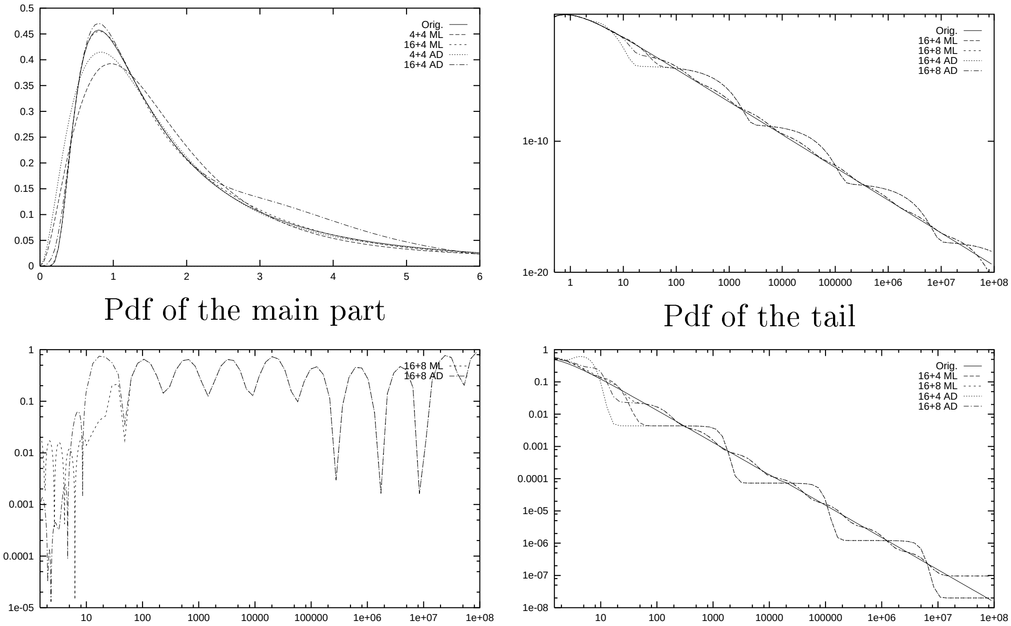
Tail behaviour

Figure 13: Queue length distribution of an M/G/1 queue (with Pareto I) and its approximate M/PH/1 queue

8 Conclusion

This paper investigates Phase type fitting techniques that are able to improve the tail fitting behaviour of the existing methods.

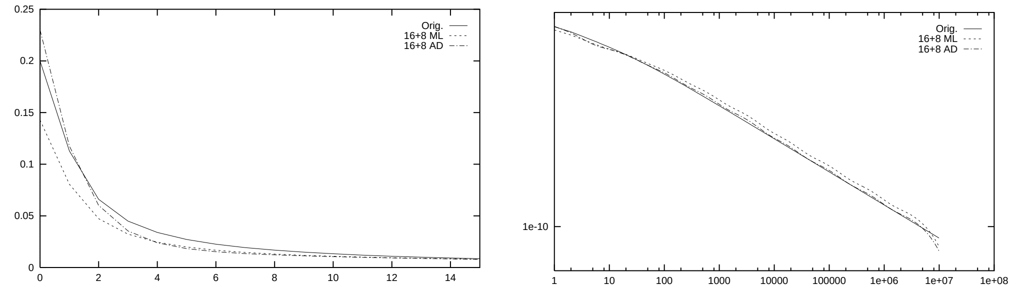
First a Phase type fitting method is presented that is able to approximate distributions based on any general distance measure. It has been shown that the properties of Phase type fitting can be tuned by choosing an appropriate distance measure for the fitting method.



Pdf of the main part | Pdf of the tail

Relative error of the pdf | Hazard rate function

Figure 14: Pareto II distribution and its PH approximation with the combined method



Main part | Tail behaviour

Figure 15: Queue length distribution of an M/G/1 queue (with Pareto II) and its approximate M/PH/1 queue

To further improve the tail fitting properties a combined Phase type fitting method is introduced. This method implements the above general method for fitting the main part of a distribution and the effective heuristic approach of Feldman and Whitt for fitting the tail behaviour.

Several numerical examples present the properties of the introduced fitting methods. The examples show that the proposed combined fitting method provides a suitable Phase type approximation of heavy tailed distributions that is also verified by the queuing behaviour of the M/G/1 queues and their approximating M/PH/1 queue.

Acknowledgements

We thank Matthew Roughan for submitting us his Matlab code for M/G/1 queuing analysis with heavy tail distributions. This work has been supported by the Hungarian OTKA under Grant No. T-30685 and F-23971.

References

[1] S. Asmussen. Phase-type distributions and related point processes: Fitting and recent advances. In S. R. Chakravarty and A. S. Alfa, editors, *Matrix-analytic methods in stochastic models*, Lecture notes in pure and applied mathematics, pages 137–149. Marcel Dekker, Inc., 1996.

[2] S. Asmussen and O. Nerman. Fitting Phase-type distributions via the EM algorithm. In *Proceedings: "Symposium i Advent Statistik"*, pages 335–346, Copenhagen, 1991.

[3] A. Bobbio and A. Cumani. ML estimation of the parameters of a PH distribution in triangular canonical form. In G. Balbo and G. Serazzi, editors, *Computer Performance Evaluation*, pages 33–46. Elsevier Science Publishers, 1992.

[4] A. Bobbio and M. Telek. A benchmark for PH estimation algorithms: results for Acyclic-PH. *Stochastic Models*, 10:661–677, 1994.

[5] D.R. Cox. A use of complex probabilities in the theory of stochastic processes. *Proceedings of the Cambridge Phylosophical Society*, 51:313–319, 1955.

[6] A. Cumani. On the canonical representation of homogeneous Markov processes modelling failure-time distributions. *Microelectronics and Reliability*, 22:583–602, 1982.

[7] John N. Daigle. Queue length distributions from probability generating functions via discrete fourier transforms. *Operations Research Letters*, 8:229–236, 1989.

[8] A. Feldman and W. Whitt. Fitting mixtures of exponentials to long-tail distributions to analyze network performance models. *Performance Evaluation*, 31:245–279, 1998.

[9] A. Lang and J. L. Arthur. Parameter approximation for phase-type distributions. In S. R. Chakravarty and A. S. Alfa, editors, *Matrix-analytic methods in stochastic models*, Lecture notes in pure and applied mathematics, pages 151–206. Marcel Dekker, Inc., 1996.

[10] M.F. Neuts. *Matrix Geometric Solutions in Stochastic Models.* Johns Hopkins University Press, Baltimore, 1981.

[11] M. Roughan, D. Veitch, and M. Rumsewicz. Numerical inversion of probability generating functions of power-law tail queues. Tech. report, 1997.

A Parameters of fitting

Test cases of the benchmark presented in [4] extended with heavy tail distributions

Density	Symbol	Numerical Cases	
Weibull			
$f(t) = \frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1} e^{-(\frac{t}{\eta})^{\beta}}$	*W1*	$\eta = 1$	$\beta = 1.5$
	W2	$\eta = 1$	$\beta = 0.5$
Lognormal			
$f(t) = \frac{1}{\sigma t \sqrt{2\pi}} \exp\left[-\frac{(\log(t/\phi)+\sigma^2/2)^2}{2\sigma^2}\right]$	*L1*	$\phi = 1$	$\sigma = 1.8$
	L2	$\phi = 1$	$\sigma = 0.8$
	L3	$\phi = 1$	$\sigma = 0.2$
Uniform on (a, b)	*U1*	$a = 0$	$b = 1$
	U2	$a = 1$	$b = 2$
Shifted Exponential			
$f(t) = \frac{1}{2} e^{-t} + \frac{1}{2} e^{-(t-1)} I(t \geq 1)$	*SE*		
Matrix Exponential			
$f(t) = \left(1 + \frac{1}{(2\pi)^2}\right)(1 - \cos(2\pi t))\, e^{-t}$	*ME*		
Pareto I			
$f(t) = \begin{cases} \alpha B^{-1} e^{-\frac{\alpha}{B} t} & \text{for } t \leq B \\ \alpha B^{\alpha} e^{-\alpha} t^{-(\alpha+1)} & \text{for } t > B \end{cases}$	*P1*	$\alpha = 1.5$	$B = 4$
Pareto II			
$f(t) = \frac{b^{\alpha} e^{-b/t}}{\Gamma(\alpha)} x^{-(\alpha+1)}$	*P2*	$\alpha = 1.5$	$b = 2$

The fitting parameters are given in the following tables. The column labeled "Appr." describes the fitting procedure: the symbol of the fitted distribution, the order of the PH distribution and the applied distance measure are given. In the case when there is one number in this column the fitting method given in Section 3 was applied. If there are two numbers of the form $x + y$, x is the number of phases to fit the main part and y is the number phases to fit the tail of the distribution. The relative errors in the 2nd and 3rd moments for the Pareto-like distributions (*P1*,*P2*) are given as defined in (3). For *U1* and *U2* the absolute error in the 3rd moment is given since their 3rd centered moments equal to 0.

Appr.	Cr. Ent. ($\hat{H}$)	Area D.	RME 1	RME 2	RME 3
W1/4	$H = 7.869e-01$				
ML	7.8714e-01	8.9011e-03	1.3734e-04	4.8522e-03	4.0876e-02
AD	7.8728e-01	7.2039e-03	2.8943e-03	2.4579e-02	1.0107e-01
W1/8					
ML	7.8702e-01	3.5172e-03	9.2808e-05	9.0348e-04	9.7635e-03
AD	7.8725e-01	3.1405e-03	2.0490e-03	1.6044e-03	1.3276e-02
W1/16					
ML	7.8696e-01	**1.8287e-03**	1.5028e-05	7.5074e-04	4.8582e-03
AD	7.8773e-01	5.3185e-03	8.2319e-04	5.6408e-04	5.5515e-04
W2/4	$H = 1.1546e+00$				
ML	1.1631e+00	**8.2830e-02**	1.2918e-02	2.2862e-01	5.7052e-01
AD	1.1933e+00	1.1162e-01	8.9610e-02	4.5281e-01	7.8482e-01
RAD	1.2945e+00	3.0036e-01	2.1358e-02	1.6789e-01	3.7917e-01
W2/8					
ML	1.1626e+00	**6.0275e-02**	9.8517e-03	1.9015e-01	4.9798e-01
AD	1.1828e+00	9.4857e-02	4.8308e-02	2.0663e-01	6.0035e-01
RAD	1.2078e+00	1.4406e-01	7.5972e-04	4.4718e-02	1.6866e-01
W2/8+3					
ML	1.1544e+00	5.3229e-02	1.0493e-02	1.0160e-01	2.7704e-01
W2/16					
ML	1.1643e+00	**5.9088e-02**	1.3257e-02	1.8740e-01	4.8400e-01
AD	1.1827e+00	8.9019e-02	2.7044e-01	8.0677e-01	9.9291e-01
RAD	1.1908e+00	1.0993e-01	1.4096e-02	1.1377e-01	3.2936e-01
L1/4	$H = 3.745e-01$				
ML	4.0322e-01	**4.1789e-02**	1.6090e-01	7.6213e-01	9.7536e-01
AD	4.2514e-01	4.2115e-02	2.8016e-01	8.7027e-01	9.9189e-01
RAD	4.7847e-01	3.3306e-01	2.1285e-02	6.6316e-01	9.4266e-01
L1/8					
ML	4.0025e-01	3.3545e-02	1.6095e-01	7.5461e-01	9.7293e-01
AD	**3.9838e-01**	2.3236e-02	1.6794e-01	7.4279e-01	9.6790e-01
RAD	3.9489e-01	2.3912e-02	1.2947e-01	6.8020e-01	9.4964e-01
L1/16					
ML	3.9767e-01	**2.2141e-02**	1.5709e-01	7.4466e-01	9.7000e-01
AD	4.1918e-01	2.6217e-02	2.7960e-01	8.6512e-01	9.9089e-01
RAD	3.9570e-01	6.3266e-02	1.3474e-01	6.7170e-01	9.4684e-01
L2/4	$H = 8.575e-01$				
ML	8.7843e-01	3.7793e-02	9.2433e-04	4.9473e-02	2.7973e-01
AD	8.8427e-01	2.6330e-02	4.7449e-02	2.9014e-01	6.5015e-01
L2/8					
ML	8.7602e-01	**7.2309e-03**	3.1555e-04	1.7827e-02	1.2628e-01
AD	8.7946e-01	1.4181e-02	2.5898e-02	1.8984e-01	5.1514e-01
L2/16					
ML	8.7608e-01	**6.6401e-03**	4.6425e-04	1.2390e-02	1.0484e-01
AD	8.8048e-01	9.9469e-03	2.6237e-02	2.0412e-01	5.4638e-01
L3/4	$H = -2.104e-01$				
ML	3.0658e-01	8.4828e-01	1.4604e-04	5.1277e+00	2.3693e+01
AD	3.0661e-01	8.4824e-01	3.3590e-03	5.0848e+00	2.3434e+01
L3/8					
ML	2.9635e-02	**5.4836e-01**	7.3630e-05	2.0634e+00	5.1719e+00
AD	7.8277e-02	6.0827e-01	1.0497e-02	2.4274e+00	6.8083e+00
L3/16					
ML	-1.3451e-01	**2.8779e-01**	3.4638e-05	7.5127e-01	1.0167e+00
AD	-1.1618e-01	3.2116e-01	1.1931e-02	8.4018e-01	1.2541e+00

Appr.	Cr. Ent. ($\check{H}$)	Area D.	RME 1	RME 2	RME 3
U1/4	$H = 0.0$				
ML	1.3891e-01	3.1453e-01	6.8431e-05	2.6496e-01	2.754e-02
AD	1.6007e-01	2.7496e-01	8.3465e-02	6.9141e-01	4.946e-02
U1/8					
ML	9.8786e-02	2.2803e-01	8.1970e-05	1.0497e-01	1.112e-02
AD	1.1882e-01	2.0160e-01	4.9737e-02	3.6159e-01	2.217e-02
U1/16					
ML	7.1136e-02	1.6767e-01	1.8894e-04	4.1393e-02	4.863e-03
AD	9.7424e-02	1.6509e-01	3.4785e-02	2.4784e-01	1.391e-02
U2/4	$H = 0.0$				
ML	7.0956e-01	9.9978e-01	1.0420e-04	5.7514e+00	7.498e-01
AD	7.1263e-01	9.9720e-01	3.8192e-02	5.2442e+00	6.672e-01
U2/8					
ML	4.7837e-01	7.3796e-01	1.5747e-04	2.8584e+00	1.378e-01
AD	4.8372e-01	7.3290e-01	3.8091e-02	2.5689e+00	1.226e-01
U2/16					
ML	2.7843e-01	4.4080e-01	8.1234e-05	1.0769e+00	5.581e-02
AD	2.8100e-01	4.3575e-01	1.9523e-02	9.9713e-01	4.965e-02
SE/4	$H = 1.295e + 00$				
ML	1.3258e+00	1.8387e-01	3.5709e-04	2.7466e-02	1.9046e-01
AD	1.3278e+00	1.7478e-01	2.6355e-02	3.3976e-02	1.2472e-01
RAD	1.3449e+00	2.3464e-01	1.0494e-02	8.9644e-04	5.6800e-02
SE/8					
ML	1.3162e+00	1.3527e-01	4.5258e-04	8.3997e-04	2.4375e-02
AD	1.3179e+00	1.2575e-01	9.7610e-03	4.8098e-03	3.6399e-02
RAD	1.3278e+00	1.7125e-01	2.6487e-02	3.5088e-02	5.6718e-02
SE/16					
ML	1.3101e+00	1.0094e-01	5.1215e-04	2.7437e-03	2.4322e-02
AD	1.3129e+00	9.8741e-02	2.9152e-03	2.9267e-02	1.1545e-01
RAD	1.3138e+00	1.0919e-01	4.5643e-04	9.5314e-03	4.2254e-03
ME/4	$H = 7.277e - 01$				
ML	8.9794e-01	4.4689e-01	1.3964e-02	6.3662e-02	1.8893e-01
AD	9.1015e-01	4.3606e-01	3.0798e-02	2.2010e-01	7.7306e-01
RAD	3.6212e+00	1.1907e+00	7.4910e-01	9.3050e-01	5.0741e-01
ME/8					
ML	8.5417e-01	3.5047e-01	1.3729e-02	6.8973e-02	2.7284e-01
AD	8.6532e-01	3.2532e-01	1.3120e-02	6.5389e-02	2.4569e-01
RAD	3.7799e+00	1.1416e+00	7.1318e-01	8.5348e-01	1.0191e+00
ME/16					
ML	8.2535e-01	2.8761e-01	1.4200e-02	4.8011e-02	1.7391e-01
AD	8.3318e-01	2.6575e-01	8.9693e-03	9.1900e-02	3.6612e-01
RAD	3.6503e+00	9.6673e-01	5.3353e-01	1.0375e+01	8.2788e+02
P1/4	$H = 2.1295e + 00$				
ML	2.2323e+00	4.4603e-02	5.6533e-02	9.7564e-01	9.9998e-01
AD	2.2635e+00	3.2399e-02	1.6432e-01	9.9199e-01	1.0000e-00
RAD	2.2338e+00	6.6204e-02	1.1868e-01	9.8327e-01	9.9999e-01
P1/4+10					
ML	2.2246e+00	**1.7881e-02**	2.6115e-03	6.7019e-04	6.1649e-04
AD	2.2247e+00	1.8119e-02	8.5872e-03	9.7804e-04	6.1535e-04
P1/8					
ML	2.2292e+00	**1.3672e-02**	7.1673e-02	9.7659e-01	9.9998e-01
AD	2.2470e+00	2.7601e-02	1.2395e-01	9.8867e-01	1.0000e-00
RAD	2.2378e+00	8.9478e-02	1.2673e-01	9.8322e-01	9.9999e-01
P1/8+10					
ML	2.2245e+00	6.5659e-03	2.5887e-03	6.4993e-04	6.1648e-04
AD	**2.2244e+00**	5.6568e-03	1.2990e-03	7.3505e-04	6.1608e-04
P1/16					
ML	2.2282e+00	1.3221e-02	3.8874e-02	9.5945e-01	9.9993e-01
AD	2.2562e+00	**2.4655e-02**	1.4066e-01	9.9123e-01	1.0000e-00
RAD	2.2644e+00	4.6493e-01	1.5008e-01	8.1545e-01	9.9891e-01
P2/4	H=1.9811e+00				
ML	2.0768e+00	2.2020e-01	6.3924e-02	9.7417e-01	9.9997e-01
AD	2.1098e+00	1.7385e-01	2.0160e-01	9.9262e-01	1.0000e-00
RAD	2.1435e+00	3.1506e-01	6.9994e-02	9.8428e-01	9.9999e-01
P2/8					
ML	2.0414e+00	**4.4020e-02**	6.0110e-02	9.7176e-01	9.9997e-01
AD	2.0752e+00	9.8591e-02	2.2906e-01	9.9466e-01	1.0000e-00
RAD	2.0954e+00	3.7922e-01	6.2820e-02	9.8458e-01	9.9999e-01
P2/8+10					
ML	2.0598e+00	**4.4354e-02**	3.0374e-02	1.3027e-01	1.3101e-01
AD	2.0778e+00	8.3044e-02	1.1807e-03	1.2998e-01	1.3101e-01
P2/16					
ML	2.0363e+00	**2.7991e-02**	7.6029e-02	9.7586e-01	9.9997e-01
AD	2.0908e+00	5.5272e-02	2.4299e-01	9.9506e-01	1.0000e-00
RAD	2.1027e+00	2.2861e-01	8.6895e-02	9.8338e-01	9.9999e-01
P2/16+10					
ML	2.0336e+00	**2.5536e-02**	3.0803e-02	1.3033e-01	1.3101e-01
AD	2.0573e+00	5.8133e-02	4.0151e-02	1.2967e-01	1.3102e-01