Fitting Phase-Type Distributions to Data from a Telephone Call Center

Research Thesis

Submitted in partial fulfillment of the requirements

for the degree of Master of Science in Industrial

Engineering and Management - Statistics

Eva Ishay

SUBMITTED TO THE SENATE OF THE TECHNION -
ISRAEL INSTITUTE OF TECHNOLOGY
HESHVAN 5763 HAIFA October 2002

Acknowledgments

The research thesis was done under the supervision of Prof. Avishai Mandelbaum and Dr. Eitan Greenshtein in the department of Industrial Engineering.

The generous financial help of the Technion is gratefully acknowledged.

I thank Prof. *Avishai Mandelbaum* and Dr. *Eitan Greenshtein* for their professional contribution to the research and for their generous support throughout my studies.

I dedicate this work to my family with love.

Contents

List of Tables

List of Figures

Abstract

We analyze *Service Times* and *Customers' Patience* at a Call Center of one of Israel's banks. This is done by modelling and fitting *Phase-type (PH)* distributions to its data. The motivation is the optimization of call center performance. The correct design and management of a call center become possible only as a result of a system modelling and deep analysis of the data supporting the model. PH-distributions are used as the statistical models that provide a compact description of the data and possibly enhance our understanding of the mechanism of the underlying processes.

PH-distributions are defined as distributions of absorption times T in Markov processes with $k < \infty$ transient states (the phases) and one absorbing state Δ. There are several reasons for using the class of PH-distributions. One important property of PH-distributions is that they are dense and can be used to approximate any kind of distribution on $[0, \infty)$. Moreover, PH-distributions are sufficiently versatile and computationally tractable that they can be used to reflect the essential qualitative features of the model and to provide, through the interpretation of numerical results, much useful information on its physical behavior. They enable us to investigate the underlying processes going through the time a customer spends in service, or to understand customers' behavior by modelling their patience.

Service time is the positive time a customer spends with an agent, until departure from the service/system. Whereas the patience is the time a customer is willing to wait in queue before being served. We refer to the service time and the waiting time variables as the *survival time*, since both are times a customer has "survived" over some follow-up period, till occurrence of a certain event. When studying the service time, the event of interest is the time of departure from service. While measuring the patience, the customer's survival time becomes incomplete at the right side of the follow-up period. For customers who abandon the system before being served, the patience is their positive waiting time in queue before abandoning the system. On the other hand, the patience of customers who get the service is larger than their waiting time in queue and hence, the corresponding data constitutes

right-censored observations. The parameters of PH-distributions, for both censored and non-censored observations, are estimated via the *EM-algorithm* using the EMpht-program.

In this research we fit various phase-type distributions to empirical data sets, referred to as service duration and patience, according to priorities and service types. Qualitative comparison of empirical survival functions with the fitted ones is done by visual examination of their plots. The empirical survival functions are computed by non-parametric techniques, constructed for censored as well as for non-censored observations. The Kaplan-Meier setup for estimating the patience and the Kernel density estimator for estimating the density of service data are implemented, using S-PLUS and Matlab softwares. Simultaneous confidence interval for the empirical cumulative distribution function (CDF) provide heuristic stopping rules for adding phases of the fitted PH-distribution. We implement the Kolmogorov-Smirnov and Anderson-Darling goodness-of-fit tests to evaluate quantitative aspects of the produced fits.

We found that the general structure of order $k = 3$ already provides a reasonable fit to the service time. Moreover, the Coxian structure of the same order is also appropriate. The PH-model that provides a perfect fit to the patience is the general Coxian structure of order $k = 30$, which captures peaks that take place at the small time-interval, around 15 and 60 seconds, while the overall time-interval is over 1000 seconds.

We fit PH-distributions to four major service types of customers and priorities. We note a stochastic ordering between types and priorities. For example, it is demonstrated that high priority customers are more patient. This pattern emerges not just visually through survival and hazard curves, but also from the Coxian structure of order 30 fitted to the customers' patience of different priorities.

In view of the fact that Service Times of our call center turn out to be Log-normal distributed, according to Mandelbaum et al. [20], we compared Phase-Type to Log-normal distribution. Method of moments has been used for comparing between these two distributions. Furthermore, we found the optimal parameters of the PH-distribution numerically, for specific parameters of Log-normal distribution and for a given order of the PH-distribution. We implemented two numerical methods for this purpose: constrained nonlinear minimization, using Matlab, and minimization of information divergence, using EMpht.

List of symbols and abbreviations

Symbol	Meaning
$A-D$	Anderson-Darling
CDF	Cumulative Distribution Function
e	vector of ones
EDF	Empirical Distribution Function
EM	Expectation Maximization
h	bandwidth
$i.i.d.$	independent identically distributed
K	Kernel function
$K-S$	Kolmogorov-Smirnov
n	sample size
PH	Phase type
st	stochastically

Chapter 1

Introduction

1.1 Motivation

The call center industry constitutes a big, complicated group of people and companies, from many different backgrounds. It spans the globe, crosses industry borders, and includes everyone from reps to technology vendors. Somewhere between 2–3% of the American workforce works in a call center. That's an enormous amount [16, 29]. There are anywhere from 20,000 to 350,000 call centers, which employ anywhere between 4 to 6.5 million people (more than the entire agriculture sector) [17].

Call centers are becoming more important in financial services, and in this research we will focus on this sector of a telephone-based human-service operation. They are of importance to retail banking operations, credit card operations and mutual fund organizations. A significant part of the dynamics of call centers in financial services is similar to call centers in other industries. Financial services institutions are providing a rapidly expanding variety of products and services; technology is making customers more mobile, and delay is unacceptable in financial transactions. These attributes of the financial services sector mean that firms must provide effective, efficient and reliable service or quickly lose customers to competitors. To avoid huge labor costs, financial services firms must develop innovative approaches to manage their workforces and their service delivery process [18].

Quoting from Mandelbaum et al. [20], "*Call center* is the common term for describing a telephone-based human-service operation. A call center provides *tele-services*, namely services in which the customers and the service agents are remote from each other. The agents, who sit in cubicles, constitute the physical embodiment of the call center: with numbers varying from very few to many hundreds, they serve customers over the phone, while facing a

computer terminal that outputs and inputs customer data. The customers, who are only virtually present, are either being served, or they are waiting in, what we call, *tele-queues*: up to possibly thousands of customers sharing a phantom queue, invisible to each other and the agents serving them, waiting and accumulating impatience until one of two things happens – an agent is allocated to serve them (through a supporting software), or they *abandon* the tele-queue, plausibly due to impatience that has built up to exceed their anticipated worth of the service".

Telecommunication call centers have become the primary channel of customer service interaction for many businesses. The level of professionalism and efficiency that call center agents deliver to customers provides a significant advantage over traditional customer service practices. The growth of call centers has been substantial over the last two decades. This growth is driven by a company's desire to lower operating costs and to increase revenues. Given analytical and simulation-based models for the design and management of call centers, the goal is to optimize their performance. The system performance can be measured with quantities such as the mean waiting-time in queue, the expected time in system, the percentage of calls answered within a given time, the waiting-time probability distribution, and the abandonment rate [18].

Call centers are growing at unprecedented rates, yet relatively little is known about customer satisfaction with this method of service delivery. One of the main questions arising is: why do customers leave? Primarily because they don't get what they want. But it has less to do with price than service level: 45% of those who leave do so because of "poor service"; another 20% because of "lack of attention" (that's 65% leaving because you've done something wrong). 15% leave because they can find a cheaper product elsewhere, another 15% because they find a better product elsewhere, and 5% for other unspecified reasons [16, 29]. Hence, a major concern for service managers is the determination of how long a customer should wait to be served. Services, due to the customer's direct interaction with the process, must face a trade-off between minimizing the cost of having a customer wait and the cost of providing good service. Then in order to determine the least number of agents that could provide a given service level, it is critical to understand customers' *(im)patience* while waiting at the phone to be served.

The motivation behind this research work is analyzing and modelling of service durations and patience by measuring this data. The correct design and management of a call center, and the optimization of call center performance become possible only as a result of system modelling and deep analysis of the data supporting the model. The source of the data used in this research is a small call center of one of Israel's banks. The center provides

several types of services: information for current and prospective customers, transactions of checking and saving accounts, stock-trading, and technical support for Internet users of the bank's site.

1.2 Objective

The purpose of this research is analyzing and modelling the call center data described above and then fitting phase-type distributions (see Chapter 5) to the data. There are several motivations for using phase-type distributions as statistical models. These distributions arise from a generalization of Erlang's method of stages in a form that is particularly well-suited for numerical computation: problems which have an explicit solution assuming exponential distributions are algorithmically tractable when one replaces the exponential distribution with a phase-type distribution. Furthermore, the class of phase-type distributions is dense and hence any distribution on $[0, \infty)$ can, at least in principle, be approximated arbitrary close by a phase-type distribution. In some applications, the phases have no physical interpretation and the phase-type modelling is purely descriptive. However, in other areas such as demography, drug kinetics, epidemiology, etc, the probabilistic interpretation fits in nicely with standard Markovian modelling.

It may often be natural to think that there are underlying processes going through a set of stages till the occurrence of certain events. For example, a primary interest is in modelling waiting times, where the waiting time is defined as a positive time in queue. It is important for understanding customer behavior - patience, and for analyzing how it might be different for customers with different priorities or different types of service. As defined in [20], *patience* is the time a customer is willing to wait in queue before being served. For customers who abandon the system before being served, patience will be estimated by their positive waiting time in queue before abandoning the system. On the other hand, for customers who get the service, their patience is larger than their waiting time in queue and hence they are right-censored observations. In order to estimate patience, it is necessary to use techniques for analyzing censored data. Another reason for being interested in the underlying processes is the need to understand better the hazard rate or intensity (see section 4.1, 4.2) of the distributions of interest. One may ask why the hazard rate assumes various typical shapes, sometimes increasing, sometimes decreasing, sometimes with a unimodal shape of increase followed by a decrease.

Modelling service time data via phase-type distributions allows us to understand a service, which consists of a random sequence of tasks such that a

phase corresponds to a task. According to Mandelbaum et al. [20], the service times are Log-normal distributed. Therefore, the comparison between Phase-type and Log-normal distributions is natural.

Once we have a model, it is important to check whether it is any good or not. Typically this is judged by comparing the empirical functions with corresponding ones obtained from the fitted model. The nonparametric methods for estimation of the density, the distribution, the survival and the hazard functions are implemented for both censored and non-censored observations. The plots of the empirical functions together with the fitted ones, simultaneous confidence interval for empirical distribution function, and goodness of fit tests allow us to assess whether a particular phase-type distribution provides an adequate fit to the data.

1.3 Outline of the Research

The organization of this research work is as follows.

Chapter 2 presents a survey of several articles where phase-type distributions play a central role. The applications of phase-type distributions in the various fields of study are described. It contains various examples of fitted phase-type models to both censored and non-censored real data-sets. The main results of previous works are pointed out as well.

Chapter 3 gives partial description of the data-base of the telephone call-center of "Anonymous Bank" in Israel, used in this research, and focuses on studying two types of the data: service durations and customer patience of several types and/or priority rules. The complete information about the data-base and the design of the telephone call-center of "Anonymous Bank" in Israel are given in [20].

Chapter 4 describes the nonparametric methods for estimation the density, the survival and distribution functions, and the hazard rate. Estimation from a sample containing right-censored observations as well as non-censored is presented. Considering right-censored observations, the Kaplan-Meier estimator is used. The kernel density estimator is described for estimating density of non-censored observations. Several S-PLUS functions are mentioned for implementing these techniques.

Chapter 5 contains the formal definition of phase-type distributions. We review early in the chapter the reasons for introducing this highly versatile class of probability distributions. We illustrate some basic dis-

tributional characteristics and properties of PH-distributions, and introduce several familiar distributions, which are simple examples of the phase-type distributions. An example of non-identifiability of the parameterization of the phase-type distribution is illustrated too.

Chapter 6 is entitled after the EMpht-program that was kindly supplied to us by Marita Olsson [26]. This is a program for fitting phase-type distributions, either to a sample (which may contain censored observations), or to another continuous distribution, using an EM-algorithm. The chapter concludes with a discussion about the EM-algorithm and its properties, and an example for illustrating an implementation of the proposed algorithm to the Hyperexponential distribution with two phases.

Chapter 7 includes a discussion about an heuristic stoping rule for adding phases into the model, that is the confidence band for empirical cumulative distribution function. It is used as a graphical method, which has an advantage being very simple and effective. In addition, the chapter describes two statistical tests for goodness-of-fit, the so-called EDF tests, based on the empirical distribution function and enabling to compare quantitatively the produced fits.

Chapter 8 analyzes two kinds of empirical data, referred to as service durations and patience. Various phase-type distributions were fitted to these data-sets, according to priorities and service types. Qualitative comparison of empirical survival functions with the fitted ones is done by visual examination of their plots. The empirical survival functions are computed by non-parametric techniques, described at Chapter 4. The parameters of fitted phase-type distributed functions are derived with EMpht-program, described at Chapter 6. The simultaneous confidence interval for empirical cumulative distribution function is referred as an heuristic stoping rule for adding phases. This sensitive graphical method allows us to select the phase-type model with the least number of phases. Goodness-of-fit techniques based on supremum and quadratic empirical distribution function statistics, namely Kolmogorov-Smirnov and Anderson-Darling tests, respectively, are implemented to compare qualitatively the produced fits.

Chapter 9 discusses the problem of minimization of the distance between Phase-type and Log-normal densities. It is of interest, since according to Mandelbaum et al. [20], the service times are Log-normal distributed. Method of moments is used for comparing between these

two distributions. Furthermore, the optimal parameters of the phase-type distribution are numerically found for specific parameters of Log-normal distribution and for specified order of phase-type distribution. Two numerical methods are implemented for this purpose: the constrained optimization, using Matlab, and the minimization of information divergence, using EMpht. The results are obtained in three cases: Log-normal($\mu = 1, \sigma = 0.5$), Log-normal($\mu = 1, \sigma = 1$) and Log-normal($\mu = 0, \sigma = 1$). In addition, an example of real data of overall service time - December is considered. The parameters of Log-normal distribution, corresponding to the data, are derived by maximum likelihood estimation. Using EMpht, we approximate this Log-normal distribution by phase-type one of a specific order.

Chapter 2

Literature review: Application and examples of phase-type distributions

Phase-type distributions have received a lot of attention in applied probability, in particular in queueing theory where they generalize the classical Erlang distributions. A survey of several articles where phase-type distributions play a central role is given herein.

2.1 "On phase-type distributions in survival analysis" by O.Aalen

O.Aalen's article [2] demonstrates that phase-type distributions should find greater application in biostatistics. This survey describes various kinds of phase type models, connecting them to problems in survival analysis. Phase type models have been used in medical statistics and other fields. Examples include studies of the incubation period of AIDS, and of the duration of genital herpes lesions. The description of the model for the incubation time of AIDS is shown in Figure 2.1. One sees that the model has two phases for developing to advanced HIV disease. In that stage treatment may be offered, described by a rate γ. The effect of treatment is to slow down the further progression to AIDS by a factor θ. The incubation period in this model is the time it takes to go from state 1 (HIV infection stage) to state 5 (AIDS stage).

The model for the incubation distribution of AIDS is an example of phase type models of *acyclic* type, that is, where no state can be visited more than once. Often, in biostatistical applications, this inexorable development in

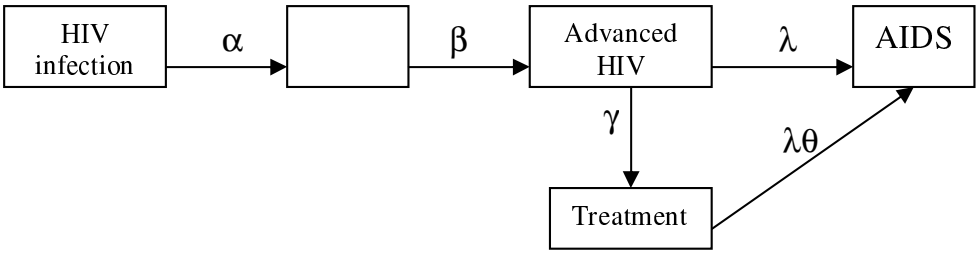


Figure 2.1: Phase-type model for incubation distribution of AIDS.

one direction will be unrealistic, and it is more natural to assume that the process moves back and forth between states even though absorption may eventually take place. Such models are so-called *models with feedback*. It is always possible to arrive from models with feedback to a particularly simple type of acyclic model – the *series model*, where all states are ordered and a transition can only go to the next state in ordering, with the last state being absorbing.

There is another example illustrating fitting phase-type distribution to real data [2]. The example concerns women having had a live-born first child, and the object is to analyze the time until their next birth, if any. The analysis of birth intervals is of interest in demographic research. A random sample of married woman who had a live-born first child during the period 1967–71 was taken. The total number of women included was 1779, and they were followed up until 1982. The number of women giving birth and the number of women that is not known whether they had another child or had not, means the number who are censored, have been registered in intervals of six months, starting nine months after the first birth.

A phase type model for the situation at hand should incorporate the following features. Since it may be assumed that the women (or couples) will generally wait a while before attempting to have another child, it is natural to incorporate at least two stages in the process. Hence, the minimal model would consist of three states, where the transition from the first to the second state might mean that the couple is ready to have another child, while transition from the second to the third state represents birth of the next child. Next, one should consider the *heterogeneity* between individual women. Some women will bear another child quite soon, while for others it may take many years. Hence there should be at least two ways to proceed through the states, a fast one and a slower one. Finally, one should incorporate the possibility that some women will never conceive another child for an individual decision, or some medical problems, or other circumstances. Hence there should be an

extra absorbing state in the state space. A simple model which incorporates these features is given in Figure 2.2. This simple and special structure is made

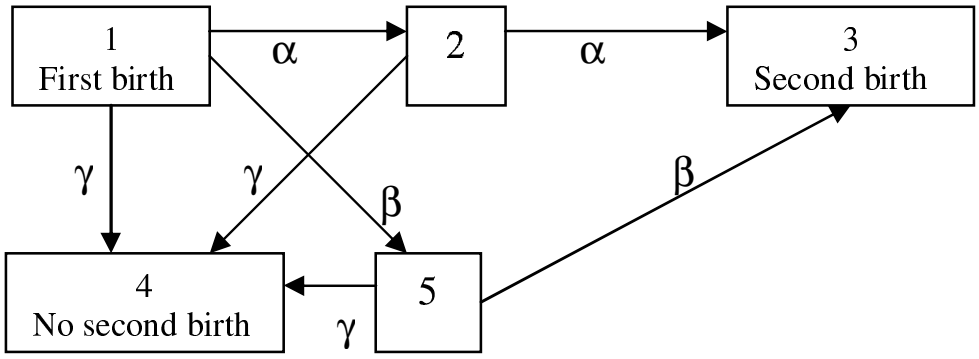


Figure 2.2: Phase-type model for birth interval example.

possible for explicit calculation of distribution, density and hazard functions. The fit appears to be reasonably good, and thus one has an illustration that phase type modelling may be a practical tool in statistical analysis.

2.2 "Fitting phase-type distributions via the EM algorithm" by S.Asmussen et al

S.Asmussen et al [5] present a general statistical approach to estimation theory for phase-type distributions. The idea of this approach is: the class of phase-type distributions may for a fixed k (the number of transient states) be viewed as a multi-parameter exponential family, provided the whole of the underlying absorbing Markov process is observed. Since the data in practice consist of i.i.d. replications of the absorption times $Y_1, ..., Y_n$ of Y, there are incomplete observations and it is given to implement the EM algorithm. The EM (expectation-maximization) algorithm is an iterative maximum likelihood method for estimation the elements of ($\boldsymbol{q}$,$\boldsymbol{R}$), the parameters of the phase-type distribution. The program for implementation of the proposed algorithm is EMPHT-program [11]. The performance and the dynamics of the algorithm illustrated in S.Asmussen et al [5] by a sequence of fits of phase-type distributions to three different theoretical distributions: Weibull, Log-normal and Erlang distribution with feedback. Furthermore, it was found a Coxian distribution to provide almost as good fit as a general phase-type distribution with the same k, for one exception: the Erlang distribution with feedback. It is also presented that all phase-type distributions

corresponding to *acyclic* distributions (that is distribution whose generator is upper triangular), coincide with the Coxian distributions.

In the same article [5] exhibited four samples of the lengths of incoming telephone calls to the service center of one of Israel's major television cable companies supplied by Professor A.Mandelbaum and Professor O.Kella. The calls are classified into types 0–10. The four types, 1, 3, 4, and 7, of incoming calls having the largest number of observations were taken for fitting phase-type distributions. The types have the following meaning: type 1) "home services"; type 3) "sales"; type 4) "billings"; type 7) "general information". All four samples have fitted both a general phase-type structure and a Coxian structure. For samples of type 1,4,7 it has not been possible to distinguish the fitted Coxian density from the fitted general phase-type density in the graphs. When fitting phase-type distributions to the sample of type 3, it was discovered that the phase-type distribution with feedback gave better fits (according to the log-likelihood) than the Coxian structure, although the difference is hard to see in plots of the densities. However, as much as the approximation is of the higher order k, the fits of the general phase-type and the Coxian structures gave the same log-likelihood value. The Coxian structure has the advantage of being much faster to fit.

2.3 "Estimation of phase-type distributions from censored data" by M.Olsson

In M.Olsson [25] it is shown how the EM algorithm can also be extended to cover estimation from censored data: right-censored and interval-censored observations. To test the performance of EM algorithm for data set of right-censored type of observations, the survival functions of the fitted phase-type distributions were compared with the Kaplan-Meier estimate. In the interval-censored examples, the distribution function fitted with EM algorithm have plotted together with the Turnbull estimate. According to survival function fitted to one of the data sets presented in this paper, the melanoma data-set, it is shown that as much as the fit of phase-type distribution is of higher order so it is much closer to the non-parametric survival estimator, the Kaplan-Meier estimator in this case. The same results are received by fitting phase-type distributions to data-set — Data on Hepatitis A in Bulgaria, which includes interval-censored observations.

2.4 "Analysis of the $\Sigma Ph_i/Ph/1$ queue" by G.R.Bitran and S.Dasu

G.R.Bitran and S.Dasu [6] have been analyzing a queue to which the arrival process is the superposition of separate arrival streams, each of whose interarrival time distributions is of phase type, and the service time distribution is also of phase type, that is $\Sigma Ph_i/Ph/1$. There are several situations in which the arrival process is the superposition of different arrival streams. For example, in multi-echelon distribution systems if the time between orders from each retailer to the central plant has a phase-type distribution, then the arrival of orders at the plant will be the superposition of phase renewal processes. Such a situation can arise if each retailer observes Poisson demand, and orders a fixed quantity k from the plant. Under these assumptions, the time between consecutive orders from each retail outlet will be distributed as an Erlang distribution of order k. The performance measures derived for $\Sigma Ph_i/Ph/1$ queue include: the distribution of the number in the system as seen by each customer class upon arrival, Laplace-Stieltjes transform of the waiting-time distribution for each customer class, characteristics of the tails of the waiting time and queue length distributions.

2.5 "Parallel-Processing Times: Extreme Values of Phase-type and Mixed Random Variables" by S.Kang and R.F.Serfozo

The purpose of S.Kang and R.F.Serfozo [12] is to determine the distribution of the time to complete a large number of tasks in parallel. That is, knowing the probability structure of the individual tasks, what is the asymptotic distribution of the maximum of the task times as the number of tasks tends to infinity? The problems discussed in the article are: a) a task consists of performing a set of randomly selected subtasks in series and the subtask durations have Erlang distributions; b) the task times are independent, identically distributed phase-type random variables; c) the tasks are dependent and their distributions are selected by a random environment process. In the paper it is shown that the distributions of the task completion times in the three setting above are the three classical extreme-value distributions — the classical Gumbel distribution in setting (a) and (b) and, any one of the three distributions in setting (c).

Chapter 3

Data Description

The source of the data [20]:
(http://iew3.technion.ac.il/serveng/callcenterdata/index.html) The call center of "Anonymous Bank" is the source of the data used in this research for constructing models and their analysis. The call center of "Anonymous Bank" provides several different service:

- Information on and transactions of checking and saving accounts, to bank-customers.
- Computer generated voice information (through Voice Response Unit(VRU)).
- Information for prospective customers.
- Support for Internet customers.

The call center consists of 8 regular-agent positions, 5 Internet-agent positions, and one shift-supervisor. Working hours are weekdays (Sunday to Thursday) from 7am to midnight; the center closes at 2pm on Friday and reopens around 8pm on Saturday. The automated service (VRU) operates 7 days a week, 24 hours a day.

The data archives all the calls handled by the call center, on a monthly basis, from January 1999 to December 1999. The number of phone calls recorded per month fluctuates between 20,000 to 40,000 calls (excluding the calls satisfied with self-service transactions at the VRU). The variables per phone call include the following:

- Service durations in seconds.
- Waiting time in queue is in seconds. An announcement is replayed every 60 seconds or so, with music, news or commercials intervened.

- Types of departure system: AGENT – a customer receives a service; HANG – a customer abandons the system after he gets tired of waiting to service.

- Priority types: "0 or 1" – indicates unidentified customer or regular customer, "2" – indicates high-priority customer. Regular customers join the "end" of the tele-queue, while high-priority customers are advanced in the queue by 1.5 minutes right upon arrival. Service is then rendered on a first-come-first-served (FCFS) basis. Customers have not been told about the existence of priorities.

- Service types: PS – regular activity, PE – regular activity in English, IN – Internet consulting, NE – stock exchange activity, NW – potential customer getting information, TT – customer who left a message asking the bank to return his call.

- Customer identification: in the most calls a customer identified by his ID.

- Server identification: in all phone calls agent who provides service is identified by his name.

The analysis in this research work focuses on studying two types of the data: service durations and customer patience of several types and/or priority rules. *Service durations* is the time a customer spends with agent, ignoring records with zero service time, since these records are mostly of customers abandoned the system directly from automated service (VRU). In order to estimate the patience, it is necessary to use waiting time, considering waiting time as a positive time in queue, for the customers who abandon the system (HANG) and for the customers who get the service (AGENT).

Let us refer to the service time and waiting time variables as *survival time*, because it gives the time that the customer has "survived" over some follow-up period. In survival analysis one studies the times to occurrence of certain events. When studying the service time, the event of interest is the time of departure from service. When measuring the patience, a customer's survival time becomes incomplete at the right side of the follow-up period, occurring when the customer get the service (AGENT), thus the survival time of this kind of data is right-censored. The distribution of survival time is usually described or characterized by three functions: (1) survival function, (2) the probability density functions, and (3) the hazard function. These three functions mathematically equivalent - if one of them is given, the other two can be derived.

Chapter 4

Nonparametric Methods for Estimation of Survival Functions

In order to see how well PH-distributions of different structure and order can fit real data, the fitted cumulative distribution function and/or survival function, the fitted density function, and the fitted hazard rate are compared to appropriate empirical functions estimated with non-parametrical methods.

4.1 Estimation of Service time

Let consider a single nonnegative random variable on (0,∞), T that denotes positive service time until departure from service. Its continuous distribution specified by a cumulative distribution function F with a density f. In a survival analysis it is more usual to work with the *survival function* $S(t) = 1 - F(t) = P(T > t)$, the *hazard function* $h(t) = \lim_{\triangle t \longrightarrow 0} P(t \leq T < t + \triangle t | T \geq t)/\triangle t$ and the *cumulative hazard function* $H(t) = \int_0^t h(s)ds$. These are all related by:

$$h(t) = \frac{f(t)}{S(t)}, \qquad H(t) = -\log_e S(t).$$

Survival function: The survival function of service time estimates as the proportion of customers surviving longer than t is:

$$\begin{aligned}\hat{S}(t) &= \frac{\text{number of calls still receiving service at time t}}{\text{total number of calls in service}} \\ &= \frac{1}{N}\sum_{i=1}^{N} 1(T_i > t), \quad \text{where} \quad 1(T_i > t) = \begin{cases} 1, & \text{if } T_i > t \\ 0, & \text{if } T_i \leq t \end{cases} \end{aligned} \tag{4.1}$$

A steep survival curve represents low survival rate or short survival time. A gradual or flat survival curve represents high survival rate or longer survival. The survival function or the survival curve (the graph of $S(t)$) is used to find the 50th percentile (the median) and other percentiles of survival time and to compare survival distributions of different types and/or priorities. The mean is used to describe the central tendency of a distribution, but by modelling data from the call center of "Anonymous Bank" the median is often better because a large number of calls with exceptionally long or short lifetimes will cause the mean survival time to be disproportionately large or small.

Density function: The probability density function of service time estimates as the proportion of calls in service ended in the short interval time:

$$\hat{f}(t) = \frac{\text{number of calls in service ending in the interval beginning at time t}}{(\text{total number of calls in service})(\text{interval width})}$$

The proportion of callers that departure from the service in any time interval and the peaks of high frequency of departure can be found from density function.

The simplest nonparametric method to estimate f that easy to construct and interpret is *histogram*, but histograms have a major disadvantage, namely they are discontinuous. This might be less crucial when the histogram is being used only as a summary of the data. However, if the histogram is being used as an intermediate step in a more complicated procedures, which require continuity of the density then the use of the histogram is not good. The estimator that has the continuity and smoothness properties is the *kernel density estimator*, and it has the form:

$$\hat{f}(t) = \frac{1}{Nh}\sum_{i=1}^{N} K\left(\frac{t - T_i}{h}\right), \tag{4.2}$$

where K is a smooth function known as *kernel function*, satisfies $\int K(u)du = 1$ and h is the bandwidth. The most widely used kernel is the Gaussian

kernel $K(u) = \phi(u) = (2\pi)^{-1/2} exp(-u^2/2)$. The kernel density estimator is sum of 'bumps' placed on t, when a kernel K determine the shape of the bumps, and the bandwidth h determine its width. The choice of bandwidth is a compromise between smoothing enough to remove insignificant bumps, *over-smoothing* and not smoothing too much to smear out real peaks, *under-smoothing*. This reflects a fundamental tradeoff in all smoothing methods – bias versus variance, when *bias* measures how close the estimator, $\hat{f}(t)$, is to the true density, while *variance* can be seen by the variability in heights of neighboring bumps. For a small bandwidth the bias is small and the variance is large and vice versa.

The function `density` is the S-PLUS command that estimates the density [31]. This function returns x and y coordinates of a non-parametric estimate of the data. Options include the choice of the window to use and the number of points at which to estimate the density. The kernel is the normal by default, with alternatives "rectangular", "triangular" and "cosine" (kernel used in this research is normal).

Hazard function: The hazard function of service time defines as the probability that callers finishing to receive their service during a very small time interval, assuming they were in service at the beginning of the interval. Then the hazard rates estimates as:

$$\hat{h}(t) = \frac{\text{number of calls in service ending at the interval beginning at time t}}{\text{(number of calls still in service at time t)(interval width)}}$$

The service time hazard rates give the local behavior of a customer. This estimate is unstable as the time increases, as the remaining population becomes smaller. To get a better picture of the hazard rate, they are smoothed with nonparametric regression method *super-smoother*, using S-PLUS function `supsmu`. The smoother `supsmu` based on a symmetric k-nearest neighbor linear least squares procedure. Cross-validation is used to choose a value of k.

There are a number of reasons why estimation of the hazard function for service time and patience may be a good idea, according to A.Mandelbaum [19]:

- The hazard rate is a *dynamic* characteristic of a distribution.
- The hazard rate is more precise "fingerprint" of a distribution than cumulative distribution function, the survival function, or density (for, example its tail need not to converge to zero; the tail can increase, decrease, converge to some constant etc.)

- The hazard rate provides a tool to compare the tail of the distribution of interest against a "benchmark": exponential distribution.
- The hazard-based models are often convenient when there is censoring.

4.2 Estimation of Patience

Let us consider a single nonnegative random variable on $(0,\infty)$, W that denotes a positive waiting time in queue until one of the following events occur: a) entering an agent for receiving service (AGENT); or b) abandoning the queue due to lack of patience (HANG). Denote by V the "virtual waiting time" and by T the "time willing to wait before abandoning", defined as the patience [32]. We observe $W = \min(V, T)$ and an indicator for observing V and T (AGENT and HANG outcomes, respectively). Since our purpose is to estimate distribution of T, we consider all calls that reached an agent as censored observations.

Survival function: The estimation of survival function of patience is implemented by introducing the product-limit (PL) method of estimating the survival function developed by Kaplan and Meier (1958) or *Kaplan-Meier* (KM) method.

The KM setup for estimating patience is as follows from Mandelbaum et al. [32]: "There is given a sample W_i of N waiting times from a call center. Some of the calls end up with abandonment ($W_i = T_i$) and the others with a service ($W_i = V_i$). Denote by $M \leq N$ the number of *distinct* abandonment times in the sample. Let $T^1 < T^2 < \cdots < T^M$ be the ordered observed abandonment times, and A_k the number of abandonment at T_k units of time. The *Kaplan-Meier estimator* $\hat{S}(t), t \geq 0$, estimates survival function $\bar{F}(t) = P(T > t)$, where T is the time to abandon (patience). It is given by

$$\hat{S}(t) = \prod_{k:B_k \leq t} (1 - \frac{A_k}{B_k}) \tag{4.3}$$

$$= \prod_{k:B_k \leq t} (1 - \hat{h}_k), \tag{4.4}$$

where B_k denotes the number of customers still present at T_k, that is neither served nor abandoned before T_k. The $\hat{h}(t)$ is the estimated hazard rate of the patience. The estimator for mean patience is then based on the tail-formula

$$\widehat{E[T]} = \int_0^\infty \hat{S}(t)dt. \tag{4.5}$$

On Independence: KM assumes independence for the observations whose distribution is to be estimated. Such an independence is plausible for patience, despite the repeat calls by the same customers. It is also requires an independence of the patience and the censoring time. It was found out that at the entrance to the queue, and thereafter every minute or so, customers are exposed to an automatic message. This message informs customers about their queue, which possibly affects their patience, hence it is not clear, how much information a customer actually gains out of it, and how strong the dependence is."

The S-PLUS command that computes an estimate of a survival curve is `survfit`. The estimate of the survival curve for uncensored data is one minus the empirical distribution function. The function `survfit` also handles censored data, and uses KM estimator by default.

Hazard function: The estimated hazard at a time interval is the number of failures during that interval (i.e. number who abandon), divided by the number at risk at the beginning of the interval (i.e. number of calls who were still waiting at the beginning of the interval). Those raw hazards very noisy and unstable for large time, as the remaining population becomes small. The raw hazard rates are the building blocks for the Kaplan-Meier estimate of the survival function (formula (4.4) above). The KM estimator is very sensitive to censored data at the upper tail of the sample. For example, if the longest wait in a customer's history ended up with an abandonment, the KM estimator has a positive mass at infinity. In order to get the continuous and smoother hazard curve we use *super-smoother* method, mentioned in section 4.1. The S-PLUS function `supsmu` smoothes the raw hazard rates above, when the hazard rates for non-failure times are zero (for correction the behavior of the tail).

Density function: Given the estimated survival function and hazard function, a natural way to estimate the density of patience is by:

$$\hat{f}(t) = \hat{S}(t) \cdot \hat{h}(t) \tag{4.6}$$

Chapter 5

Phase-type distributions

5.1 Definitions

The continuous distribution of phase-type (PH-distribution) is the distribution of *time to absorption* of an absorbing Markov process. Correspondingly, the discrete PH-distribution is the distribution of *number of steps to absorption* in an absorbing Markov chain. Let us define the continuous PH-distribution precisely.

Definition 5.1 *Consider a finite Markov process* $\{X(t), t \geq 0\}$ *on the state space* $\{1, 2, ..., K, \Delta\}$ *with infinitesimal generator* $\boldsymbol{Q}$. *The* $\{1, ..., K\}$ *are all transient and* Δ *is the only absorbing state of the process. Let* q_k *be the probability of starting in the transient state* k, *for* $1 \leq k \leq K$, *when* $q_k = P(X_0 = k)$, *and the probability of starting in the absorbing state,* q_Δ, *is zero. That is,* $\boldsymbol{q} = (q_1, ..., q_K, 0)$ *be some initial distribution. Write*

$$\boldsymbol{Q} = \begin{pmatrix} \boldsymbol{R} & \boldsymbol{r} \\ 0, ..., 0 & 0 \end{pmatrix}, \tag{5.1}$$

where r_k *(the* k*-th element of* $\boldsymbol{r}$, *the* exit-rate vector*) is the conditional intensity of absorption in* Δ *from state* k. *The* $(K \times K)$*-dimensional matrix* $\boldsymbol{R}$ *is called the* phase-type generator. *The matrix* $\boldsymbol{R}$ *satisfies* $\boldsymbol{R}_{kk} < 0$, *for* $1 \leq k \leq K$, *and* $\boldsymbol{R}_{kj} \geq 0$, *for* $k \neq j$. *Since every row in* $\boldsymbol{Q}$ *sums to zero, it follows that* $\boldsymbol{r} = -\boldsymbol{R}\boldsymbol{1}$, *where* $\boldsymbol{1} = (1, ..., 1)'$, *a column-vector of* K *ones. Let* $T = inf\{t > 0 : X(t) = \Delta\}$ *be the absorption time in the state* Δ. *Then* $F_T(t) = P_q[T \leq t]$ *is a PH-distribution.*

The pair ($\boldsymbol{q}$,$\boldsymbol{R}$) is called a *representation* of $F(\cdot)$ and k is the *order* of the representation. The class of phase distributions (as k, q, and Q vary) is called the *PH-class*.

5.2 Characteristics of PH-distributions

Some basic distributional characteristics of PH-distributions are:

- distribution function $F_T(t) = 1 - \boldsymbol{q} exp\{\boldsymbol{R}t\}\mathbf{1}$, (5.2)
- density $f_T(t) = \boldsymbol{q} exp\{\boldsymbol{R}t\}\boldsymbol{r}$, (5.3)
- Laplace transform $\int_0^\infty exp\{-xt\}F_T(dt) = \boldsymbol{q}(x\boldsymbol{I} - \boldsymbol{R})^{-1}\boldsymbol{r}$, (5.4)
- rth moment $E(T^r) = \int_0^\infty t^r F_T(dt) = (-1)^r r!\boldsymbol{q}\boldsymbol{R}^{-r}\mathbf{1}$. (5.5)

5.3 The special cases of PH-distributions

There are several reasons for using the class of phase-type distributions as statistical models. The most established ones come from their role as the computational vehicle of much of applied probability. These distributions are much applied in queueing theory, where they include commonly used distributions like the exponential distributions, the Erlang distribution, sum of exponential, and mixtures of exponential distributions. An Erlang distribution, that is the distribution of the sum of independent identically distributed exponential random variables, and Hyperexponential distribution are simple examples of the PH-distributions:

a) Erlang distribution: K independent and identically distributed tasks, given in Figure 5.1. It can be represented as a PH-distribution of order K, where the underlying Markov-process starts in state 1 (implying $\boldsymbol{q}= (1, 0, \ldots, 0)$), then visits each state $2, \ldots, K$, and terminates when leaving state K.

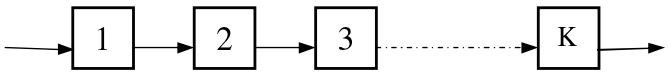


Figure 5.1: Erlang distribution.

b) Hyperexponential distribution: k tasks in parallel, given in Figure 5.2. The Markov-process can start in any state (all elements in $\boldsymbol{q}$ are allowed to be non-zero), but terminates in absorbing state without visiting any other states (all elements in $\boldsymbol{R}$ off the main diagonal is zero).

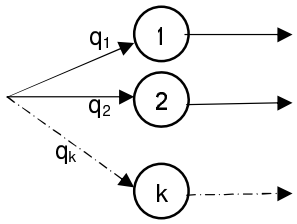


Figure 5.2: Hyperexponential distribution.

c) Coxian distribution is constructed as a sum of exponential, or generalized Erlang-distribution, where the Markov-process starts in state 1, and then visits each state $2, \ldots, k$, with the exception that the absorbing state can be reached from all the transient states in the process, see Figure 5.3.

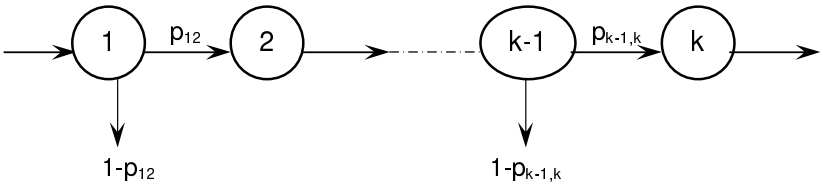


Figure 5.3: Coxian distribution.

d) Erlang mixtures:

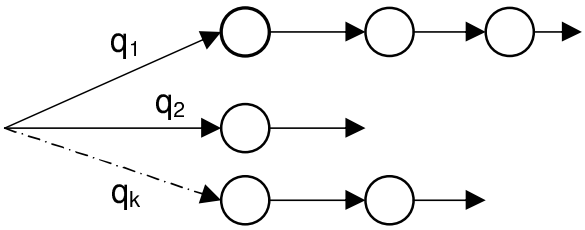


Figure 5.4: Erlang mixtures.

5.4 Properties of PH-distributions

There are several important properties of PH-distributions:

- *dense*: PH-distributions can be used to approximate all probability distributions on $[0, \infty)$ and applied in the large area of statistical fitting. For every non-negative distribution G, there exists a sequence of phase-type distributions $F_n \ni F_n \Rightarrow G$.
- *Markov modelling*: PH-distributions arise from a generalization of Erlang's method of stages in a form that is particularly well-suited for numerical computation: problems which have an explicit solution assuming exponential distributions are algorithmically tractable when one replaces the exponential distribution with a phase-type distribution.
- *structurally informative*: PH-distributions are sufficiently versatile and computationally tractable that they can be used to reflect the essential qualitative features of the model and to provide, through the interpretation of numerical results, much useful information on its physical behavior.

5.5 The non-identifiability of PH-distributions

There are usually several different representations of a PH-distribution: different setups of parameters can correspond to the same distribution, the parameters $(\boldsymbol{q}, \boldsymbol{R})$ are thus not identifiable. Therefore, we should be cautious in giving a physical interpretation to the k phases of the process $\boldsymbol{Q}$. The following are some examples of different representations of a PH-distribution.

General Erlang distribution of second order: This distribution can be represented with two different setups of parameters $(\boldsymbol{q}, \boldsymbol{R})$ and $(\boldsymbol{q}, \boldsymbol{R})'$:

(1) $\longrightarrow \boxed{\lambda_1} \longrightarrow \boxed{\lambda_2} \longrightarrow$

$$\mathbf{q} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \mathbf{R} = \begin{bmatrix} -\lambda_1 & \lambda_1 \\ 0 & -\lambda_2 \end{bmatrix}$$

(2) $\longrightarrow \boxed{\lambda_2} \longrightarrow \boxed{\lambda_1} \longrightarrow$

$$\mathbf{q}' = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad \mathbf{R}' = \begin{bmatrix} -\lambda_1 & 0 \\ \lambda_2 & -\lambda_2 \end{bmatrix}$$

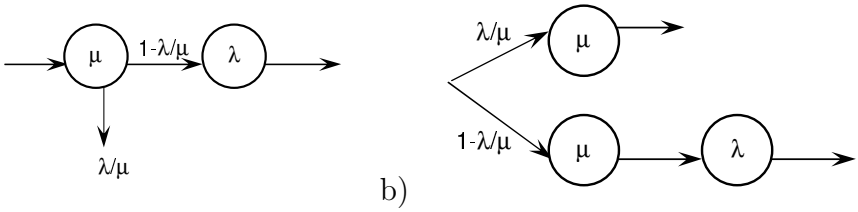
Let X_1, X_2 be two independent random variables exponentially distributed with parameters $\lambda_i, i = 1, 2$. Then $Y = X_1 + X_2$ is time to absorption, and in both cases:

$$f_Y(y) = \frac{\lambda_1 \lambda_2}{(\lambda_1 - \lambda_2)} \left(e^{-\lambda_2 y} - e^{-\lambda_1 y} \right), y > 0.$$

Figure 8.3 (section 8.1.1, p. 44) shows another example of two different structures of PH-distribution of order $k = 3$ with the same log likelihood function.

Exponential distribution with parameter λ: According to S. Asmussen, any PH-distribution with a complicated PH-generator of order $k \geq 2$, but with absorption rates independent of the phase, have an extremely representation of the exponential distribution. Figure 5.5 a) presents the special case of this statement - the PH-distribution of the Coxian structure of order $k = 2$ with the absorption rate λ from each phase. Figure 5.5 b) shows another rep-

Figure 5.5: The nicest examples of the Exponential distribution with parameter λ.



resentation of the same distribution. Let X_1, X_2 - two independent random variables exponentially distributed with parameters μ and λ, respectively, and an indicator

$$I = \begin{cases} 1 & \text{w.p. } 1 - \frac{\lambda}{\mu} \\ 0 & \text{w.p. } \frac{\lambda}{\mu} \end{cases}$$

independent of $X_i, i = 1, 2$. Then $Y = (1 - I)X_1 + I(X_1 + X_2)$ is time to absorption, exponentially distributed with parameter λ.

The *non-identifiability* of the parameterization of the PH-distributions implies that there can be different set-ups of estimates $(\hat{\boldsymbol{q}}, \hat{\boldsymbol{R}})$ corresponding to the same PH-distribution. Typically, when different fits to a sample or a distribution are performed using EMpht-program (see section 6), they will result in different parameter values. However, the value of the likelihood function are most often almost the same for different fits, and the corresponding densities are, when plotted, seldom possible to distinguish from each other [26].

Chapter 6

The EMpht-program

EMpht is a program for fitting phase-type distributions. This program was kindly supplied by Marita Olsson [26] and greatly appreciated. It can be used either to fit a phase-type distribution to a sample (which may contain censored observations), or to make a phase-type approximation of another continuous distribution. The fitting procedure consists of an iterative estimation of the parameters of the phase-type distribution, using an *EM-algorithm*. The program is an implementation of the EM-algorithm presented in [25, 5]. The EMpht-program is a C-program. It is complemented by a Matlab program, PHplot, for graphical display of the fitted phase-type distribution.

6.1 The EM Algorithm

All data augmentation algorithms, including the EM (expectation maximization) algorithm, are used to *locate the mode or modes* of the likelihood function or of the posterior density. (The quantity $p(\theta|Y)$, which describes what is known about given data Y, is called the *posterior density of* θ). They share a common approach to problems: rather than performing a complicated maximization or simulation, one augments the observed data with latent data to simplify the computations in the analysis. This latent data can be the "missing" data, parameter values or values of sufficient statistics. Thus, in the EM algorithm, one augments the observed data with latent data such that one complicated maximization is replaced by an iterative series of simple maximizations. The principle of data augmentation can then be stated as follows: augment the observed data Y with latent data Z so that the *augmented posterior distribution* $p(\theta|Y, Z)$ is "simple". Making use of this simplicity in maximizing/marginalizing/calculating/sampling the *observed posterior* $p(\theta|Y)$.

More specifically, *the EM algorithm is an iterative method for locating posterior modes*. Each iteration consists of two steps: the E-step (expectation step) and the M-step (maximization step). Formally, let θ^i denote the current guess to the mode of the observed posterior $p(\theta|Y)$; let $p(\theta|Y,Z)$ denote the augmented posterior, i.e. the posterior of the augmented data; and let $p(Z|\theta^i,Y)$ denote the conditional predictive distribution of the latent data Z, conditional on the current guess to the posterior mode. In the most general setting, the E-step consists of computing *the expected log likelihood*

$$Q(\theta|\theta^i,Y) = \int_Z \log[p(\theta|Z,Y)]p(Z|\theta^i,Y)dZ, \tag{6.1}$$

i.e. the expectation of $log[p(\theta|Z,Y)]$ with respect to $p(Z|\theta^i,Y)$. In the M-step the Q function is maximized with respect to θ to obtain θ^{i+1}. The algorithm is iterated until $\|\theta^{i+1}-\theta^i\|$ or $|Q(\theta^{i+1}|\theta^i,Y)-Q(\theta^i|\theta^i,Y)|$ is sufficiently small.

6.2 An Example

An observation y of the time to absorption can be regarded as an incomplete observation of the Markov process $\{X(t), t \geq 0\}$. It is incomplete in the sense that it only tells us when the process hits Δ, and does not provide any information about how it got there, where it started, which of the states it visited and for how long. For example, consider the Hyperexponential distribution with two phases in parallel, as given in Figure 6.1, where q is the probability to start in state one with rate λ_1 and $1-q$ is the probability to start in state two with rate λ_2.

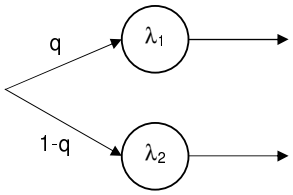


Figure 6.1: An example of Hyperexponential distribution.

6.2.1 The general approach of EM-algorithm

Let $y_i, i = 1,2,3$ represents the observed data, and the augmented data is given by $x_i, i = 1,2,3$ that represents to which state the process is jumped,

either to state one or to state number two. Then, the $z_i = (y_i, x_i)$ be the complete information of an underlying Markov process. The vector of parameters, for which maximum likelihood estimator have to be found, is $\boldsymbol{\theta} = (q, \lambda_1, \lambda_2)$.

The goal is to maximize the likelihood function $L(\boldsymbol{\theta}|\boldsymbol{Y})$ or the observed posterior $f(\boldsymbol{\theta}|\boldsymbol{Y})$. Notice that

$$log[f(\boldsymbol{\theta}|\boldsymbol{Y})] = log[f(\boldsymbol{\theta}|\boldsymbol{Y}, \boldsymbol{X})] - log[f(\boldsymbol{X}|\boldsymbol{\theta}, \boldsymbol{Y})]. \tag{6.2}$$

With an initial guess $\boldsymbol{\theta_0}$ (to start the iterations), the expected log likelihood is

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta_0}, \boldsymbol{Y}) = \sum_{\boldsymbol{X}} [\log[f(\boldsymbol{Y}, \boldsymbol{X}|\boldsymbol{\theta})] f(\boldsymbol{X}|\boldsymbol{Y}, \boldsymbol{\theta_0}), \tag{6.3}$$

where according to Bayes' theorem

$$f(\boldsymbol{X}|\boldsymbol{\theta_0}, \boldsymbol{Y}) = \frac{f(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{\theta_0}) P(\boldsymbol{X}|\boldsymbol{\theta_0})}{f(\boldsymbol{Y}|\boldsymbol{\theta_0})}, \tag{6.4}$$

and

$$f(\boldsymbol{Y}, \boldsymbol{X}|\boldsymbol{\theta}) = f(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{\theta}) P(\boldsymbol{X}|\boldsymbol{\theta}). \tag{6.5}$$

When three observed data (y_1, y_2, y_3) are given, there are eight possible cases that may happen, and the computation of $f(\boldsymbol{Y}, \boldsymbol{X}|\boldsymbol{\theta})$ for all cases is given:

$$\begin{aligned}
f(\boldsymbol{Y}, X_1 = 1, X_2 = 1, X_3 = 1|\boldsymbol{\theta}) &= (q\lambda_1)^3 e^{-\lambda_1 \sum_{i=1}^3 y_i} \\
f(\boldsymbol{Y}, X_1 = 0, X_2 = 0, X_3 = 0|\boldsymbol{\theta}) &= ((1-q)\lambda_2)^3 e^{-\lambda_2 \sum_{i=1}^3 y_i} \\
f(\boldsymbol{Y}, X_1 = 0, X_2 = 0, X_3 = 1|\boldsymbol{\theta}) &= q(1-q)^2 \lambda_1 \lambda_2^2 e^{-\lambda_1 y_3} e^{-\lambda_2 (y_1 + y_2)} \\
f(\boldsymbol{Y}, X_1 = 1, X_2 = 0, X_3 = 0|\boldsymbol{\theta}) &= q(1-q)^2 \lambda_1 \lambda_2^2 e^{-\lambda_1 y_1} e^{-\lambda_2 (y_2 + y_3)} \\
f(\boldsymbol{Y}, X_1 = 1, X_2 = 1, X_3 = 0|\boldsymbol{\theta}) &= q^2(1-q) \lambda_1^2 \lambda_2 e^{-\lambda_1 (y_1 + y_2)} e^{-\lambda_2 y_3} \\
f(\boldsymbol{Y}, X_1 = 1, X_2 = 0, X_3 = 1|\boldsymbol{\theta}) &= q^2(1-q) \lambda_1^2 \lambda_2 e^{-\lambda_1 (y_1 + y_3)} e^{-\lambda_2 y_2} \\
f(\boldsymbol{Y}, X_1 = 0, X_2 = 1, X_3 = 1|\boldsymbol{\theta}) &= q^2(1-q) \lambda_1^2 \lambda_2 e^{-\lambda_1 (y_2 + y_3)} e^{-\lambda_2 y_1} \\
f(\boldsymbol{Y}, X_1 = 0, X_2 = 1, X_3 = 0|\boldsymbol{\theta}) &= q(1-q)^2 \lambda_1 \lambda_2^2 e^{-\lambda_1 y_2} e^{-\lambda_2 (y_1 + y_3)}
\end{aligned} \tag{6.6}$$

Then, the expected log likelihood function is

$$
\begin{aligned}
Q \quad & (\boldsymbol{\theta}|\boldsymbol{\theta_0}, \boldsymbol{Y}) = (3\log[q\lambda_1] - \lambda_1 \sum_{i=1}^{3} y_i) f(\boldsymbol{X} = (1,1,1)|\boldsymbol{Y}, \boldsymbol{\theta_0}) + \\
+ \quad & (3\log[(1-q)\lambda_2] - \lambda_2 \sum_{i=1}^{3} y_i) f(\boldsymbol{X} = (0,0,0)|\boldsymbol{Y}, \boldsymbol{\theta_0}) + \\
+ \quad & (2\log\lambda_2 - \lambda_2(y_1+y_2) + \log\lambda_1 - \lambda_1 y_3 + 2\log(1-q) + \log q) f(\boldsymbol{X} = (0,0,1)|\boldsymbol{Y}, \boldsymbol{\theta_0}) + \\
+ \quad & (\log\lambda_1 + 2\log\lambda_2 - \lambda_1 y_1 - \lambda_2(y_2+y_3) + 2\log(1-q) + \log q) f(\boldsymbol{X} = (1,0,0)|\boldsymbol{Y}, \boldsymbol{\theta_0}) + \\
+ \quad & (2\log\lambda_1 + \log\lambda_2 - \lambda_1(y_1+y_2) - \lambda_2 y_3 + 2\log q + \log(1-q)) f(\boldsymbol{X} = (1,1,0)|\boldsymbol{Y}, \boldsymbol{\theta_0}) + \\
+ \quad & (2\log\lambda_1 + \log\lambda_2 - \lambda_1(y_1+y_3) - \lambda_2 y_2 + 2\log q + \log(1-q)) f(\boldsymbol{X} = (1,0,1)|\boldsymbol{Y}, \boldsymbol{\theta_0}) + \\
+ \quad & (2\log\lambda_1 + \log\lambda_2 - \lambda_1(y_2+y_3) - \lambda_2 y_1 + 2\log q + \log(1-q)) f(\boldsymbol{X} = (0,1,1)|\boldsymbol{Y}, \boldsymbol{\theta_0}) + \\
+ \quad & (\log\lambda_1 + 2\log\lambda_2 - \lambda_1 y_2 - \lambda_2(y_1+y_3) + \log q + 2\log(1-q)) f(\boldsymbol{X} = (0,1,0)|\boldsymbol{Y}, \boldsymbol{\theta_0}).
\end{aligned}
\tag{6.7}
$$

The M-step consists of maximizing $Q(\boldsymbol{\theta}|\boldsymbol{\theta_0}, \boldsymbol{Y})$. By differentiating the Q function with respect to $\boldsymbol{\theta}$ in order to obtain the 3×1 vector $\partial Q(\boldsymbol{\theta}|\boldsymbol{\theta_0}, \boldsymbol{Y})/\partial\boldsymbol{\theta}$:

$$
\begin{aligned}
\frac{\partial Q(\boldsymbol{\theta}|\boldsymbol{\theta_0}, \boldsymbol{Y})}{\partial q} \quad = \quad & \frac{3}{q} f(\boldsymbol{X} = (1,1,1)|\boldsymbol{Y}, \boldsymbol{\theta_0}) + \\
+ \quad & \left(-\frac{3}{1-q}\right) f(\boldsymbol{X} = (0,0,0)|\boldsymbol{Y}, \boldsymbol{\theta_0}) + \\
+ \quad & \left(\frac{1}{q} - \frac{2}{1-q}\right) f(\boldsymbol{X} = (0,0,1)|\boldsymbol{Y}, \boldsymbol{\theta_0}) + \\
+ \quad & \left(\frac{1}{q} - \frac{2}{1-q}\right) f(\boldsymbol{X} = (1,0,0)|\boldsymbol{Y}, \boldsymbol{\theta_0}) + \\
+ \quad & \left(\frac{2}{q} - \frac{1}{1-q}\right) f(\boldsymbol{X} = (1,1,0)|\boldsymbol{Y}, \boldsymbol{\theta_0}) + \\
+ \quad & \left(\frac{2}{q} - \frac{1}{1-q}\right) f(\boldsymbol{X} = (1,0,1)|\boldsymbol{Y}, \boldsymbol{\theta_0}) + \\
+ \quad & \left(\frac{2}{q} - \frac{1}{1-q}\right) f(\boldsymbol{X} = (0,1,1)|\boldsymbol{Y}, \boldsymbol{\theta_0}) + \\
+ \quad & \left(\frac{1}{q} - \frac{2}{1-q}\right) f(\boldsymbol{X} = (0,1,0)|\boldsymbol{Y}, \boldsymbol{\theta_0}),
\end{aligned}
\tag{6.8}
$$

$$
\begin{aligned}
\frac{\partial Q(\boldsymbol{\theta}|\boldsymbol{\theta_0}, \boldsymbol{Y})}{\partial \lambda_1} &= \left(\frac{3}{\lambda_1} - \sum_{i=1}^{3} y_i\right) f(\boldsymbol{X} = (1,1,1)|\boldsymbol{Y}, \boldsymbol{\theta_0}) + \\
&+ \left(\frac{1}{\lambda_1} - y_3\right) f(\boldsymbol{X} = (0,0,1)|\boldsymbol{Y}, \boldsymbol{\theta_0}) + \\
&+ \left(\frac{1}{\lambda_1} - y_1\right) f(\boldsymbol{X} = (1,0,0)|\boldsymbol{Y}, \boldsymbol{\theta_0}) + \\
&+ \left(\frac{2}{\lambda_1} - (y_1 + y_2)\right) f(\boldsymbol{X} = (1,1,0)|\boldsymbol{Y}, \boldsymbol{\theta_0}) + \\
&+ \left(\frac{2}{\lambda_1} - (y_1 + y_3)\right) f(\boldsymbol{X} = (1,0,1)|\boldsymbol{Y}, \boldsymbol{\theta_0}) + \\
&+ \left(\frac{2}{\lambda_1} - (y_2 + y_3)\right) f(\boldsymbol{X} = (0,1,1)|\boldsymbol{Y}, \boldsymbol{\theta_0}) + \\
&+ \left(\frac{1}{\lambda_1} - y_2\right) f(\boldsymbol{X} = (0,1,0)|\boldsymbol{Y}, \boldsymbol{\theta_0}), \qquad (6.9)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial Q(\boldsymbol{\theta}|\boldsymbol{\theta_0}, \boldsymbol{Y})}{\partial \lambda_2} &= \left(\frac{3}{\lambda_2} - \sum_{i=1}^{3} y_i\right) f(\boldsymbol{X} = (0,0,0)|\boldsymbol{Y}, \boldsymbol{\theta_0}) + \\
&+ \left(\frac{2}{\lambda_2} - (y_1 + y_2)\right) f(\boldsymbol{X} = (0,0,1)|\boldsymbol{Y}, \boldsymbol{\theta_0}) + \\
&+ \left(\frac{2}{\lambda_2} - (y_2 + y_3)\right) f(\boldsymbol{X} = (1,0,0)|\boldsymbol{Y}, \boldsymbol{\theta_0}) + \\
&+ \left(\frac{1}{\lambda_2} - y_3\right) f(\boldsymbol{X} = (1,1,0)|\boldsymbol{Y}, \boldsymbol{\theta_0}) + \\
&+ \left(\frac{1}{\lambda_2} - y_2\right) f(\boldsymbol{X} = (1,0,1)|\boldsymbol{Y}, \boldsymbol{\theta_0}) + \\
&+ \left(\frac{1}{\lambda_2} - y_1\right) f(\boldsymbol{X} = (0,1,1)|\boldsymbol{Y}, \boldsymbol{\theta_0}) + \\
&+ \left(\frac{2}{\lambda_2} - (y_1 + y_3)\right) f(\boldsymbol{X} = (0,1,0)|\boldsymbol{Y}, \boldsymbol{\theta_0}). \qquad (6.10)
\end{aligned}
$$

Then, by setting the three simultaneous expected log likelihood equation to 0 and solving for $\boldsymbol{\theta}$, one derives the maximum likelihood estimate $\hat{\boldsymbol{\theta}}$, based on the sample $\sum y_i, i = 1, 2, 3$. The $j+1$th step consists of calculating the expected log likelihood 6.3, with $\boldsymbol{\theta_0}$ replaced by $\hat{\boldsymbol{\theta}}_{\boldsymbol{j}}$, and then maximizing it.

Suppose that $\boldsymbol{Y} = (y_1, y_2, y_3) = (1, 10, 10)$. Starting at $q = 0.5$, $\lambda_1 = 0.5$,

$\lambda_2 = 0.9$, after 20 iterations the EM algorithm converges to

$$\hat{\mathbf{q}}' = \begin{pmatrix} 0.970101 \\ 0.029899 \end{pmatrix} \quad \hat{\mathbf{R}} = \begin{bmatrix} -0.139186 & 0.000000 \\ 0.000000 & -0.990838 \end{bmatrix}$$

with log-likelihood = -8.840220, that is, $\hat{q} = 0.97, \hat{\lambda}_1 = 0.14, \hat{\lambda}_2 = 0.99$. And after 500 iterations the EM algorithm converges to $\hat{q} = 1, \hat{\lambda}_1 = 0.14, \hat{\lambda}_2 = 0.99$ with log-likelihood = -8.837730. The derived phase-type structure includes only one phase with rate $\hat{\lambda} = 0.14$. The corresponding mean is 7, and there is also the average of three observed observations $(y_1, y_2, y_3) = (1, 10, 10)$. However, with observed observations $(y_1, y_2, y_3) = (1, 10, 10)$, we expect to derive $q = 0.33, \lambda_1 = 1, \lambda_2 = 0.1$. The results of EM algorithm are not what it is expected. But when we tried to apply the same procedure to observations $(y_1, y_2, y_3) = (1, 100, 100)$, the results of EM algorithm were exactly what we are expected. For example, with starting values $q = 0.5$, $\lambda_1 = 0.5$, $\lambda_2 = 0.9$, after 50 iterations the EM algorithm converges to

$$\hat{\mathbf{q}}' = \begin{pmatrix} 0.685633 \\ 0.314367 \end{pmatrix} \quad \hat{\mathbf{R}} = \begin{bmatrix} -0.010282 & 0.000000 \\ 0.000000 & -1.000003 \end{bmatrix}$$

with log-likelihood = -14.064556, that is, $\hat{q} = 0.69, \hat{\lambda}_1 = 0.01, \hat{\lambda}_2 = 1$. For comparison, the expected parameter values are $q = 0.67$, $\lambda_1 = 0.01$, $\lambda_2 = 1$. So, with observations (1,10,10) the EM algorithm converges to structure with one phase that is the average of these observations, apparently, because the distance between 1 and 10 is not sufficient.

6.2.2 The EM-algorithm for the exponential family

The observed sample $y_i, i = 1, 2, 3$ of the time to absorption is an incomplete information of the Markov process $X(t)$. Suppose that we have 3 independent replications of the process, $X_t^{[1]}, X_t^{[2]}, X_t^{[3]}$. Then, the density of complete sample $x_i, i = 1, 2, 3$ can be written in the form

$$f(\boldsymbol{x}; q, \boldsymbol{R}) = q^{B_1} e^{-\lambda_1 Z_1} \lambda_1^{B_1} \cdot (1-q)^{B_2} e^{-\lambda_2 Z_2} \lambda_2^{B_2}$$

where

B_i = the number of Markov processes starting in state $i, i = 1, 2$.

Z_i = the total time spent in state $i, i = 1, 2$ of all Markov processes.

The density $f(\boldsymbol{x}; q, \lambda_1, \lambda_2)$ is a member of a multi-parameter exponential family with sufficient statistic

$$\boldsymbol{S} = ((B_i)_{i=1,2}, (Z_i)_{i=1,2}).$$

E-step: The calculation of the conditional expectation of the sufficient statistic $\boldsymbol{S}$, given the observed sample $y_i, i = 1, 2, 3$ and the current estimates of $(q, \lambda_1, \lambda_2)$.
Let $B_1 = \sum I_j, j = 1, 2, 3$ where I_j is an indicator that the process starting in state 1. Then $B_1 + B_2 = 3$.

$$E_{(\hat{q},\hat{\lambda}_1,\hat{\lambda}_2)}(B_1|y_1, y_2, y_3) = \sum_{j=1}^{3} E(I_j|y_j) = \sum_{j=1}^{3} P_{(\hat{q},\hat{\lambda}_1,\hat{\lambda}_2)}(\text{the process visits state } 1|y_j),$$

where

$$P_{(\hat{q},\hat{\lambda}_1,\hat{\lambda}_2)}(\text{the process visits state } 1|y_1) = \frac{\hat{q}\hat{\lambda}_1 e^{-\hat{\lambda}_1 y_1}}{\hat{q}\hat{\lambda}_1 e^{-\hat{\lambda}_1 y_1} + (1-\hat{q})\hat{\lambda}_2 e^{-\hat{\lambda}_2 y_1}}.$$

$$E_{(\hat{q},\hat{\lambda}_1,\hat{\lambda}_2)}(Z_1|y_1, y_2, y_3) = \sum_{j=1}^{3} y_j \cdot P_{(\hat{q},\hat{\lambda}_1,\hat{\lambda}_2)}(\text{the process visits state } 1|y_j).$$

M-step: The new estimates are given by

$$\hat{q} = \frac{B_1}{3}, \qquad \hat{\lambda}_1 = \frac{B_1}{Z_1}, \qquad \hat{\lambda}_2 = \frac{B_2}{Z_2}.$$

By implementing this approach of the EM-algorithm to the same data sets as in the previous section 6.2.1, the results are seen to coincide.

6.3 The closure properties of the EM-algorithm

The following is the key property of the sequence $\{\hat{\theta}_{(j)}\}$:

- The EM algorithm increases the posterior $p(\theta|Y)$ at each iteration, that is, $p(\theta^{i+1}|Y) \geq p(\theta^i|Y)$, with equality holding if and only if $Q(\theta^{i+1}|\theta^i, Y) = Q(\theta^i|\theta^i, Y)$.

  In addition, the following results are available:

- The EM algorithm may not converge to the global maximum, when there are multiple stationary points (local maxima or saddle points) of $p(\theta|Y)$.

- The EM algorithm converges at a linear rate, with the rate depending on the proportion of information about θ in $p(\theta|Y)$ that is observed. This implies that the convergence can be quite slow if a large proportion of the data is missing.

- By implementing the EM algorithm to fit PH-distributions, the mean of the fitted phase-type distribution is the same as the mean of the sample (or the theoretical mean if it fits another distribution).

6.4 Fitting continuous distributions

For the approximation of a theoretical density by a phase-type density, it is considered the infinite analogue of maximum likelihood estimation: minimization of the information divergence (relative entropy or Kullback-Leibler information). Computationally, this turns out to be almost equivalent to implementing the EM algorithm for a sample. Let $f(\cdot; \boldsymbol{q}, \boldsymbol{R})$ be a density of a phase-type distribution and $h(\cdot)$ the density of the distribution to be approximated. Fitting $f(\cdot; \boldsymbol{q}, \boldsymbol{R})$ to $h(\cdot)$ means minimizing the information divergence, then the information divergence of f with respect to h is

$$\int \log \left[\frac{h(y)}{f(y; \boldsymbol{q}, \boldsymbol{R})} \right] h(y) dy, \tag{6.11}$$

which is equivalent to maximizing

$$\int log[f(y; \boldsymbol{q}, \boldsymbol{R})] h(y) dy. \tag{6.12}$$

Details and further theoretical motivations are given in [5].

In each iteration of the EM-algorithm, in the EMpht-program, when fitting phase-type distribution to the sample or to another distribution, the new parameter estimates are calculated by solving of homogeneous linear differential equations (of dimension $k^2 + 2k$, when fitting general phase-type structure). This is done numerically with Runge-Kutta method of fourth order [26, p. 12].

6.5 The PHplot-program

PHplot is a Matlab program for graphical display of the fitted phase-type distribution [26, p. 13]. It is modified here for more convenient way to obtain the graphical display (postscript files) and the results of sample characteristics with fitted results, when one fits PH-distribution to sample (text-file). The graphical display contains the fitted distribution function and the fitted survival function together with empirical distribution (or survival) function (the Kaplan-Meier estimator is used for censored observations), the fitted density and hazard functions together with appropriate empirical functions

estimated with non-parametrical methods derived in S-PLUS (described in 4.1, 4.2). The text-file contains sample characteristics of the data to be fit, such as the total number of observations, the minimum and maximum, the sample mean, the median, the standard deviation and the coefficient of variation (CV). Given a data sample $\{X_i\}_{i=1}^N$, the main sample statistics are:

- *Average:* $\hat{m} = \sum_{i=1}^N X_i/N$. (Function `mean` in Matlab.)
- *Standard deviation:* $\hat{\sigma} = \sqrt{\sum_{i=1}^N (X_i - \hat{m})^2/(N-1)}$. (Function `std` in Matlab.)
- *Coefficient of Variation(CV):* $\hat{c} = \hat{\sigma}/\hat{m}$.

It also gives the number of transient states fitted, the transition probability matrix, the vector of the probability of starting in the transient states $[1, \ldots, K]$, the absorption probability vector, the vector of a length of time spends in states $[1, \ldots, K]$ in the seconds and the minutes, the fitted mean and standard deviation of PH-distribution and, consequently, the coefficient of variation. The fitted quantities derives from $(\hat{\boldsymbol{q}}, \hat{\boldsymbol{R}})$, which estimated with EMpht-program. For example, the transition probability matrix is calculated by:

$$P_{jk} = \begin{cases} -R_{jk}/R_{jj} & \text{if } j \neq k, 1 \leq k \leq K \\ 0 & \text{if } j = k, 1 \leq k \leq K \end{cases} \tag{6.13}$$

the absorption probability vector is:

$$P_{k\triangle} = -r_k/R_{kk}, \quad \text{for } 1 \leq k \leq K, \tag{6.14}$$

and the length of time spends in states $[1, \ldots, K]$ in the seconds is:

$$m_k = -1/R_{kk}, \quad \text{for } 1 \leq k \leq K. \tag{6.15}$$

Chapter 7

Graphical methods and Goodness-of-fit tests based on the EDF, applied to Service times

7.1 Heuristic stopping rule for adding phases

Based on the data, we have an empirical CDF, estimated with non-parametric methods (see Chapter 4). Usually, as the number of parameters in the model is larger, the corresponding fitted model is better. Thus, in order to get a perfect fit of PH-distribution to the empirical CDF we need infinitely many phases. But we want to fit PH-model to the true distribution, not to the empirical CDF. The true distribution is known to be in $\frac{1}{\sqrt{n}}$ neighborhood. That is, we know that the true distribution is up to $\pm\frac{1}{\sqrt{n}}$ resolution. Therefore, we keep adding phases until we get the fit under this resolution. More observations lead us to add more phases. This heuristic guides us in Chapter 8. The next section provides the foundation to these heuristical arguments.

7.2 Construction of simultaneous confidence interval to the CDF

Confidence interval is an important graphical method to convey some idea of the probable accuracy of the estimator, by specifying a random set covering the true parameter value with some specified (high)probability. For construction a simultaneous confidence interval for cumulative distribution function

F (CDF) with non-parametric methods, one can use Kolmogorov-Smirnov (K-S) statistic [27] (denoted as D, formula (7.14) below). Given a sample of n iid random variables $\{X_1, X_2, \ldots, X_n\}$ and significance level γ, the upper γ-percent point of the distribution of D – $D_{n,\gamma}$, can be found such as

$$P\{D \geq D_{n,\gamma}\} \leq \gamma. \tag{7.1}$$

Let

$$L_n(x) = \max\{\frac{1}{n} - D_{n,\gamma}, 0\} \tag{7.2}$$

and

$$U_n(x) = \min\{\frac{1}{n} + D_{n,\gamma}, 1\}. \tag{7.3}$$

Then the region between $L_n(x)$ and $U_n(x)$ can be used as a confidence band for $F(x)$ with associated confidence coefficient $1 - \gamma$.
The exact distribution of D for selected values of γ and $n = 1, ..., 40$, and approximation for $n > 40$ is given in [27], p.661. Table 7.2 gives the values of $D_{n,\gamma}$ for some selected γ and for $n > 40$.

Table 7.1: Critical Values of the Kolmogorov-Smirnov One sample Test Statistics.

Significance level γ	0.20	0.10	0.05	0.02	0.01
Approximation for $n > 40$	$\frac{1.07}{\sqrt{n}}$	$\frac{1.22}{\sqrt{n}}$	$\frac{1.36}{\sqrt{n}}$	$\frac{1.52}{\sqrt{n}}$	$\frac{1.63}{\sqrt{n}}$

Table 7.2 presents values computed using the theorem, derived by Kolmogorov, about the large-sample distribution of D. It states that for any continuous distribution function F,

$$\lim_{n\to\infty} P\{D \leq zn^{-1/2}\} = L(z), \quad z \geq 0, \tag{7.4}$$

where

$$L(z) = 1 - 2\sum_{i=1}^{\infty}(-1)^{i-1}e^{-2i^2z^2}. \tag{7.5}$$

From (7.1),(7.4) and (7.5) followed

$$2\sum_{i=1}^{\infty}(-1)^{i-1}e^{-2i^2z^2} \leq \gamma. \tag{7.6}$$

For each specific value of γ, the value of z can be found such that (7.6) holds. Then

$$D_{n,\gamma} = \frac{z}{\sqrt{n}} \tag{7.7}$$

is computed to get the confidence interval $[L_n(x),\ U_n(x)]$ for $F(x)$. For example, for significance level $\gamma = 0.05$ the value of z is 1.36.

7.3 EDF tests

The confidence interval is the subjective method which determines whether the assumed distribution fitting the data is based on visual examination rather than statistical test.

A statistical test involves calculation of a test statistic from the data and the probability of obtaining the statistic if the correct distribution is chosen. Statistical tests that measure the discrepancy between an empirical cumulative distribution function (CDF) and a hypothesized CDF F are the EDF tests [10, 7, 14]. If the probability of obtaining the calculated statistic is very low, we conclude that the assumed distribution does not provide an adequate fit to the data. This procedure allows to reject an inadequate distribution but never allows to prove that the distribution is correct. It gives a probabilistic statement about the assumed distribution. The outcome of a statistical test of hypothesis depends on the amount of data available – n; the more data there is, the better chances of rejecting an inadequate distribution.

Let $F(t)$ be the underlying distribution from which the sample of our data is taken. Then, the general null hypothesis is

$$H_0 : F(t) = F_0(t), \tag{7.8}$$

where $F_0(t)$ is a specific distribution.

The two famous goodness-of-fit tests, Kolmogorov-Smirnov (K-S) and Anderson-Darling (A-D), are implemented in this research work. According to E. Chlebus [7], the A-D test is more powerful then the K-S test, nevertheless the K-S test is presented in order to compare between them.
Given an observed sample $\{x_1, x_2, \ldots, x_n\}$ of size n that is sorted in ascending order the supremum D^* and quadratic A^2 statistics associated with the K-S and A-D tests, respectively, are given by

$$D^* = D(\sqrt{n} + 0.12 + \frac{0.11}{\sqrt{n}}) \tag{7.9}$$

and

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)[\ln z_i + \ln(1 - z_{n+1-i})], \tag{7.10}$$

where

$$z_i = F(x_i), \tag{7.11}$$

$$D^+ = \max_{1 \le i \le n} \{\frac{1}{n} - z_i\}, \tag{7.12}$$

$$D^- = \max_{1 \le i \le n} \{z_i - \frac{i-1}{n}\} \tag{7.13}$$

and

$$D = \max\{D^+, D^-\}. \tag{7.14}$$

The hypothesis that the observed sample $\{x_1, x_2, \ldots, x_n\}$ comes from the assumed probability distribution with CDF F is accepted with significance level γ if $D^*(A^2) \leq c_\gamma(a_\gamma)$, where $c_\gamma(a_\gamma)$ is the critical value of EDF test $D^*(A^2)$, respectively. Otherwise, the hypothesis is rejected. Smaller values of $D^*(A^2)$ indicate the better fit. These critical points are given in following table:

Table 7.2: Critical values $c_\gamma(a_\gamma)$ for the K-S (A-D) test statistic.

	Significance level γ							
	0.250	0.150	0.100	0.05	0.025	0.010	0.005	0.001
D^*	1.019	1.138	1.224	1.358	1.480	1.628	1.731	1.950
A^2	1.248	1.610	1.933	2.492	3.070	3.857	4.500	6.000

Null hypothesis, the distribution of service time is PH-distribution of a specific order ($k = 2, 3, 4, 5, 6$), is tested. According to the outcome of a statistical test, the PH-model with the least number of phases is selected, for which the H_0 - hypothesis is not rejected with appropriate significance level γ.

Chapter 8

Analysis, Modelling and Fitting PH-distributions to the call center data: Service times and Patience

When fitting PH-distributions to the data, we are faced with the problem of choosing the appropriate order of the PH-model. It is well known that a perfect fit requires choosing the highest possible dimension. Since the size of the data-set, we used, is large – 30,000-40,000, then theoretically we could estimate PH-model of order $k = 100$ (number of transient states), say. However, using PH-distributions for modelling, the more the phases the larger the state space, hence the less clear the model is. In addition, there is no interpretation to the phases and one could not give a physical meaning to the phase, if there are $k = 100$. Besides, it is impossible to fit such a model in a practical way. Therefore, we should estimate the model with the least number of parameters, in order to understand what is going on, and what is most appropriate to the data.

There are two issues considering the modelling of PH-distributions:

- complexity – how many phases are there.
- structure – how are the phases related.

The following are reasonable steps for selecting an appropriate model:

1. Find a PH-model with the least number of phases that fits into a confidence band around the empirical distribution function.

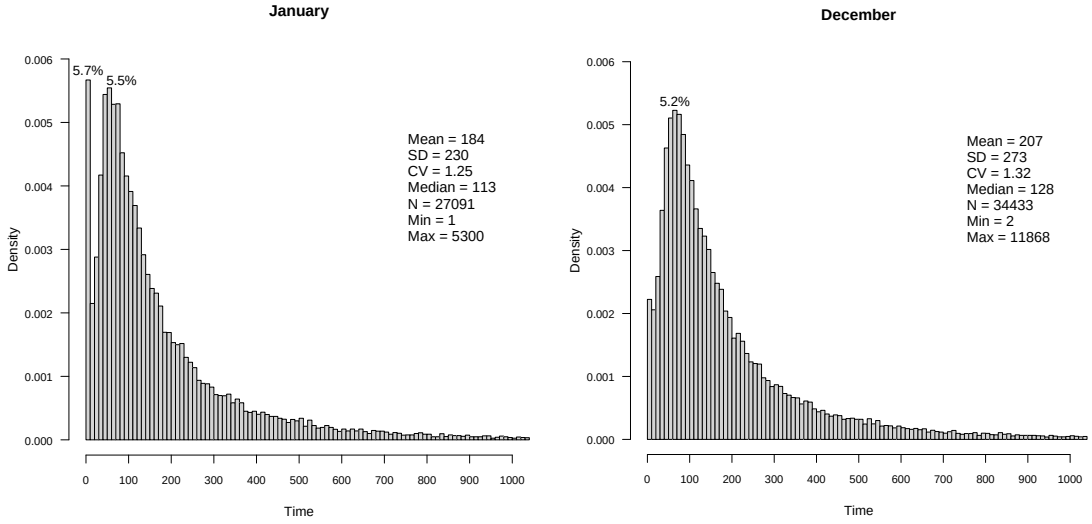
2. Given this number, presumably there will be more than one possibility, that is, structure. Then, we should analyze possible structures, and perhaps see if there are very different structures when we add, say a phase or two.

This way we can infer structure from the data, for example, what are phases of service, and what are phases of (im)patience.

8.1 Fitting PH-distribution to Service-time durations (service time > 0)

The histograms of service time of December 1999 and January 1999 are given in Figure 8.1. Service times for January exhibited a QUICK-HANG phenomena [20]: there is high percentage of calls with service times between 1 and 10 seconds, while service times for December are free of this problem. Service time shorter than 10 seconds is questionable. And indeed, questioning the manager of the call center revealed that short service times were caused by agents that simply *hung-up* on customers, to obtain extra rest-time. The phenomenon was only discovered in late, October 1999, after unreasonably many customers had complained about being disconnected. Corrective managerial action was taken and the problem was fixed towards the end of 1999.

Figure 8.1: Distribution of service time.



December service times for the four major service types: IN, NE, NW and PS, for low and high priorities, and overall, 34433 service transactions are analyzed. The distribution according to major service types is: IN – 3117, NE – 3725, NW – 2723, PS – 23811. The distribution according to priority types is: LOW – 22699, HIGH – 11734.

8.1.1 Overall service time – December

Figure 8.2 (p. 43) shows the service time histogram, where 98.07% of the calls with positive service time are in the range of the histogram, with the kernel density estimator of width = 30. Standard queueing theory often assumes that service times are "exponentially distributed". However, we see from the histogram below that service times from call center of "Anonymous Bank" does not have the shape of an exponential distribution.

Theory suggests that the bandwidth h should be proportional to $n^{-1/5}$. But this is an example in which this value, 0.12 ($n = 34433$), is too small. Considering a few bandwidths and plotting the resulting histograms, we chose $h = 10$ that gives a "smooth" histogram. The chosen value of bandwidth is very close to Sturges' rule [7], according to which $\lfloor 1 + \log_2 n \rfloor = \lfloor 1 + 3.322 \log_{10} n \rfloor$, where $\lfloor \circ \rfloor$ denotes the largest integer number not greater than the argument. Applying this rule to our data, the result is 16, meaning that the Sturges' rule does work here. For exponential distribution the optimal bandwidth is $E * \sqrt[3]{\frac{12}{n}}$, where E is an expectation. By applying this to our data ($n = 34433$) the result is 15 that almost coincide with Sturges' rule.

Figure 8.3 (p. 44) gives the actual hazard rates with superimposed super smoother hazard curve that smoothes the actual hazard rates up to 1000, when the hazard rates for non-failure times are zero (for correction the behavior of the tail). It was truncated at 1000 seconds because otherwise the shape of the hazard curve at the lower times being more flat than it has to be according to the actual hazard rates. Figure 8.3 shows the hazard rate of a unimodal shape of increase followed by decrease. Since the hazard rate of Hyperexponential distribution is always decreasing and the hazard rate of Generalized Erlang distribution is always increasing, the mode of the hazard rate of service time can be described by some kind of Generalized Erlang mixture distribution.

When starting to analyze what is the structure that fits the data, firstly it is considered to fit the PH-distribution of General structure for $k = 2, 3, 4, 5, 6$. Figure 8.4 (p. 45) compares the fitted PH-distributions of the density and the survival functions. It can be seen from the graphs that the fitted PH-distribution of order $k = 2$ cannot describe the mode of the distri-

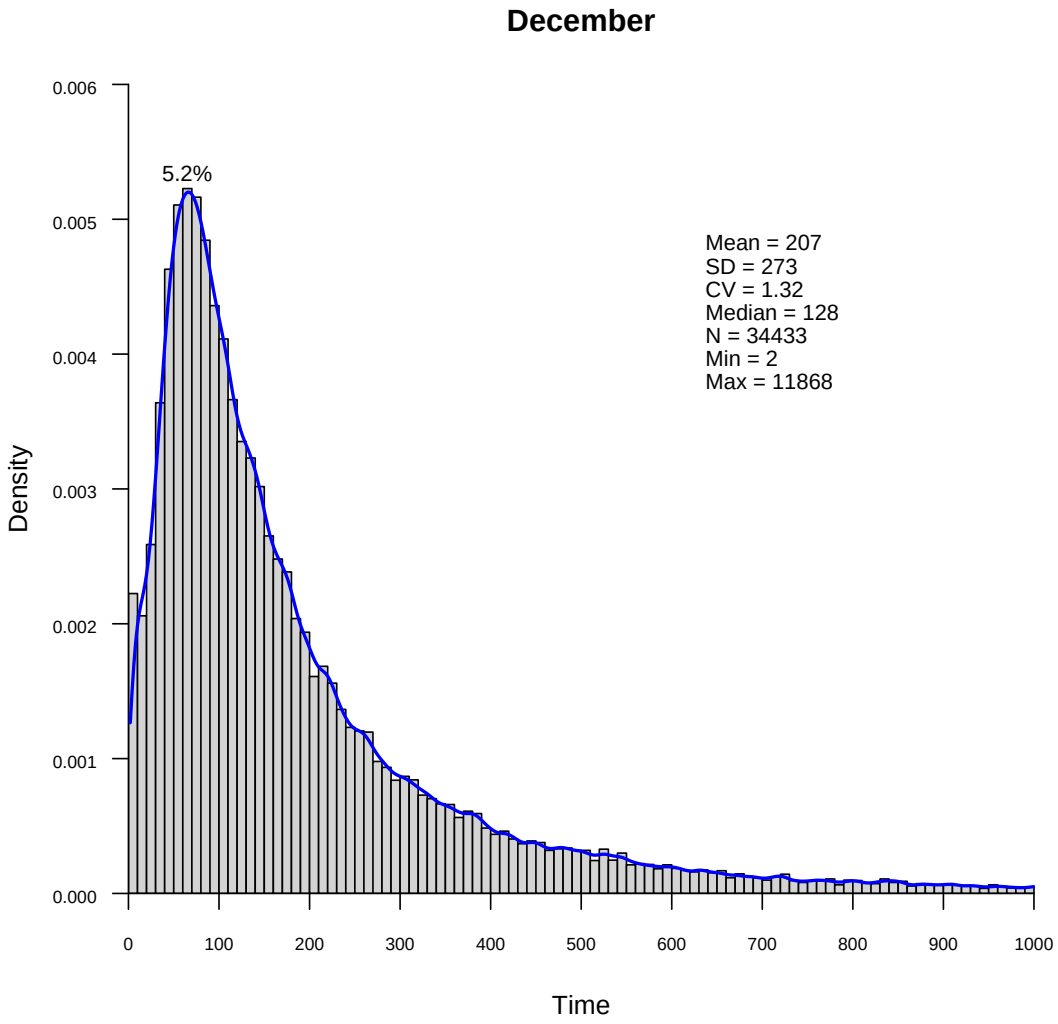


Figure 8.2: Distribution of overall service time - December. Superimposed on a histogram is the kernel density estimator with a Gaussian kernel of width = 30.

bution. The general structure of order $k = 2$ end up in the Hyperexponential distribution, because the probabilities to move from phase one to phase two and vice versa are negligible. The process starts with probability 0.94 in the state with a length time 175, and with probability 0.06 in the state with a length time 680. The state with a longer length time, 680, apparently related to customers with a longer service time. Indeed, according to the histogram of service time, 4.47% of the calls spend over 680 seconds (over 11 hours) in the service. Phase-type fits of a general structure of order $k = 3$ and $k = 4$,

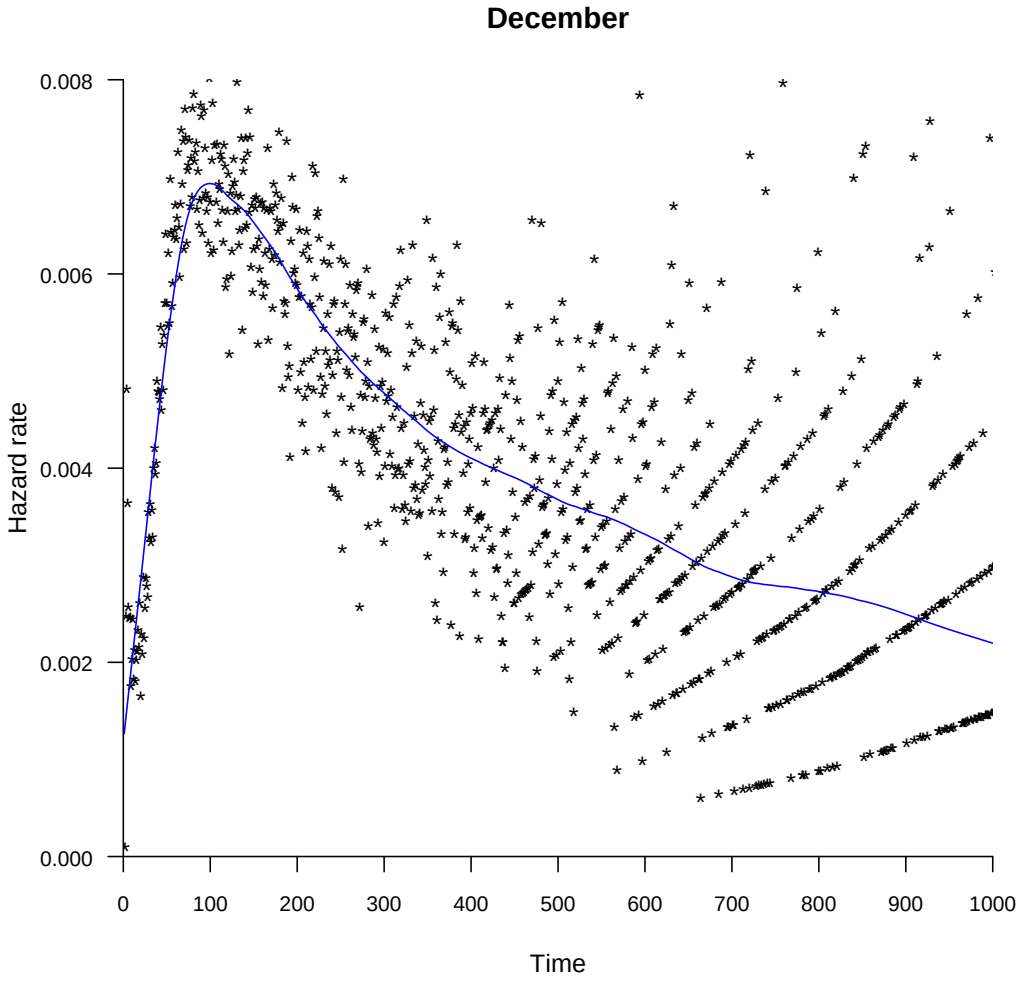


Figure 8.3: Hazard rate for overall service time - December. Superimposed on a hazard rates is a smoother hazard curve.

$k = 5$ and $k = 6$ almost coincide, therefore the phase-type fits of order $k = 4$ and $k = 6$ are not given in the graphs above. In general, as the order of PH-distribution is higher, the fits of PH-distributions is better, as seen from the graphs.

As it is demonstrated in section 5.5 (p. 25) there are several different representations of a PH-distribution. For example, fits of PH-distributions of order $k = 3$ to service time, from different initial values, gives the different setups of parameters of PH-distribution, $(\boldsymbol{q},\boldsymbol{R})$, but the maximum

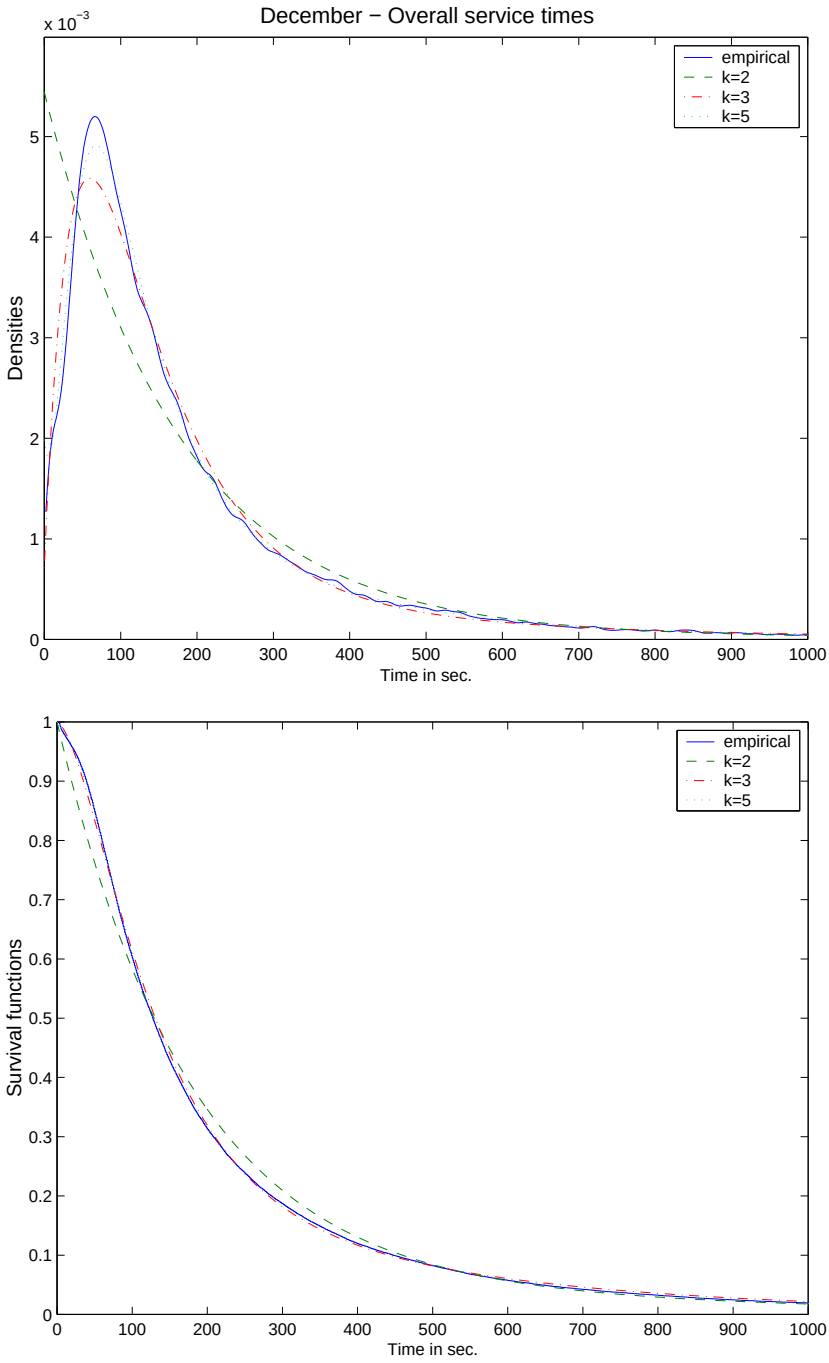
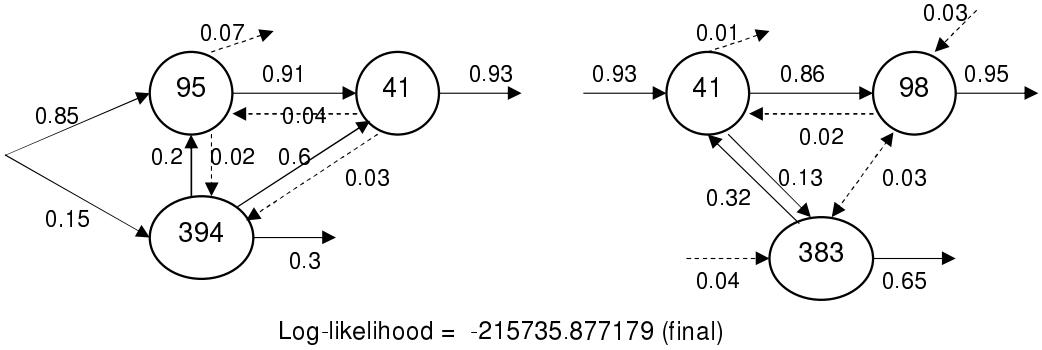


Figure 8.4: Phase-type fits to December service time by a general structure of order $k = 2 - -$, $k = 3 - \cdot -$, $k = 5 \cdots$. In the top plot, the solid line is the kernel density estimator, given as a comparison to the fitted densities. In the bottom plot, the solid line is the empirical survival function.

likelihood function is coincide. Figure 8.5 shows three different structures of PH-distribution of order $k = 3$ with the same log likelihood function. The

a)



b)

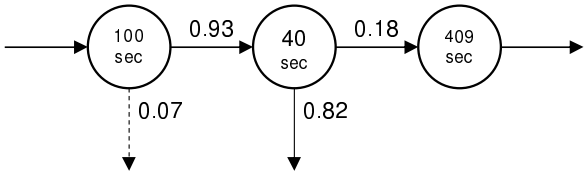
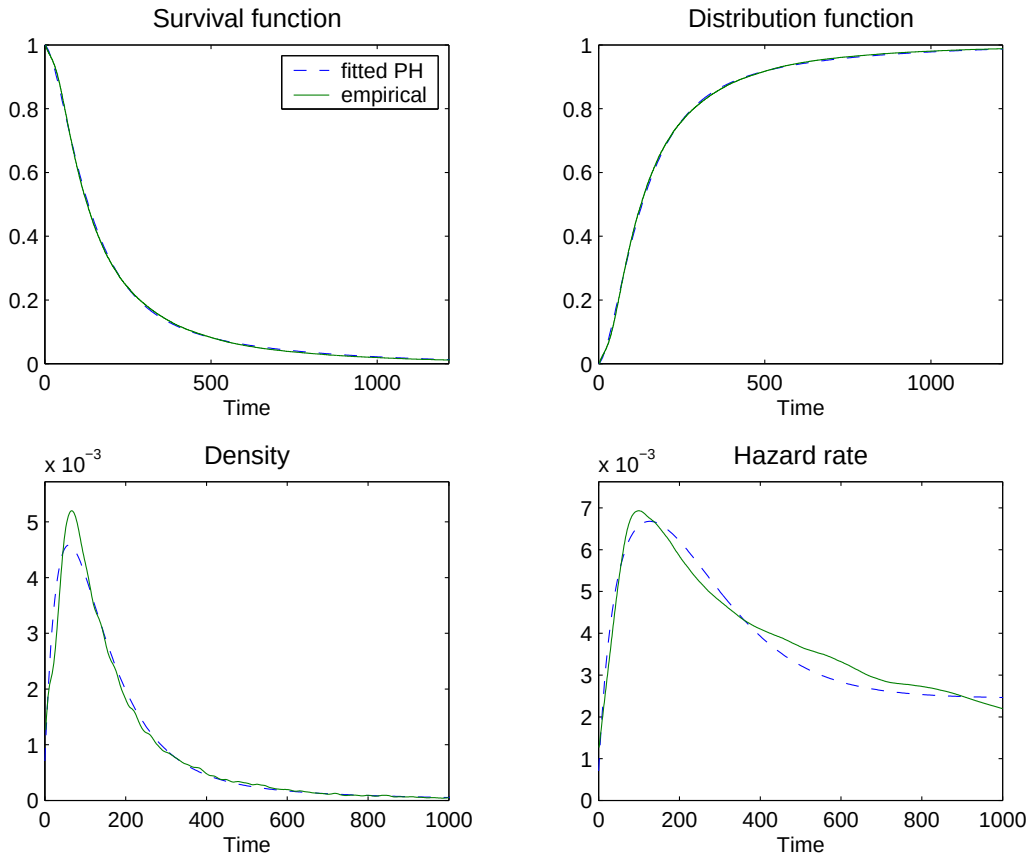


Figure 8.5: Two different structures of the same order, $k = 3$, of PH-type fit to the service time - December, starting with different initial values, Figure a) above. The fitted Coxian structure of the same order, Figure b) above.

corresponding densities are, when plotted, difficult to distinguish from each other. Then, Figure 8.6 (p. 47) demonstrates the fitted distribution, survival, density and hazard functions together with corresponding empirical functions for the PH-distribution of order $k = 3$ of the structure at left in Figure 8.5. In addition, Figure 8.5 b) demonstrates the fitted Coxian structure of order $k = 3$. It has the same log-likelihood function and, correspondingly, the same fitted mean and standard deviation.

When one looks at the two structures above by ignoring the small probabilities (the dashed arrows in Figure 8.5 a)), it can be seen that despite of different estimated set-up of the parameters at first sight, there are similar length time in the states. Moreover, these two structures can be simplified to the following, showed in Figure 8.7 (p. 48), with corresponding two different setups of parameters $(\mathbf{q}, \mathbf{R})$ and $(\boldsymbol{q}, \boldsymbol{R})'$:

Figure 8.6: PH-type fit of order $k = 3$ of general structure (dashed curve) with empirical functions (solid curve).



(1)

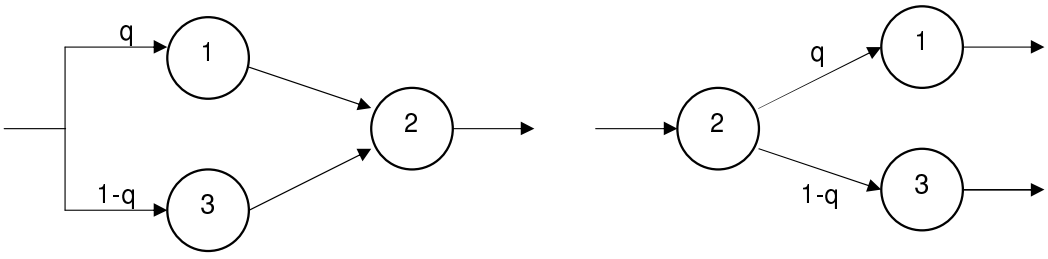
$$\mathbf{q} = \begin{pmatrix} q \\ 0 \\ 1-q \end{pmatrix} \quad \mathbf{R} = \begin{bmatrix} -\lambda_1 & \lambda_1 & 0 \\ 0 & -\lambda_2 & 0 \\ 0 & \lambda_3 & -\lambda_3 \end{bmatrix}$$

(2)

$$\mathbf{q}' = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \quad \mathbf{R}' = \begin{bmatrix} -\lambda_1 & 0 & 0 \\ q\lambda_2 & -\lambda_2 & (1-q)\lambda_2 \\ 0 & 0 & -\lambda_3 \end{bmatrix}$$

Let X_1, X_2, X_3 - three independent random variables exponentially distributed

Figure 8.7: An example of PH-distribution of third order represented by two different setups of parameters.



with parameters $\lambda_i, i = 1, 2, 3$, and an indicator

$$I = \begin{cases} 1 & \text{w.p. } q \\ 0 & \text{w.p. } 1-q \end{cases}$$

independent of $X_i, i = 1, 2, 3$. Then $Y = X_2 + IX_1 + (1 - I)X_3$ is time to absorption, and in both cases:

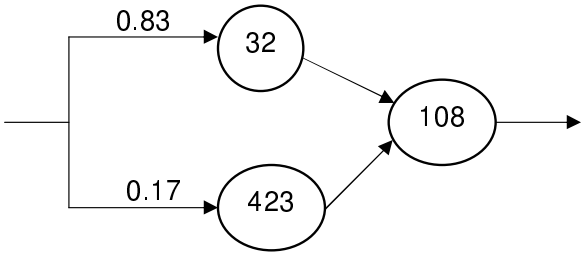
$$f_Y(y) = \frac{\lambda_1\lambda_2 q}{(\lambda_1 - \lambda_2)}\left(e^{-\lambda_2 y} - e^{-\lambda_1 y}\right) + \frac{\lambda_2\lambda_3(1-q)}{(\lambda_3 - \lambda_2)}\left(e^{-\lambda_2 y} - e^{-\lambda_3 y}\right), y > 0.$$

The PH-distribution of order $k = 3$ of the structure at left in Figure 8.7 is fitted to compare its results with the fitted PH-distribution of the same order of the general structure, given in figure 8.5. Figure 8.8 (p. 49) shows the derived specified structure with corresponding log-likelihood function and the graph of fitted PH-density functions of general and specified structures together with kernel density estimator. It is difficult to distinguish between the two structures, according to the graph of their survival functions. It can be seen, according to their fitted density functions that there is a little difference between them at the top of the mode, with preference to specified structure. However, according to the likelihood function, the general structure has larger likelihood (or consequently, the smaller log-likelihood).

Table 8.1 presents the fitted PH-distribution mean (Mean), standard-deviation (SD), coefficient of variation (CV) and log-likelihood function (Log-L) for the fitted general structure of order $k = 2, 3, 4, 5, 6$ to the service time - December.

Table 8.2 (p. 50) shows the results of applying EDF tests – the D^* and A^2 statistics associated with the K-S and A-D tests, respectively. These statistics heavily depend on size of the sample data. In table at top, the

Figure 8.8: The specified PH-structure of order $k = 3$ fitted to the service time - December (at top). The fitted PH-density functions with empirical one (at bottom).



Log-likelihood = -215792.369456 (final).

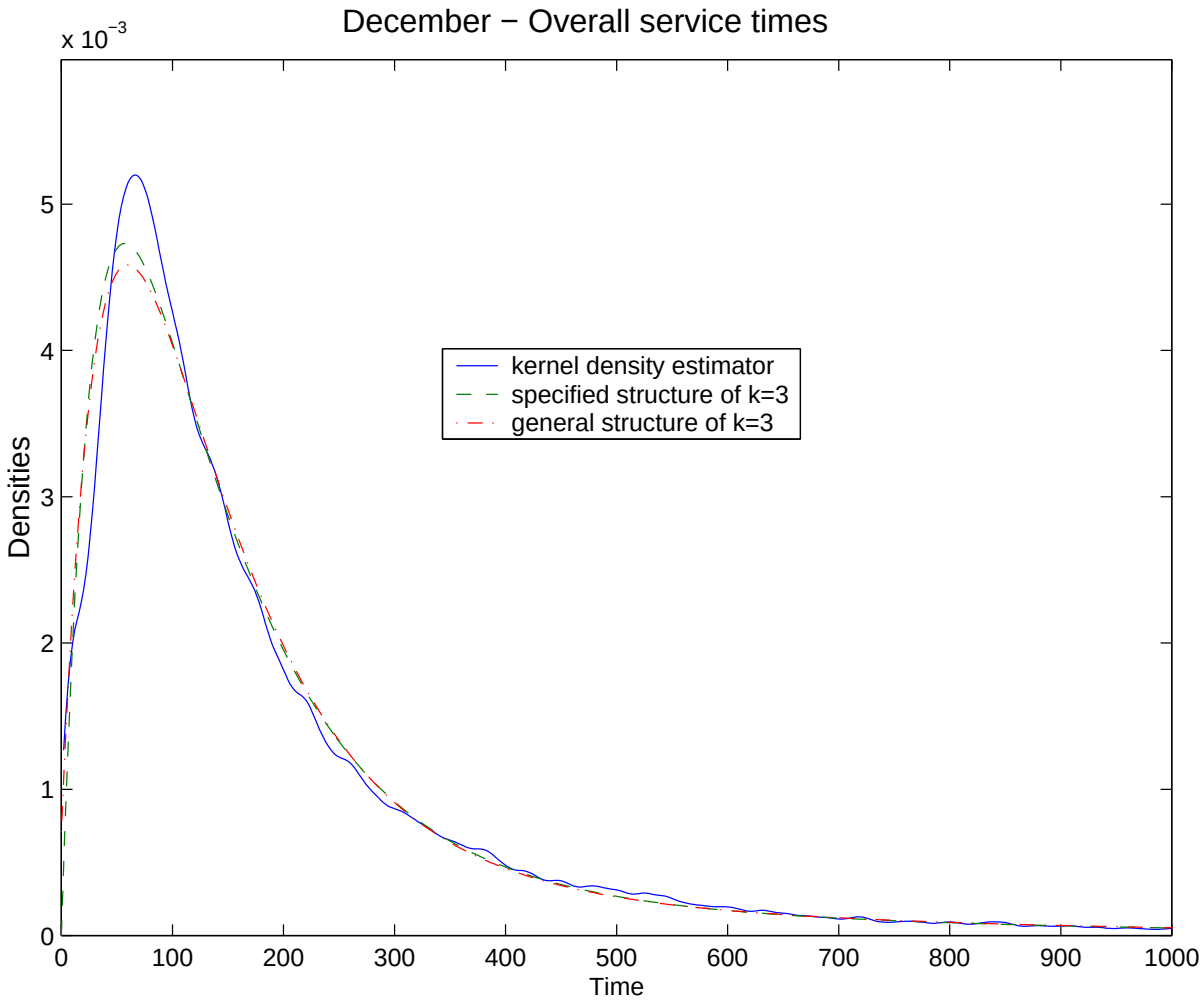


Table 8.1: Service time - December. Statistics.

	k=2	k=3	k=4	k=5	k=6
Mean	207	207	207	207	207
SD	270	253	265	265	264
CV	1.31	1.22	1.28	1.28	1.28
Log-L	-217288	-215736	-215648	-215544	-215425

Table 8.2: Service time - December. EDF tests.

Overall - 34433	k=2	k=3	k=4	k=5	k=6
D^*	17.503	3.754	3.613	**1.708**	1.799
A^2	459.214	20.417	15.294	**3.492**	3.408
D	0.094	0.020	0.019	0.009	0.010

10% of overall - 3443	k=2	k=3	k=4	k=5	k=6
D^*	5.537	**1.182**	1.134	0.535	0.564
A^2	45.795	**2.046**	1.530	0.350	0.341
D	0.094	0.020	0.019	0.009	0.010

D^* and A^2 statistics calculated for overall data - 34433 observations, while in table at bottom, the D^* and A^2 statistics calculated for 10% of overall data, that is 3443 observations. The selected model in the two cases above is different. In the first case, 34433 observations, the PH-model with 5 phases is selected (bold signed), while in the second case, with 3443 observations, the selected PH-model is with only 3 phases. Therefore, it was decided to include in each table the calculation of D, which does not depend on the size of the sample.
There are some explanations to the specified decision:

- For null hypothesis: the distribution of service time is PH-distribution of order 2, it is obtained that $D^* = 17.503$ and $A^2 = 459.214$, associated with the K-S and A-D tests, respectively. According to the critical points of K-S and A-D tests, given in Table 7.3, the null hypothesis above is rejected for any significance level γ. (The same decision received for null hypothesis: the distribution of service time is

PH-distribution of order $k = 3, 4$).

- For null hypothesis: the distribution of service time is PH-distribution of order 5, it is obtained that $D^* = 1.708$. For significance level $\gamma = 0.25, 0.15, 0.1, 0.05, 0.025, 0.01$; $1.708 > c_\gamma$, that is, the null hypothesis above is rejected, by applying K-S test. But for significance level $\gamma = 0.005, 0.001$; $1.708 < c_\gamma$, that is, the null hypothesis above is accepted. The same analysis applied to obtained A-D statistic, $A^2 = 3.492$. For significance level $\gamma = 0.25, 0.15, 0.1, 0.05, 0.025$; $3.492 > a_\gamma$, that is, the null hypothesis above is rejected, by applying K-S test. But for significance level $\gamma = 0.01, 0.005, 0.001$; $3.492 < a_\gamma$, that is, the null hypothesis above is accepted.

- For null hypothesis: the distribution of service time is PH-distribution of order 6, it is obtained that $D^* = 1.799$ and $A^2 = 3.408$. According to K-S test, the null hypothesis is accepted for $\gamma = 0.001$ and rejected for $\gamma = 0.25, 0.15, 0.1, 0.05, 0.025, 0.01, 0.005$. According to A-D test, the null hypothesis is accepted for $\gamma = 0.01, 0.005, 0.001$ and rejected for $\gamma = 0.25, 0.15, 0.1, 0.05, 0.025$.
  However, the PH-model with the least number of phases is selected and this is the PH-model of order 5. This is also the selected PH-model with the least number of phases that fits into a simultaneous confidence band ($\pm\frac{1}{\sqrt{n}}$ and $\pm\frac{1.36}{\sqrt{n}}$) around the empirical CDF. Figure 8.9 presents the simultaneous confidence interval around the empirical CDF of resolution $\pm\frac{1.36}{\sqrt{n}}$. In order to demonstrate how fitted PH-distribution trapped into the bands on each time-interval there are five plots.

Figure 8.10 (p. 55) shows fitted density, distribution, survival and hazard PH-functions together with empirical functions of selected PH-model of order 5, according to the confidence interval and goodness-of-fit tests.

There are estimated parameters of selected PH-model of order 5:

- The Probability of Starting in the state [1,..,5]:

$$\hat{\mathbf{q}} = \begin{bmatrix} 0.09 & 0.91 & 0 & 0 & 0 \end{bmatrix}$$

- The Transition Probability Matrix:

$$\hat{\mathbf{P}} = \begin{bmatrix} 0 & 0.26 & 0 & 0 & 0.12 \\ 0 & 0 & 1 & 0 & 0 \\ 0.99 & 0 & 0 & 0 & 0.01 \\ 0.03 & 0.33 & 0.39 & 0 & 0.26 \\ 0 & 0.58 & 0.39 & 0.02 & 0 \end{bmatrix}$$

- A length of time spent in state [1,..,5] in seconds and in minutes, respectively:

$$\hat{\mathbf{m}} = \begin{bmatrix} 27 & 37 & 37 & 718 & 169 \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} 0.45 & 0.61 & 0.61 & 12 & 2.8 \end{bmatrix}$$

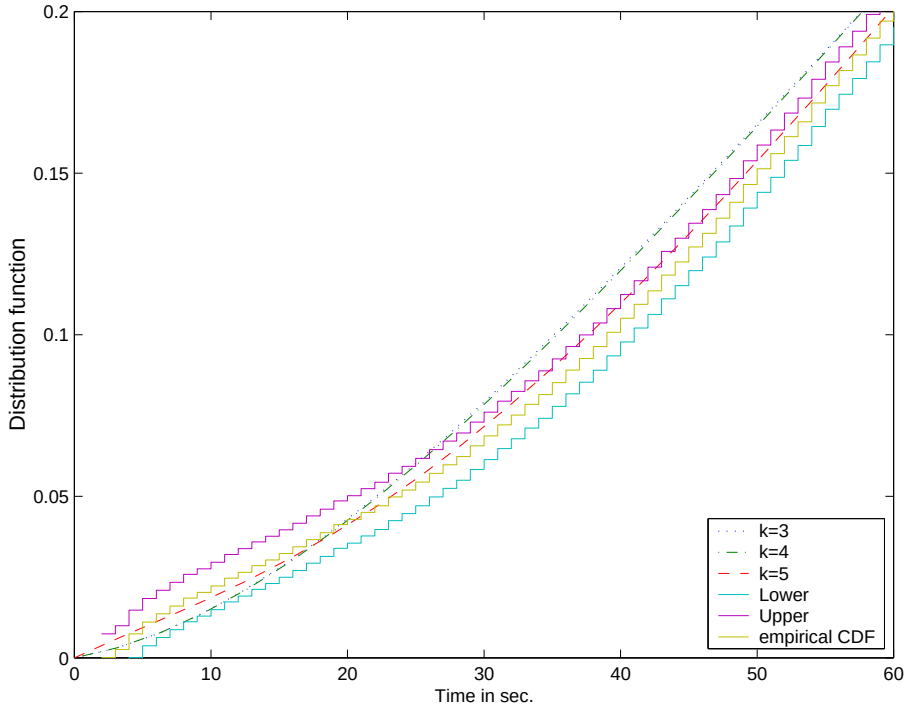
- The Absorption Probability Vector:

$$\hat{\mathbf{p}}_{\triangle} = \begin{bmatrix} 0.61 & 0 & 0 & 0 & 0 \end{bmatrix}$$

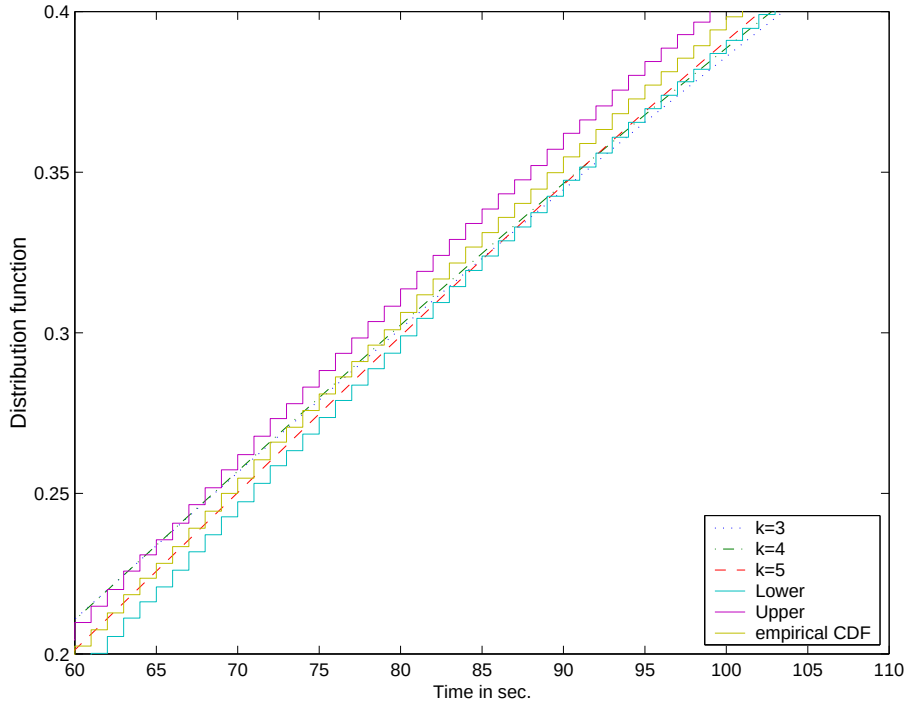
Figure 8.11 (p. 56) shows the PH-structures of order $k = 2, 3, 4, 5, 6$ derived by fitting PH-distributions of general structure to service time - December.

Figure 8.9: The simultaneous confidence interval around the empirical CDF of resolution $\pm\frac{1.36}{\sqrt{n}}$ with fitted PH-distributions of general structure of order $k = 3, 4, 5$.

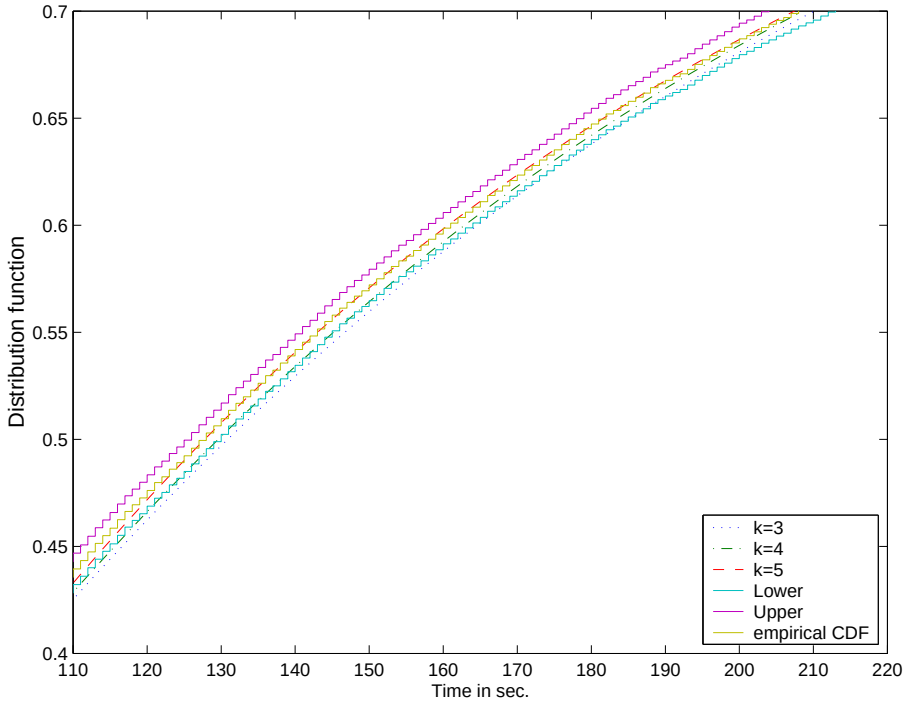
a)



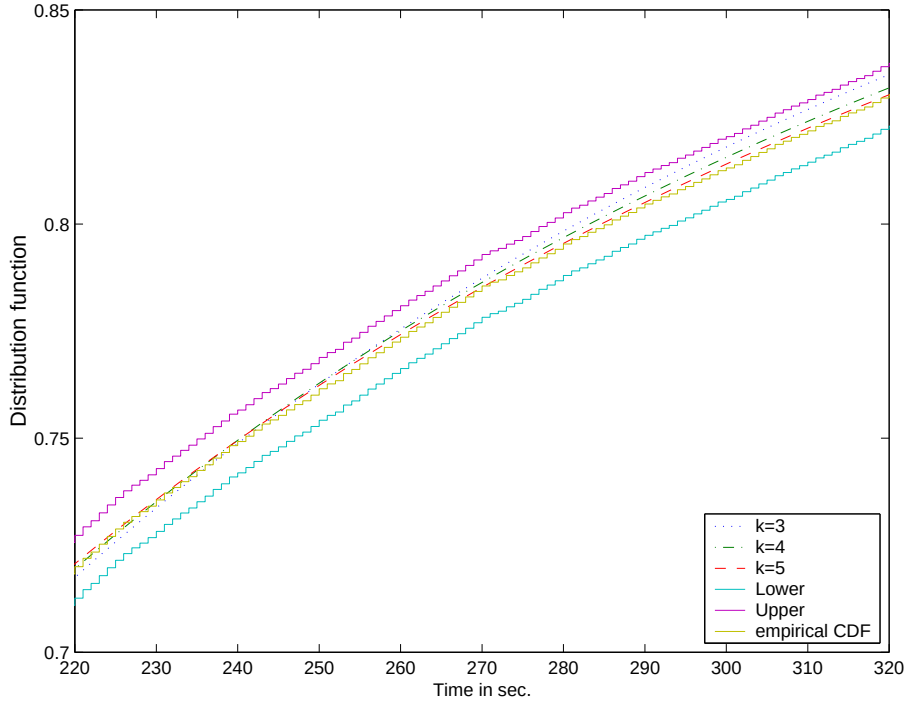
b)



c)


0.7
0.65
0.6
0.55
0.5
0.45
0.4
Distribution function
110
120
130
140
150
160
170
180
190
200
210
220
Time in sec.
k=3
k=4
k=5
Lower
Upper
empirical CDF

d)


0.85
0.8
0.75
0.7
Distribution function
220
230
240
250
260
270
280
290
300
310
320
Time in sec.
k=3
k=4
k=5
Lower
Upper
empirical CDF

e)

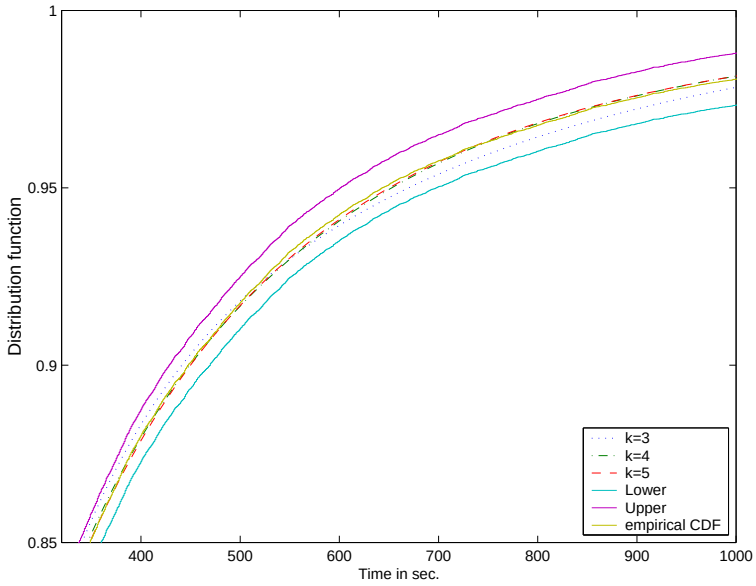


Figure 8.10: PH-type fit of order $k = 5$ of general structure (dashed curve) with empirical functions (solid curve).

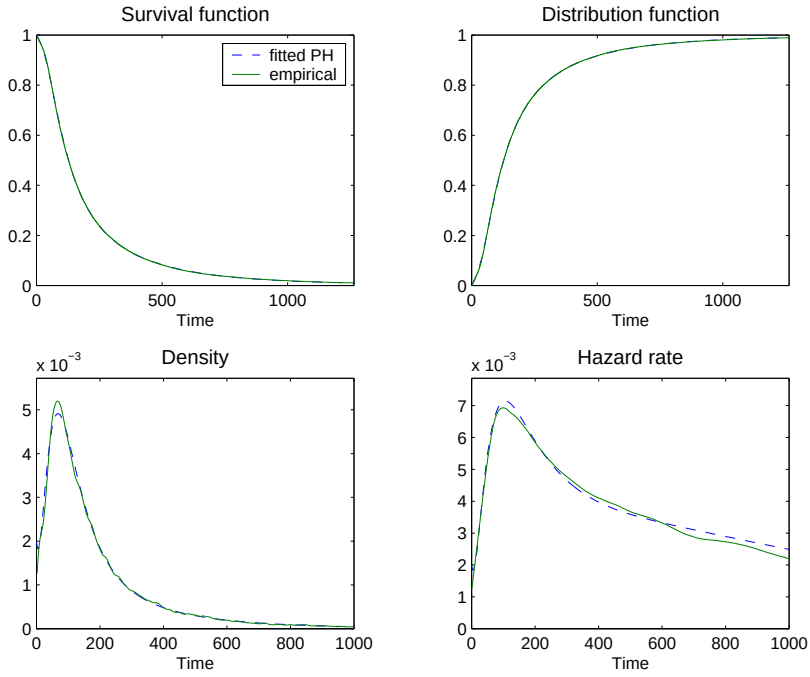
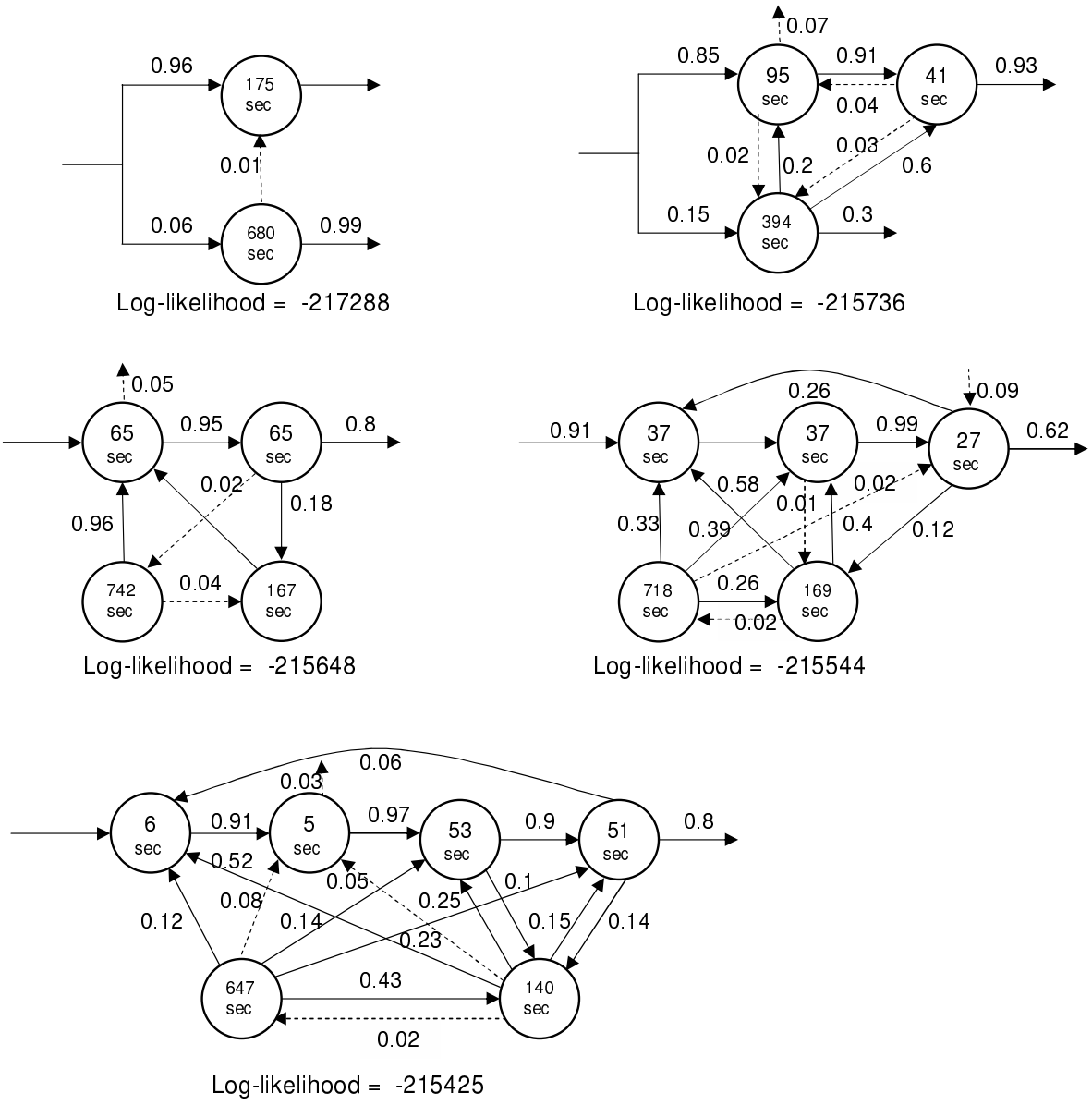


Figure 8.11: Overall service time - December. PH-type structures of order $k = 2, 3, 4, 5, 6$.



8.1.2 Service time – December, by priorities

Figure 8.12 presents histograms of service time – December, by LOW and HIGH priorities. The peak of high frequency of departure of customers with

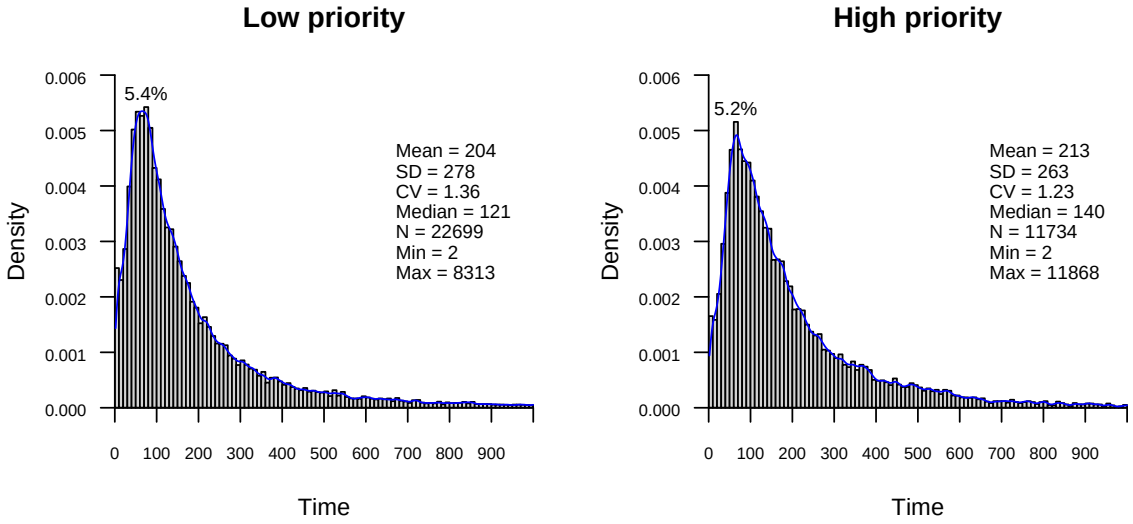


Figure 8.12: Distribution of service time, by priorities. Superimposed on a histogram is the kernel density estimator with a Gaussian kernel of width = 30.

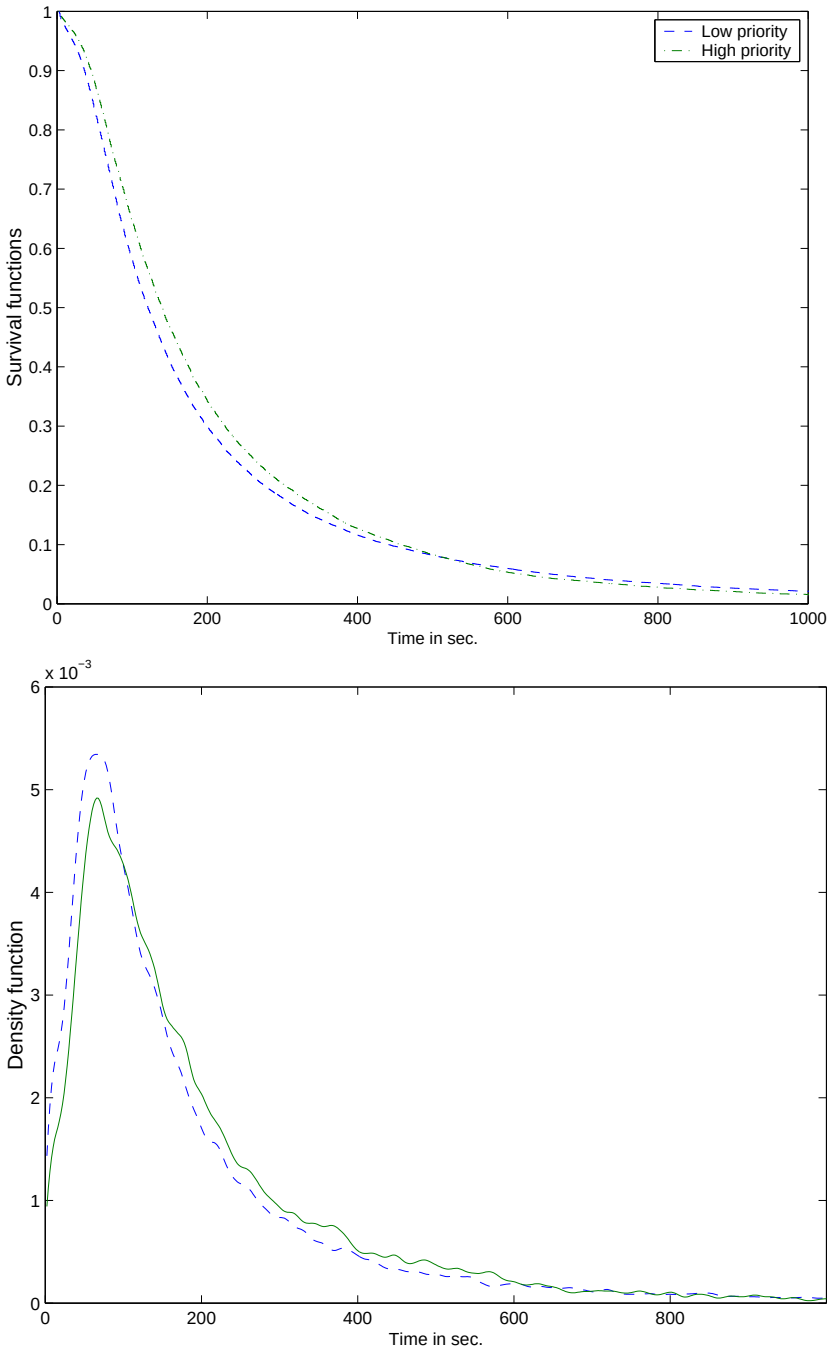
LOW priority is 5.4% at the time-interval [70,80], which is greater than of customers with HIGH priority, 5.2% at the time-interval [60,70].

Figure 8.13 (p. 58) shows the empirical survival and density functions, estimated with non-parametric methods (section 4.1), plotted for LOW and HIGH priorities. In Figure 8.13 we note a stochastic ordering between LOW and HIGH priorities. That is, given the two random variables Y_1 (for LOW priority) and Y_2 (for HIGH priority) with distributions F_1 and F_2, Y_1 stochastically dominates Y_2, denoted $Y_1 \geq_{st} Y_2$, if $S_1(t) \geq S_2(t)$ for all t, where $S(t) = 1 - F(t)$. So, not only the mean of service time for HIGH priority is larger than the mean of service time for LOW priority. In addition, according to survival functions, the proportion of HIGH-priority customers, surviving longer than t, is larger than that of LOW-priority customers.

Figure 8.14 (p. 59) compares the fitted PH-distributions of general structure for $k = 3, 4, 5$ of the density and survival functions, by priorities.

Tables 8.3 and 8.4 (p. 60) present the fitted PH-distribution mean (Mean), standard-deviation (SD), coefficient of variation (CV) and log-likelihood function (Log-L) for the fitted general structure of order $k = 2, 3, 4, 5$ to the

Figure 8.13: Service time - December, by priorities. Empirical results.



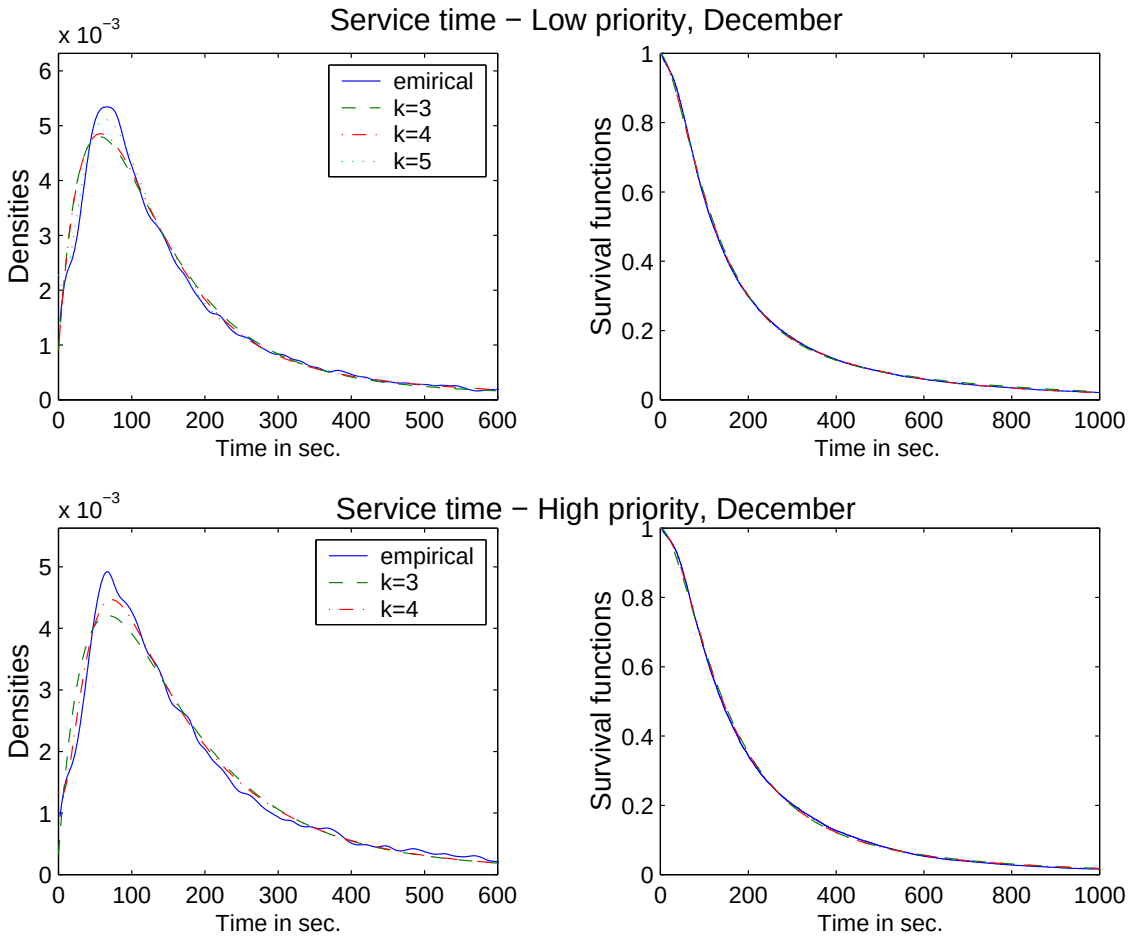


Figure 8.14: Service time - December, by priorities. Phase-type fits of general structure of order $k = 3 - -$, $k = 4 - \cdot -$, and $k = 5 \cdots$. In the density plots, the solid line is the kernel density estimator, given as a comparison to the fitted densities. In the plots of survival functions, the solid line is the empirical survival function.

service time - December, by priorities.

Tables 8.5 and 8.6 (p. 60) present the results of applying EDF tests – the D^* and A^2 statistics associated with the K-S and A-D tests, respectively.

According to the outcome of goodness-of-fit tests, the PH-model for LOW priority with the least number of phases is the selected PH-model of order 5. This is also the selected model that fits into a simultaneous confidence band ($\pm\frac{1}{\sqrt{n}}$) around the empirical CDF.

The null hypothesis: the distribution of service time - LOW priority

Table 8.3: Low priority - Service time, December. Statistics.

	k=2	k=3	k=4	k=5
Mean	204	204	204	204
SD	279	262	273	273
CV	1.37	1.28	1.34	1.34
Log-L	-142708	-141738	-141690	-141626

Table 8.4: High priority - Service time, December. Statistics.

	k=2	k=3	k=4
Mean	213	213	213
SD	256	236	237
CV	1.20	1.11	1.11
Log-L	-74509	-73884	-73831

Table 8.5: Low priority - Service time, December. EDF tests.

Low priority - 22699	k=2	k=3	k=4	k=5
D^*	13.912	3.019	2.889	**1.480**
A^2	285.086	12.882	9.923	**2.526**
D	0.092	0.020	0.019	0.010

Table 8.6: High priority - Service time, December. EDF tests.

High priority - 11734	k=2	k=3	k=4
D^*	11.292	2.481	**1.319**
A^2	189.816	8.678	**3.032**
D	0.104	0.023	0.012

is PH-distribution of order 5, is accepted for $\gamma = 0.025, 0.01, 0.005, 0.001$, according to K-S test, and rejected for $\gamma = 0.25, 0.15, 0.1, 0.05$. According to A-D test, the null hypothesis is accepted for $\gamma = 0.025, 0.01, 0.005, 0.001$ and

rejected for $\gamma = 0.25, 0.15, 0.1, 0.05$.

For HIGH priority, the selected PH-model is of order 4, according to the EDF tests and the simultaneous confidence band ($\pm\frac{1}{\sqrt{n}}$) around the empirical CDF.

The null hypothesis: the distribution of service time - HIGH priority is PH-distribution of order 4, is accepted for $\gamma = 0.05, 0.025, 0.01, 0.005, 0.001$, according to K-S test, and rejected for $\gamma = 0.25, 0.15, 0.1$. According to A-D test, the null hypothesis is accepted for $\gamma = 0.025, 0.01, 0.005, 0.001$ and rejected for $\gamma = 0.25, 0.15, 0.1, 0.05$.

Figure 8.15 shows fitted density, distribution, survival and hazard PH-functions together with empirical functions of selected PH-model of order 5, according to the confidence interval and goodness-of-fit tests, to Low-priority customers.

There are estimated parameters of selected PH-model of order 5 fitted to Low priority - Service time, December:

- The Probability of Starting in the state [1,..,5]:

$$\hat{\mathbf{q}} = \begin{bmatrix} 0 & 0.09 & 0 & 0.91 & 0 \end{bmatrix}$$

- The Transition Probability Matrix:

$$\hat{\mathbf{P}} = \begin{bmatrix} 0 & 0 & 0.07 & 0.93 & 0 \\ 0.05 & 0 & 0 & 0.28 & 0 \\ 0.68 & 0 & 0 & 0.24 & 0.09 \\ 0 & 0 & 0 & 0 & 1 \\ 0.09 & 0.91 & 0 & 0 & 0 \end{bmatrix}$$

- A length of time spent in state [1,..,5] in seconds and in minutes, respectively:

$$\hat{\mathbf{m}} = \begin{bmatrix} 169 & 26 & 591 & 35 & 35 \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} 2.8 & 0.4 & 9.9 & 0.6 & 0.6 \end{bmatrix}$$

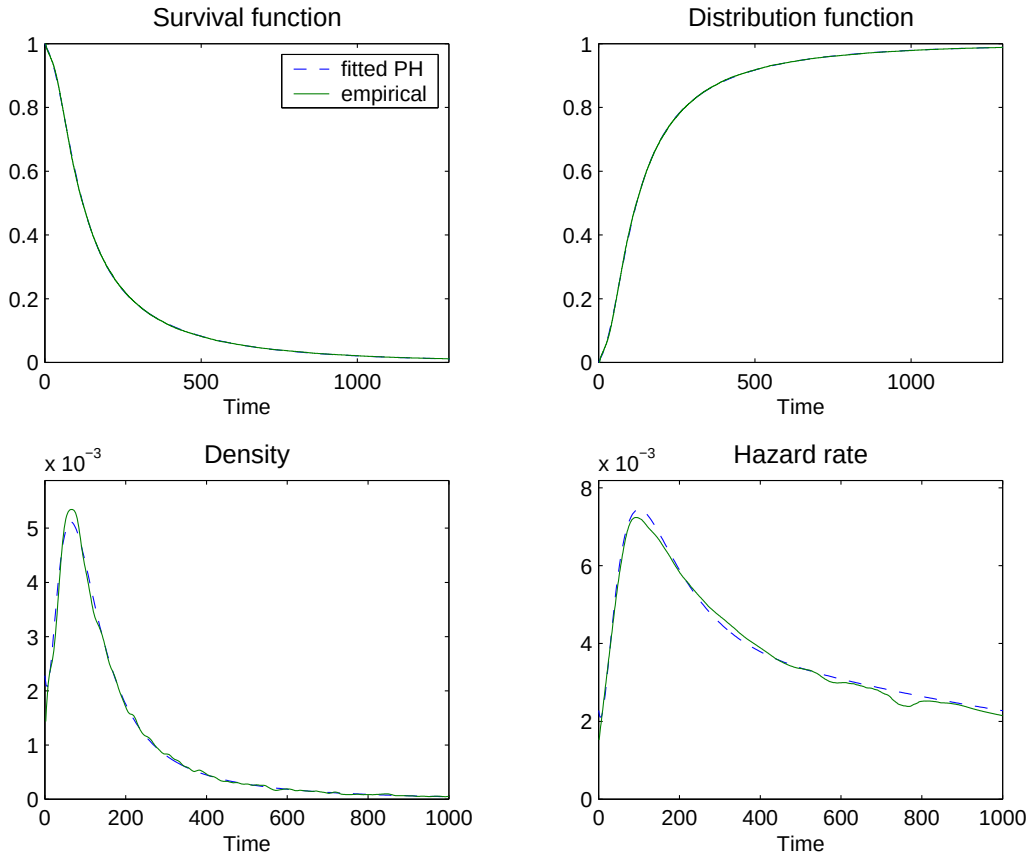
- The Absorption Probability Vector:

$$\hat{\mathbf{p}}_{\triangle} = \begin{bmatrix} 0 & 0.66 & 0 & 0 & 0 \end{bmatrix}$$

Figure 8.16 (p. 63) shows fitted density, distribution, survival and hazard PH-functions together with empirical functions of selected PH-model of order 4, according to the confidence interval and goodness-of-fit tests, to High-priority customers.

There are estimated parameters of selected PH-model of order 4 fitted to High priority - Service time, December:

Figure 8.15: Low priority - Service time, December. PH-type fit of order $k = 5$ of general structure (dashed curve) with empirical functions (solid curve).



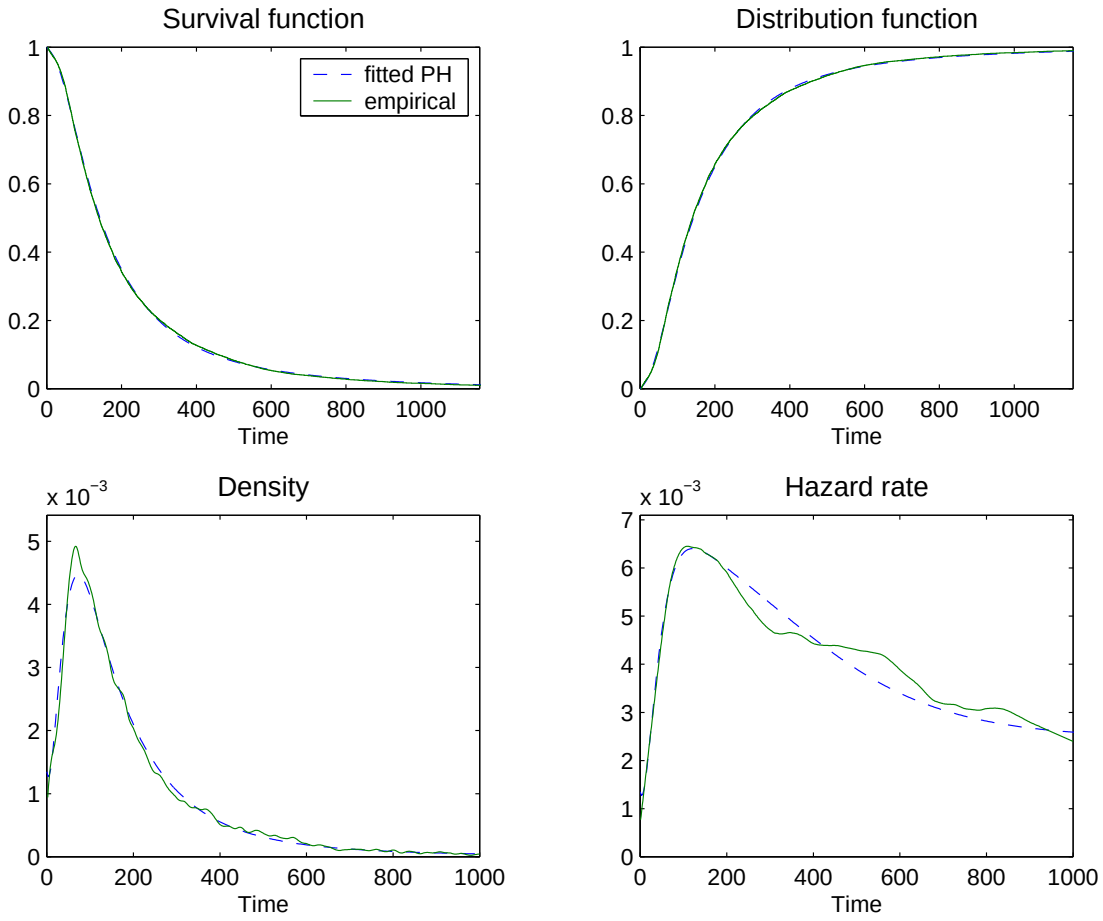
- The Probability of Starting in the state [1,..,4]:

$$\hat{\mathbf{q}} = \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}$$

- The Transition Probability Matrix:

$$\hat{\mathbf{P}} = \begin{bmatrix} 0 & 0.02 & 0.98 & 0 \\ 0.01 & 0 & 0 & 0.99 \\ 0 & 0.03 & 0 & 0.44 \\ 0.96 & 0 & 0 & 0 \end{bmatrix}$$

Figure 8.16: High priority - Service time, December. PH-type fit of order $k = 4$ of general structure (dashed curve) with empirical functions (solid curve).



- A length of time spent in state [1,..,4] in seconds and in minutes, respectively:

$$\hat{\mathbf{m}} = \begin{bmatrix} 36 & 354 & 36 & 29 \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} 0.6 & 5.9 & 0.6 & 0.48 \end{bmatrix}$$

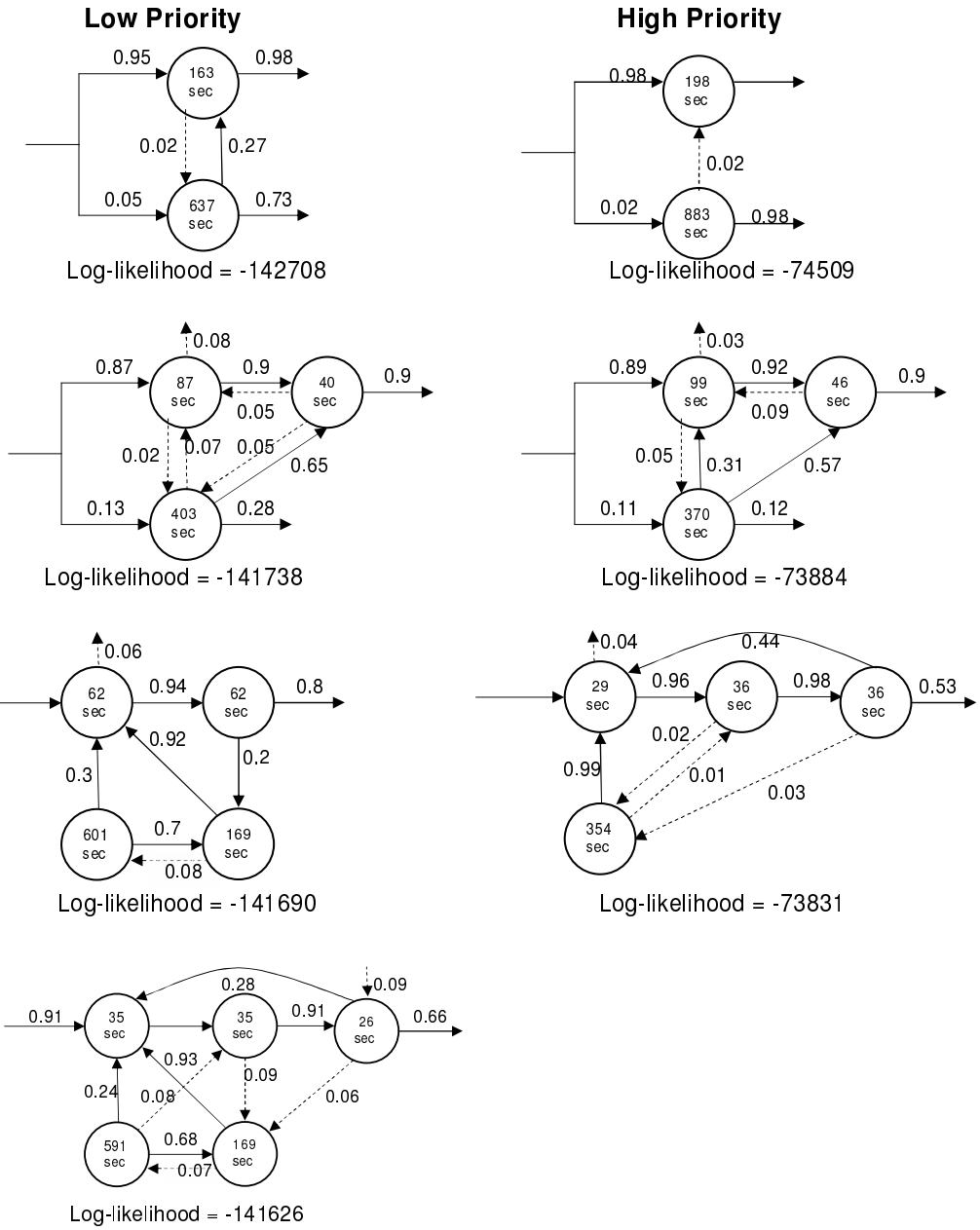
- The Absorption Probability Vector:

$$\hat{\mathbf{p}}_{\triangle} = \begin{bmatrix} 0 & 0 & 0.53 & 0.04 \end{bmatrix}$$

Figure 8.17 (p. 64) shows the PH-structures of order $k = 3, 4, 5$ derived

by fitting PH-distributions of general structure to service time - December, by priorities.

Figure 8.17: Service time - December, by priorities. PH-type structures of order $k = 3, 4, 5$.



8.1.3 Service time – December, by types

Figure 8.18 presents histograms of service time –December, by four major service types - PS, NE, IN, NW.

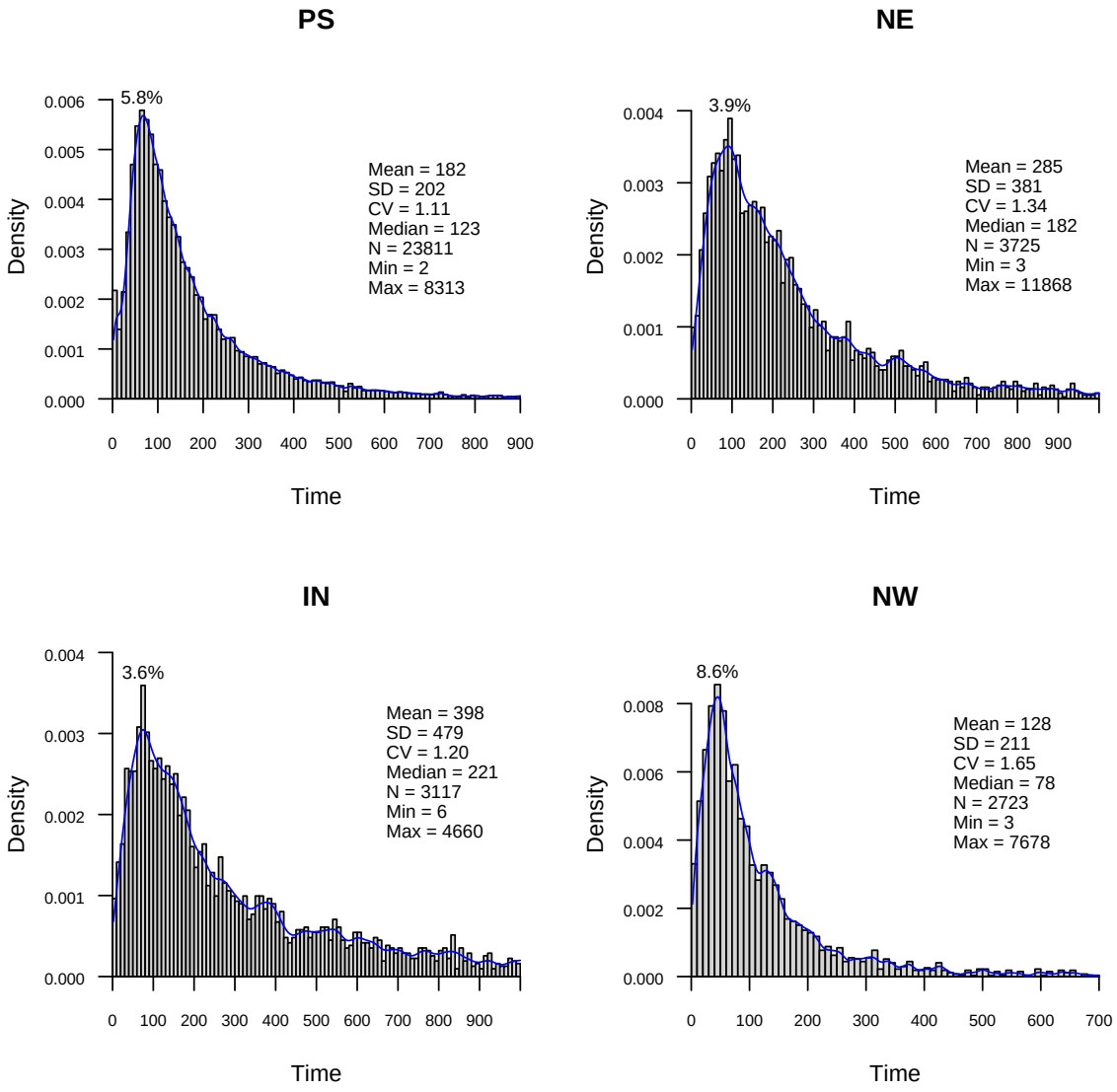
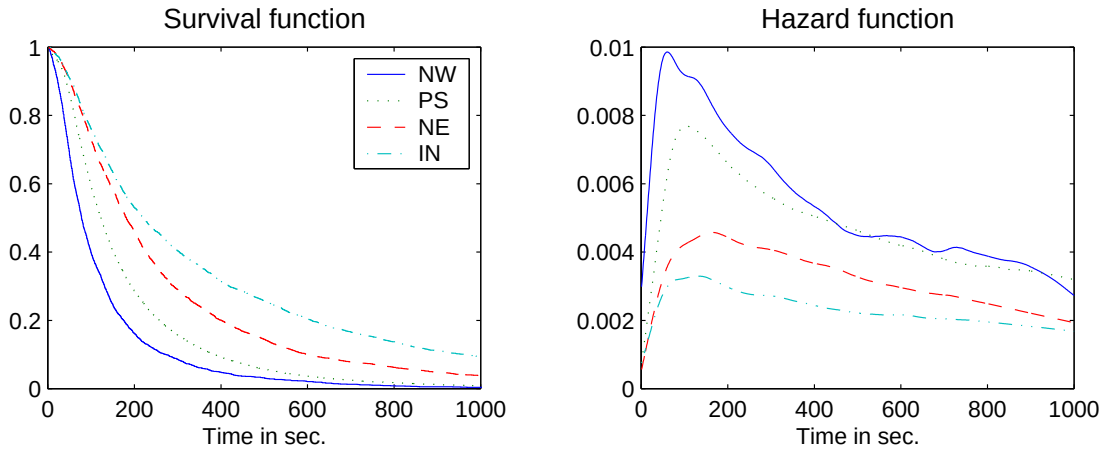


Figure 8.18: Distribution of service time, by four major service types. Superimposed on a histogram is the kernel density estimator with a Gaussian kernel of width=30.

Figure 8.19 compares the service time distribution of the four major service types: IN, NE, NW and PS, by estimating their densities, survival, haz-

ard and distribution functions with non-parametric methods (section 4.1). According to figure 8.19, there is a stochastic ordering between the four ma-

Figure 8.19: Service time - December, by types. Empirical results.



jor service types.

Figure 8.20 (p. 67) compares the fitted PH-distributions of general structure for $k = 2, 3, 4, 5, 6$ of the density and survival functions, by types. In Figure 8.20 a), in the plot of densities of PS service type the fitted PH-distribution of order $k = 2, 3, 5, 6$ is presented with kernel density estimator of width = 30, since the fitted PH-distribution of order 4 and 5 are coincide. The fitted PH-distribution of order 6 is almost coincide with the kernel density estimator. In the plot of survival functions, fitted survival functions of order $k = 3, 4, 5, 6$ are coincide with empirical one, according to the graph of the presented scale.

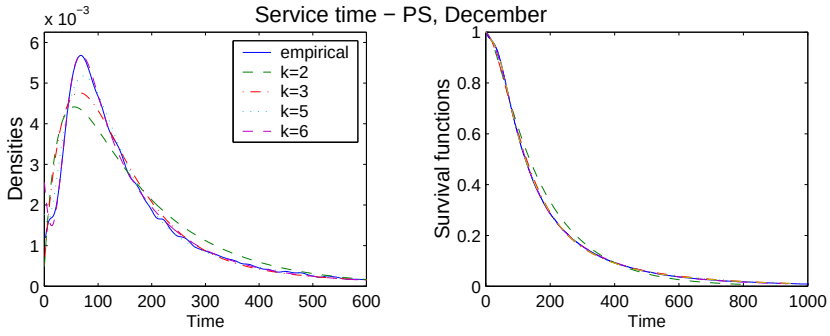
In Figure 8.20 b) and c), in the plot of densities of NE and IN service types the fitted PH-distributions of order $k = 2, 3, 4$ are presented with kernel density estimator of width=60. The fitted PH-distribution of order 3 and 4 are almost coincide. In the plot of survival functions, fitted survival functions of order 3 and 4 are coincide with empirical one, according to the graph of the presented scale.

In Figure 8.20 d), in the plot of densities of NW service type the fitted PH-distributions of order $k = 2, 3, 4$ are presented with kernel density estimator of width=30. The fitted PH-distribution of order 3 and 4 are almost coincide. In the plot of survival functions, fitted survival functions of order 3 and 4 are coincide with empirical one, according to the graph of the presented scale.

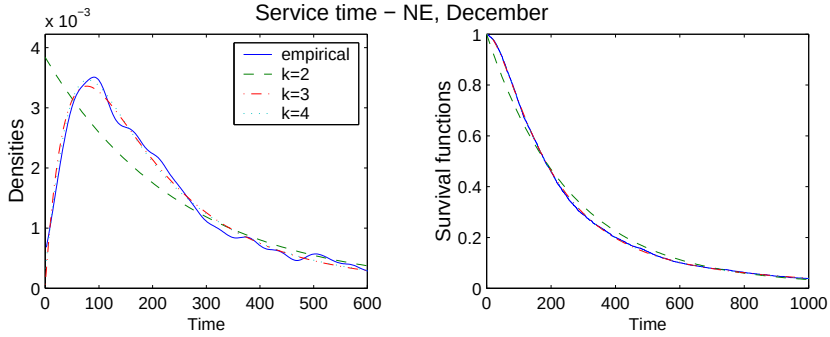
Tables 8.1.3 - 8.1.3 (p. 68 - 69) present the fitted PH-distribution mean

Figure 8.20: Service time - December, by types. Phase-type fits of general structure of order k. In the density plots, the solid line is the kernel density estimator, given as a comparison to the fitted densities. In the plots of survival functions, the solid line is the empirical survival function.

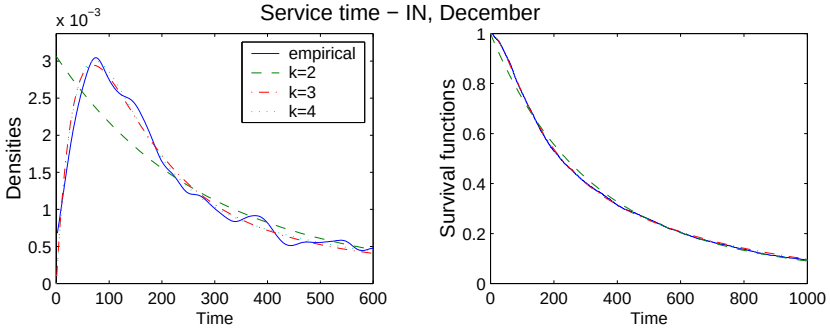
a)



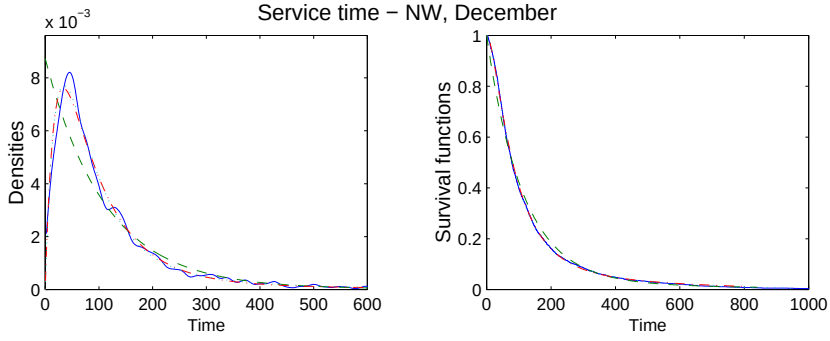
b)



c)



d)



(Mean), standard-deviation (SD), coefficient of variation (CV) and log-likelihood function (Log-L) for the fitted general structure of order $k = 2, 3, 4, 5, 6$ to the service time - December, by types:

Table 8.7: Service time - PS, December. Statistics.

	k=2	k=3	k=4	k=5	k=6
Mean	182	182	182	182	182
SD	159	190	190	195	192
CV	0.87	1.04	1.04	1.07	1.05
Log-L	-146640	-146093	-145902	-145857	-145785

Table 8.8: Service time - NE, December. Statistics.

	k=2	k=3	k=4
Mean	285	285	285
SD	361	341	342
CV	1.27	1.19	1.20
Log-L	-24715	-24544	-24542

Table 8.9: Service time - IN, December. Statistics.

	k=2	k=3	k=4
Mean	398	398	398
SD	484	470	480
CV	1.22	1.18	1.21
Log-L	-21723	-21628	-21625

Tables 8.11 - 8.14 (p. 69 - 70) presents the results of applying EDF tests – the D^* and A^2 statistics associated with the K-S and A-D tests, respectively.

According to the outcome of goodness-of-fit tests, the selected PH-model for PS service type is of order 6 (from Table 8.11, bold signed). This is also the PH-model that fits into a simultaneous confidence band ($\pm\frac{1}{\sqrt{n}}$) around the empirical cumulative distribution function. However, the PH-model of order 5 and 4 almost fits into a simultaneous confidence band ($\pm\frac{1.36}{\sqrt{n}}$).

Table 8.10: Service time - NW, December. Statistics.

	k=2	k=3	k=4
Mean	128	128	128
SD	178	166	206
CV	1.39	1.29	1.60
Log-L	-15858	-15730	-15713

Table 8.11: Service time - PS, December. EDF tests.

PS - 23811	k=2	k=3	k=4	k=5	k=6
D^*	8.390	4.668	2.582	2.391	**1.077**
A^2	129.472	29.937	9.303	6.509	**1.661**
D	0.054	0.030	0.017	0.015	0.007

Table 8.12: Service time - NE, December. EDF tests.

NE - 3725	k=2	k=3	k=4
D^*	4.994	**0.732**	0.682
A^2	43.690	**0.545**	0.352
D	0.082	0.012	0.011

Table 8.13: Service time - IN, December. EDF tests.

IN - 3117	k=2	k=3	k=4
D^*	2.997	**0.598**	0.424
A^2	15.464	**0.428**	0.246
D	0.054	0.011	0.008

For NE, IN, NW service type, the selected PH-model is of order 3, according to the EDF tests and the simultaneous confidence band $\pm\frac{1}{\sqrt{n}}$ (consequently, $\pm\frac{1.36}{\sqrt{n}}$) around the empirical CDF.

The null hypothesis: the distribution of service time - PS type is PH-distribution of order 6, accepted for $\gamma = 0.15, 0.1, 0.05, 0.025, 0.01, 0.005, 0.001,$

Table 8.14: Service time - NW, December. EDF tests.

NW - 2723	k=2	k=3	k=4
D^*	4.562	**1.010**	0.937
A^2	33.678	**1.294**	0.826
D	0.087	0.019	0.018

according to K-S test, and rejected for $\gamma = 0.25$. According to A-D test, the null hypothesis is accepted for $\gamma = 0.1, 0.05, 0.025, 0.01, 0.005, 0.001$ and rejected for $\gamma = 0.25, 0.15$. For NE service type, the null hypothesis: the distribution of service time is PH-distribution of order 3, accepted for any significance level γ, according to K-S test and A-D test. The same conclusion derived for IN service type. For NW service type, the null hypothesis: the distribution of service time is PH-distribution of order 3, accepted for $\gamma = 0.15, 0.1, 0.05, 0.025, 0.01, 0.005, 0.001$, according to K-S test, and rejected for $\gamma = 0.25$. According to A-D test, the null hypothesis is accepted for any significance level γ.

Figure 8.21 shows fitted density, distribution, survival and hazard PH-functions together with empirical functions of selected PH-model of order 6, according to the goodness-of-fit tests, to PS service type. There are estimated parameters of selected PH-model of order 6 fitted to Service time - PS, December:

- The Probability of Starting in the state [1,..,6]:

$$\hat{\mathbf{q}} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

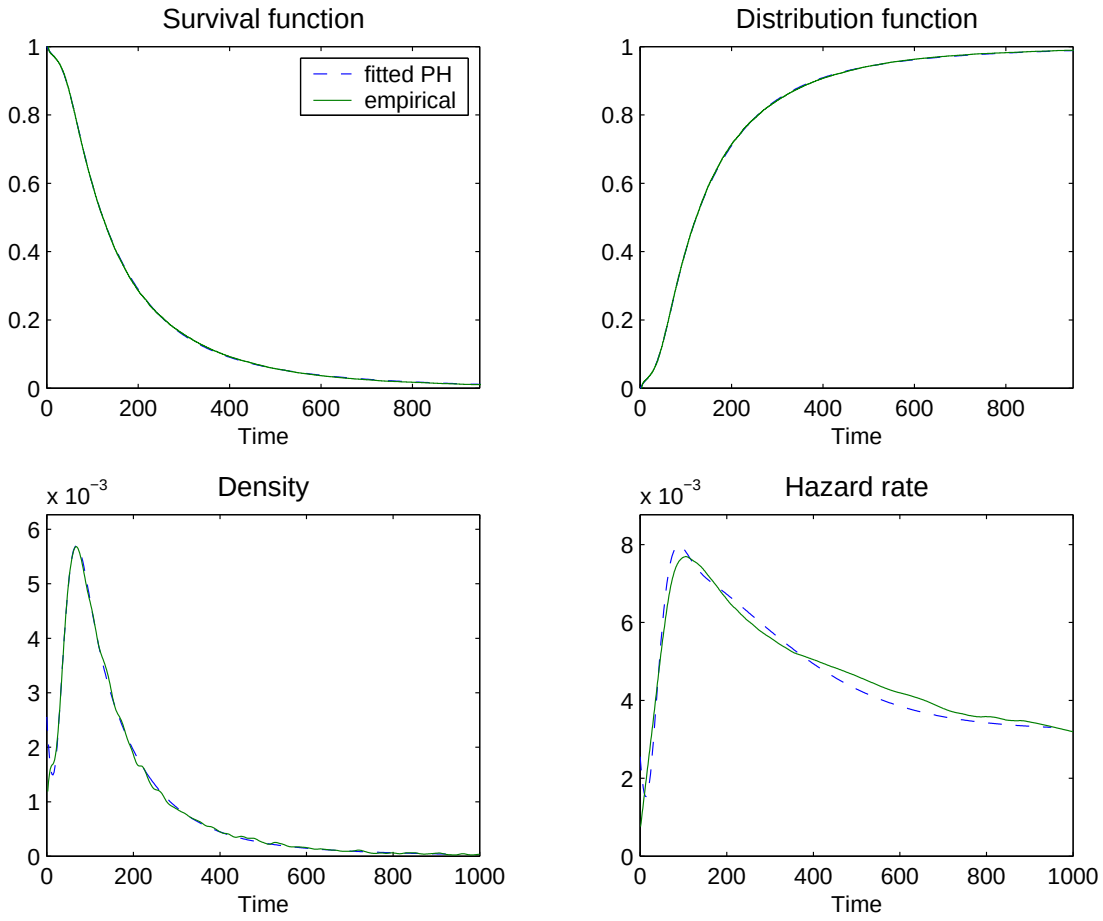
- The Transition Probability Matrix:

$$\hat{\mathbf{P}} = \begin{bmatrix} 0 & 0 & 0 & 0.01 & 0 & 0.99 \\ 0.1 & 0 & 0.86 & 0 & 0 & 0 \\ 0.91 & 0 & 0 & 0.02 & 0 & 0.07 \\ 0.15 & 0.32 & 0.11 & 0 & 0.37 & 0.04 \\ 0 & 0.48 & 0 & 0.03 & 0 & 0 \\ 0 & 0 & 0 & 0.01 & 0.99 & 0 \end{bmatrix}$$

- A length of time spent in state [1,..,6] in seconds and in minutes, respectively:

$$\hat{\mathbf{m}} = \begin{bmatrix} 15 & 17 & 14 & 269 & 20 & 20 \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} 0.2 & 0.3 & 0.2 & 4.5 & 0.3 & 0.3 \end{bmatrix}$$

Figure 8.21: Service time - PS, December. PH-type fit of order $k = 6$ of general structure (dashed curve) with empirical functions (solid curve).

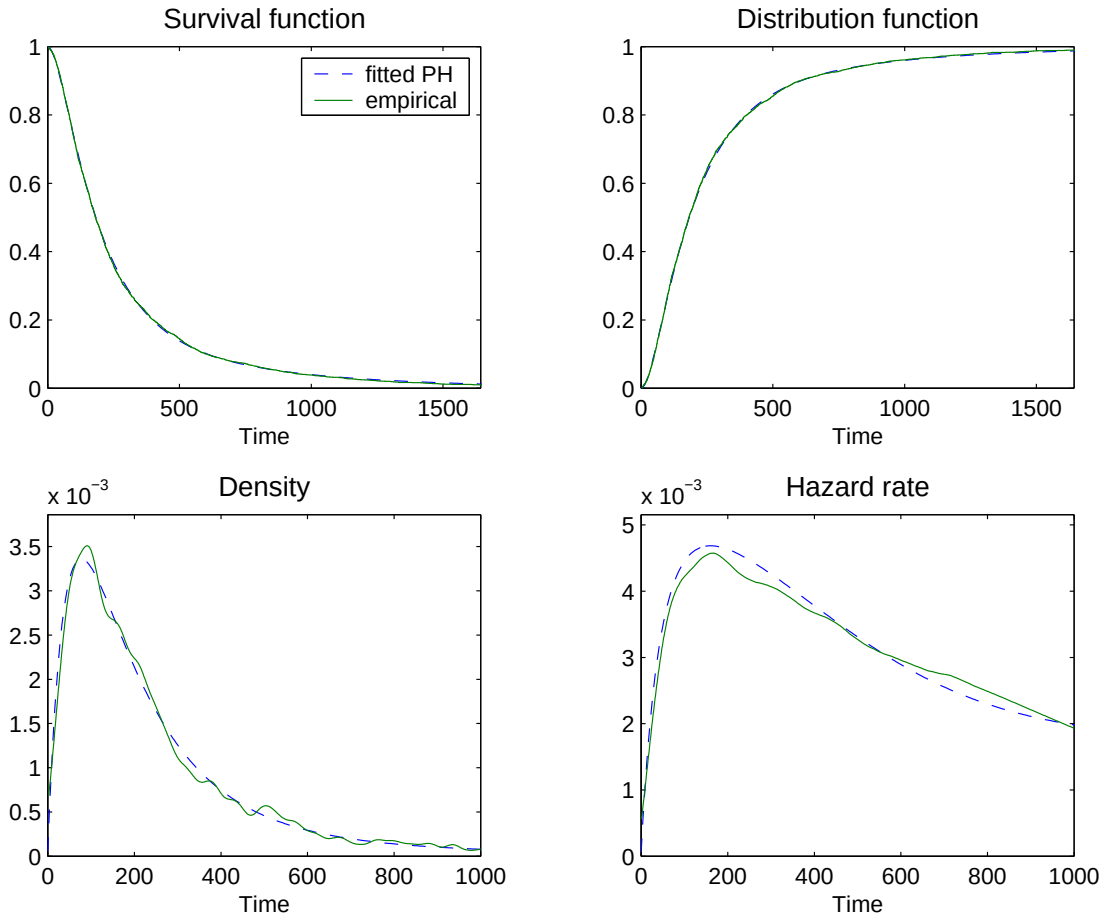


- The Absorption Probability Vector:

$$\hat{\mathbf{p}}_{\triangle} = \begin{bmatrix} 0 & 0.04 & 0 & 0.01 & 0.49 & 0 \end{bmatrix}$$

Figure 8.22 (p. 72) shows fitted density, distribution, survival and hazard PH-functions together with empirical functions of selected PH-model of order 3, according to the goodness-of-fit tests, to NE service type. There are estimated parameters of selected PH-model of order 3 fitted to Service time - NE, December:

Figure 8.22: Service time - NE, December. PH-type fit of order $k = 3$ of general structure (dashed curve) with empirical functions (solid curve).



- The Probability of Starting in the state [1,..,3]:

$$\hat{\mathbf{q}} = \begin{bmatrix} 0.91 & 0 & 0.09 \end{bmatrix}$$

- The Transition Probability Matrix:

$$\hat{\mathbf{P}} = \begin{bmatrix} 0 & 0.98 & 0.02 \\ 0.24 & 0 & 0.04 \\ 0.08 & 0.7 & 0 \end{bmatrix}$$

- A length of time spent in state [1,..,3] in seconds and in minutes, re-

spectively:

$$\hat{\mathbf{m}} = \left[\begin{array}{ccc} 109 & 48 & 546 \end{array}\right] \quad \text{or} \quad \left[\begin{array}{ccc} 1.8 & 0.8 & 6.1 \end{array}\right]$$

- The Absorption Probability Vector:

$$\hat{\mathbf{p}}_{\triangle} = \left[\begin{array}{ccc} 0 & 0.72 & 0.22 \end{array}\right]$$

Figure 8.23 shows fitted density, distribution, survival and hazard PH-functions together with empirical functions of selected PH-model of order 3, according to the goodness-of-fit tests, to IN service type. There are estimated parameters of selected PH-model of order 3 fitted to Service time - IN, December:

- The Probability of Starting in the state [1,..,3]:

$$\hat{\mathbf{q}} = \left[\begin{array}{ccc} 0 & 0 & 1 \end{array}\right]$$

- The Transition Probability Matrix:

$$\hat{\mathbf{P}} = \left[\begin{array}{ccc} 0 & 0.16 & 0.01 \\ 0.04 & 0 & 0.6 \\ 0.66 & 0.34 & 0 \end{array}\right]$$

- A length of time spent in state [1,..,3] in seconds and in minutes, respectively:

$$\hat{\mathbf{m}} = \left[\begin{array}{ccc} 121 & 369 & 37 \end{array}\right] \quad \text{or} \quad \left[\begin{array}{ccc} 2 & 6.1 & 0.6 \end{array}\right]$$

- The Absorption Probability Vector:

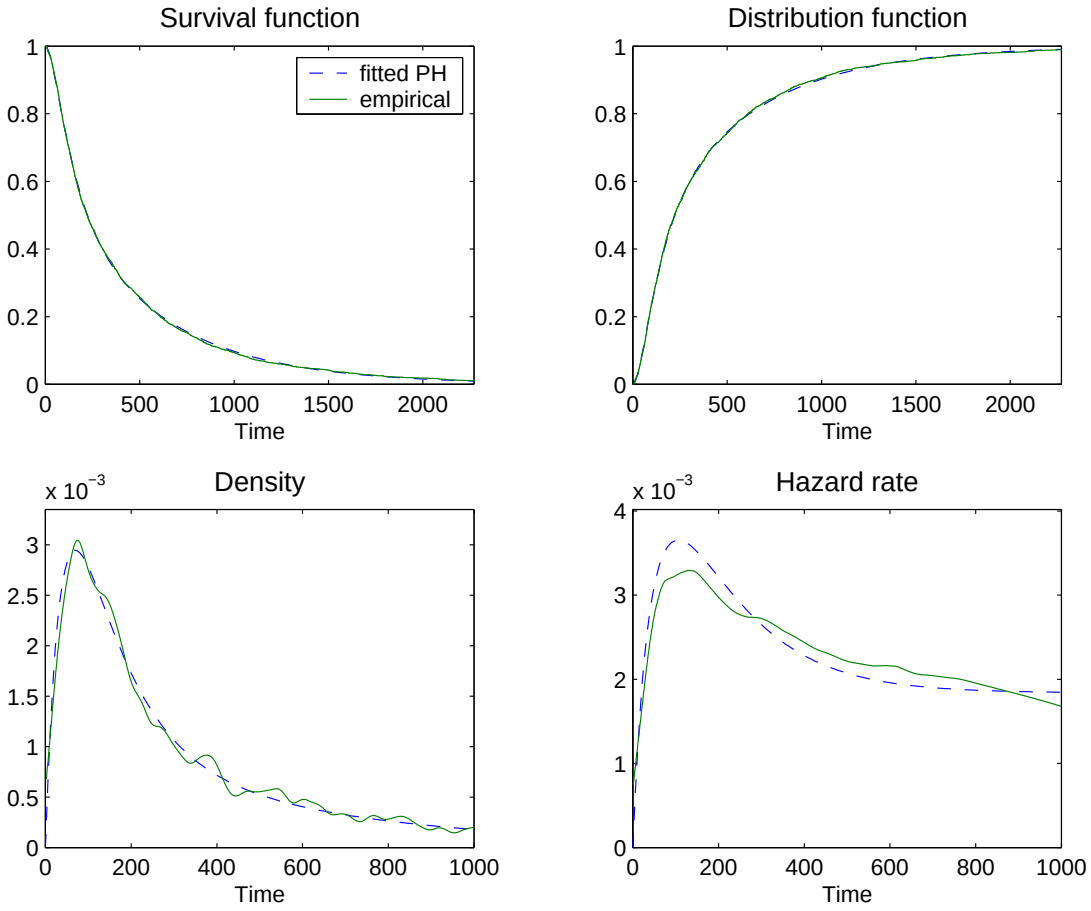
$$\hat{\mathbf{p}}_{\triangle} = \left[\begin{array}{ccc} 0.83 & 0.36 & 0 \end{array}\right]$$

Figure 8.24 (p. 75) shows fitted density, distribution, survival and hazard PH-functions together with empirical functions of selected PH-model of order 3, according to the goodness-of-fit tests, to NW service type. There are estimated parameters of selected PH-model of order 3 fitted to Service time - NW, December:

- The Probability of Starting in the state [1,..,3]:

$$\hat{\mathbf{q}} = \left[\begin{array}{ccc} 0.03 & 0 & 0.97 \end{array}\right]$$

Figure 8.23: Service time - IN, December. PH-type fit of order $k = 3$ of general structure (dashed curve) with empirical functions (solid curve).



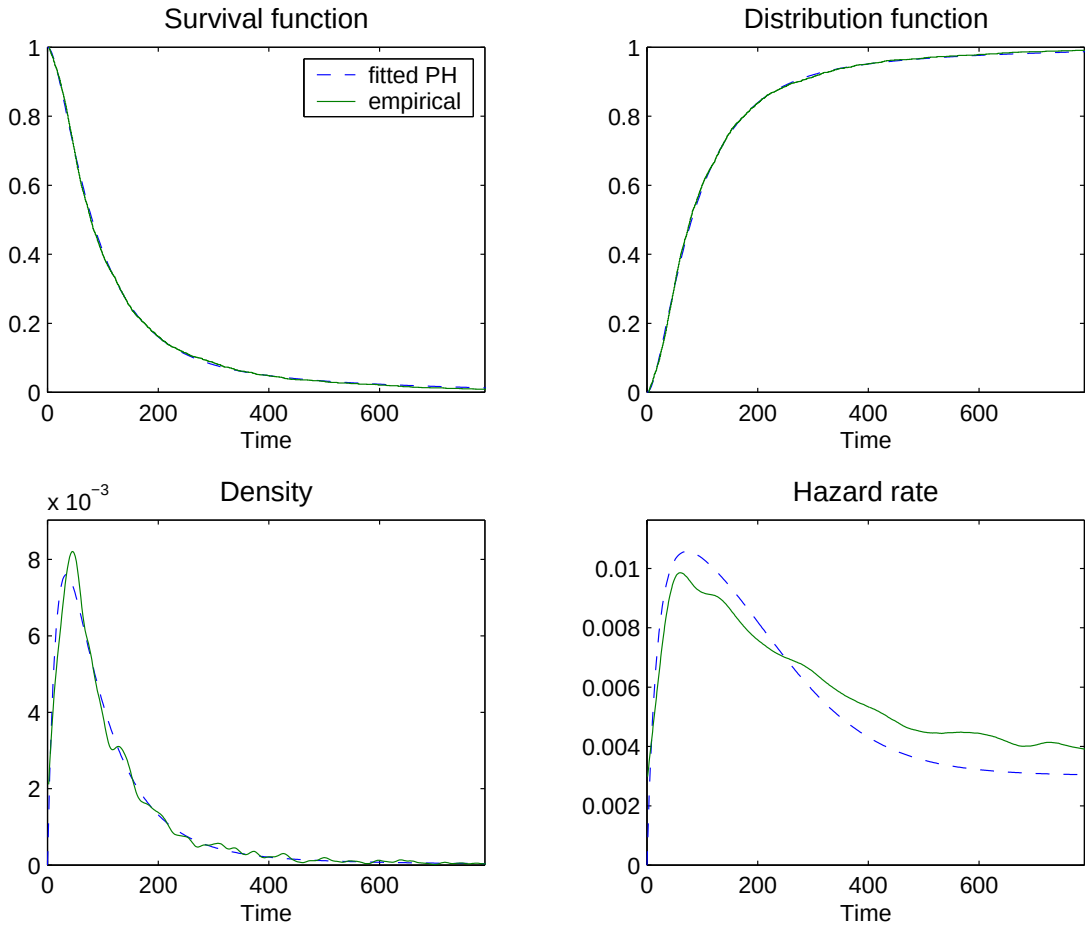
- The Transition Probability Matrix:

$$\hat{\mathbf{P}} = \begin{bmatrix} 0 & 0.86 & 0.14 \\ 0.03 & 0 & 0.35 \\ 0.01 & 0.99 & 0 \end{bmatrix}$$

- A length of time spent in state [1,..,3] in seconds and in minutes, respectively:

$$\hat{\mathbf{m}} = \begin{bmatrix} 308 & 41 & 22 \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} 5.1 & 0.7 & 0.4 \end{bmatrix}$$

Figure 8.24: Service time - NW, December. PH-type fit of order $k = 3$ of general structure (dashed curve) with empirical functions (solid curve).

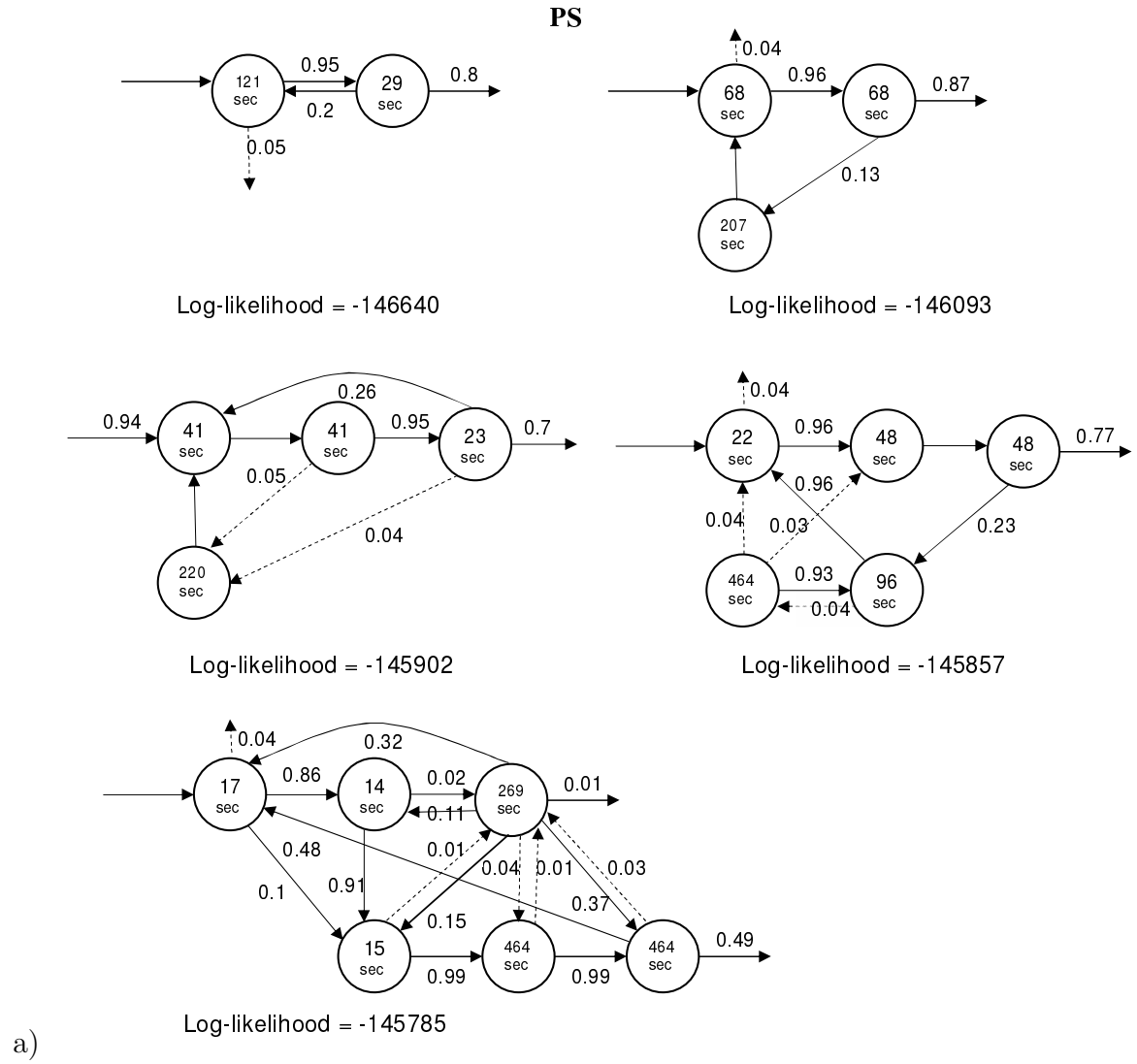


- The Absorption Probability Vector:

$$\hat{\mathbf{p}}_{\triangle} = \begin{bmatrix} 0 & 0.62 & 0 \end{bmatrix}$$

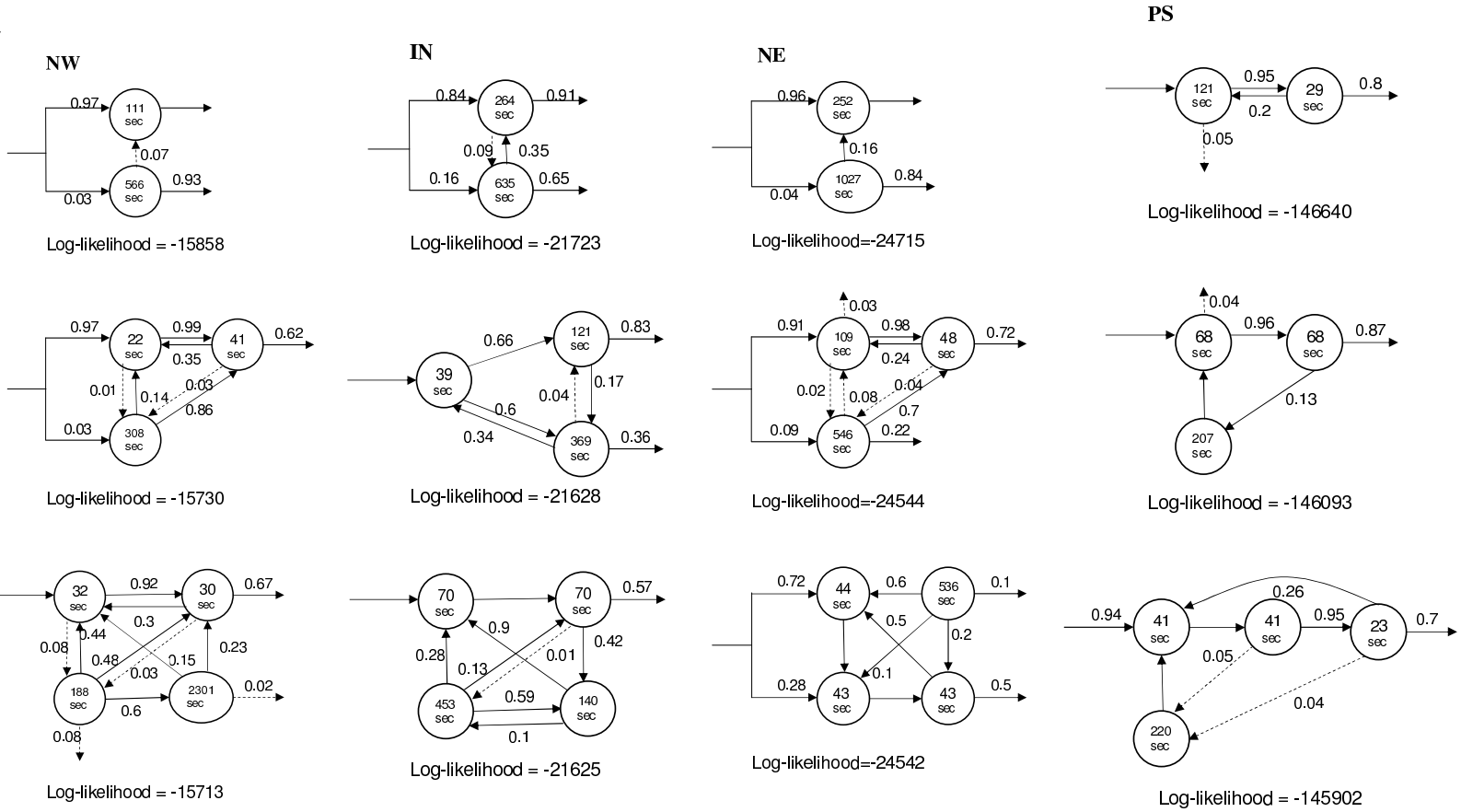
Figure 8.25 a) shows the PH-structures of order $k = 2, 3, 4, 5, 6$ derived by fitting PH-distributions of general structure to service time - PS, December. Figure 8.25 b) compares the PH-structures of order $k = 2, 3, 4$ for the four main service types.

Figure 8.25: Service time - December, by types. PH-type structures of order $k = 2, 3, 4, 5, 6$.



a)

b)


NW
0.97
111 sec
0.07
566 sec
0.93
0.03
Log-likelihood = -15858
IN
0.84
264 sec
0.91
0.09
0.35
0.16
635 sec
0.65
Log-likelihood = -21723
NE
0.96
252 sec
0.16
0.04
1027 sec
0.84
Log-likelihood=-24715
PS
121 sec
0.95
0.2
29 sec
0.8
0.05
Log-likelihood = -146640
0.97
22 sec
0.99
41 sec
0.62
0.35
0.01
0.14
0.03
0.86
0.03
308 sec
Log-likelihood = -15730
0.66
121 sec
0.83
39 sec
0.04
0.17
0.6
0.34
369 sec
0.36
Log-likelihood = -21628
0.03
0.91
109 sec
0.98
48 sec
0.72
0.24
0.02
0.08
0.04
0.7
0.09
546 sec
0.22
Log-likelihood=-24544
0.04
68 sec
0.96
68 sec
0.87
0.13
207 sec
Log-likelihood = -146093
32 sec
0.92
30 sec
0.67
0.3
0.44
0.08
0.48
0.15
0.23
0.03
188 sec
2301 sec
0.02
0.6
0.08
Log-likelihood = -15713
70 sec
70 sec
0.57
0.9
0.28
0.01
0.42
0.13
0.59
453 sec
140 sec
0.1
Log-likelihood = -21625
0.72
44 sec
0.6
536 sec
0.1
0.5
0.2
0.1
0.28
43 sec
43 sec
0.5
Log-likelihood=-24542
0.26
0.94
41 sec
41 sec
0.95
23 sec
0.7
0.05
0.04
220 sec
Log-likelihood = -145902

8.2 Fitting PH-distribution to customer patience (waiting time > 0)

Let us consider December positive waiting times for customers of regular type - PS, and for the same type, by LOW and HIGH priorities. Their histograms are given in Figure 8.26.

Figure 8.26: Distribution of positive waiting time.

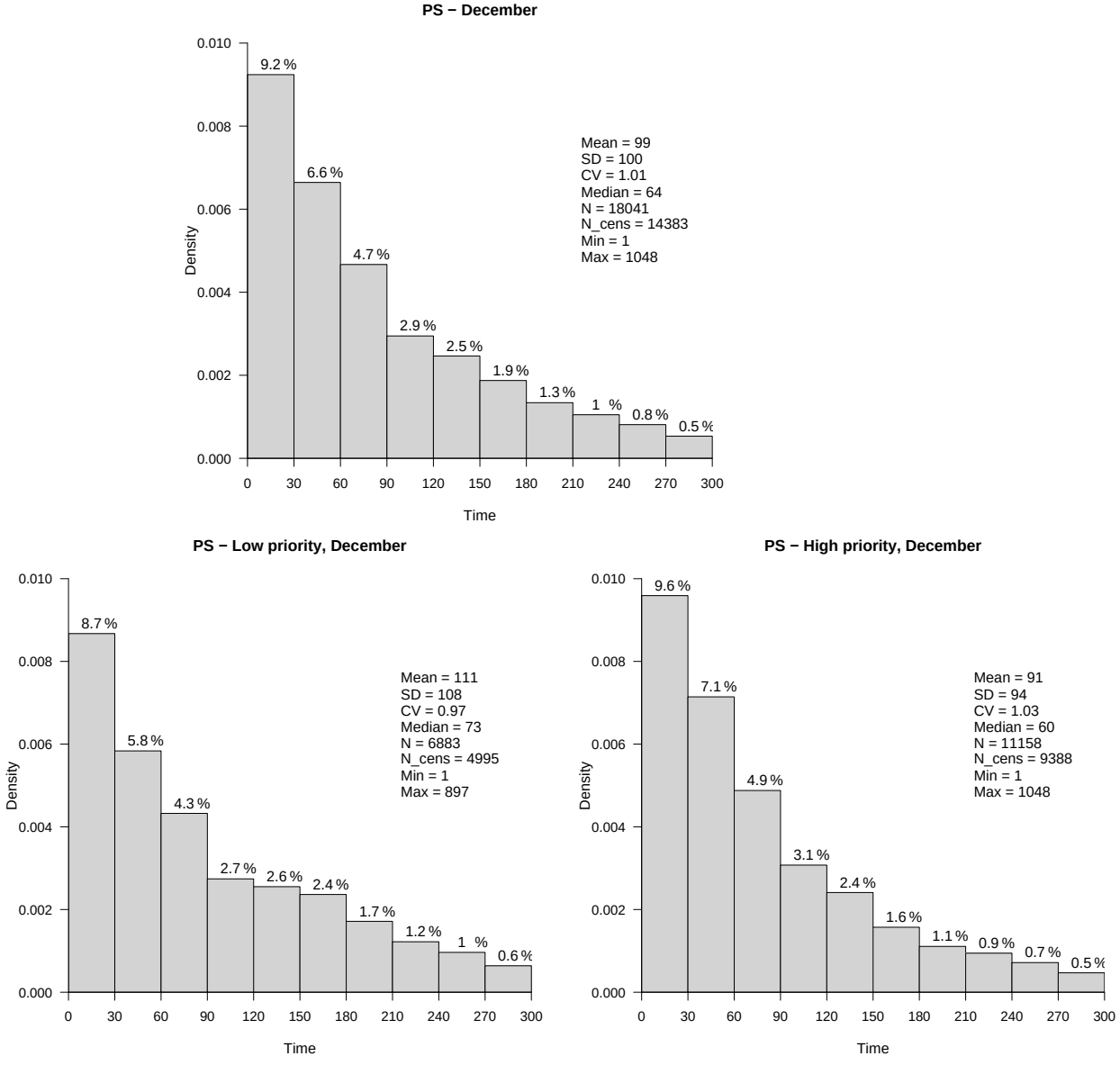
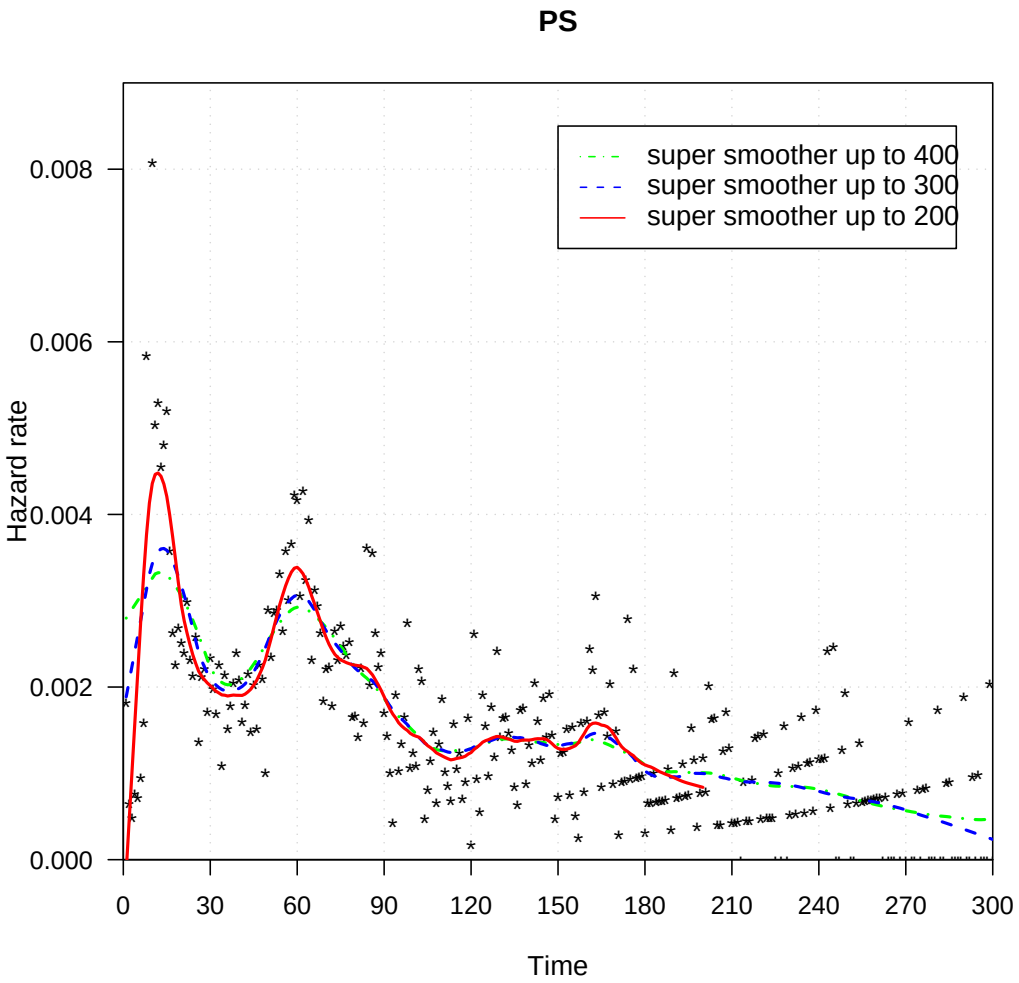


Figure 8.27 shows the smooth S-PLUS estimators of the hazard superimposed on the raw hazard rates, estimated with non-parametric methods (section 4.2), for service type PS. There are three smoothed hazard estimators that show peaks around 15 and 60 seconds, and smooth the raw hazard rates up to time 200, 300, 400 in order to distinguish clearly this interesting pattern at small times. These local peaks in the hazard rates of time willing to wait, manifest systematic tendency to abandon, while constant hazard rates indicates that the tendency to abandon remains the same.

Figure 8.27: Hazard rate for Patience - December.



There are also occasional peaks at other multiples of 60. This suggestes that some systematic phenomenon is lurking in the background during waiting [20]. And indeed, upon joining the queue, and about every minute or so thereafter, customers are exposed to an automatic message, which causes customers to abandon.

In Figure 8.28, the regular service customers are separated, according to priorities. The empirical hazard and survival functions of High and Low priorities for customers of type PS are presented. The hazard curves are derived by smoothing their raw rates up to time 200. There is stochastic ordering between overall PS, PS for High and Low priorities. The hazard and survival functions are placed in upside down order, because they are related according to the formula $H(t) = -\log_e S(t)$. High priority customers tend to be more patient then Low priority customers. According to the hazard rates, the probability of failure in small interval $[t, t + \triangle t]$, given that the customer has waited up to time t till now, is greater for customers with Low priority. Also, according to the survival functions, the probability not to hang up up to time t is greater for customers with High priority.

Figure 8.29 (p. 82) presents fitted PH-distributions of a general Coxian structures of order 20, 25, 30 for service type PS. The general Coxian structure is constructed as a sum of exponentials that can reach the absorbing state from all transient states and the Markov process is allowed to start in any transient state.

Table 8.15 presents the fitted PH-distribution mean (Mean), standard-deviation (SD), coefficient of variation (CV) and log-likelihood function (Log-L) of the fitted general Coxian structure of order $k = 20, 25, 30$ to the Patience - PS, December:

Table 8.15: Patience - PS, December. Statistics.

	k=20	k=25	k=30
Mean	762	1118	928
SD	662	1323	952
CV	0.87	1.18	1.03
Log-L	-25788	-25719	-25679

There is a censored version of EDF tests, according to [10]. However, as we have seen in section 8.1, in most cases the PH-model selected by EDF tests was the model selected by simultaneous confidence interval. Therefore, the simultaneous confidence band around the empirical cumulative distribution function is the sufficient criterion for model selection, implemented herein.

Figure 8.28: Patience - December - PS, PS for High and Low priorities.

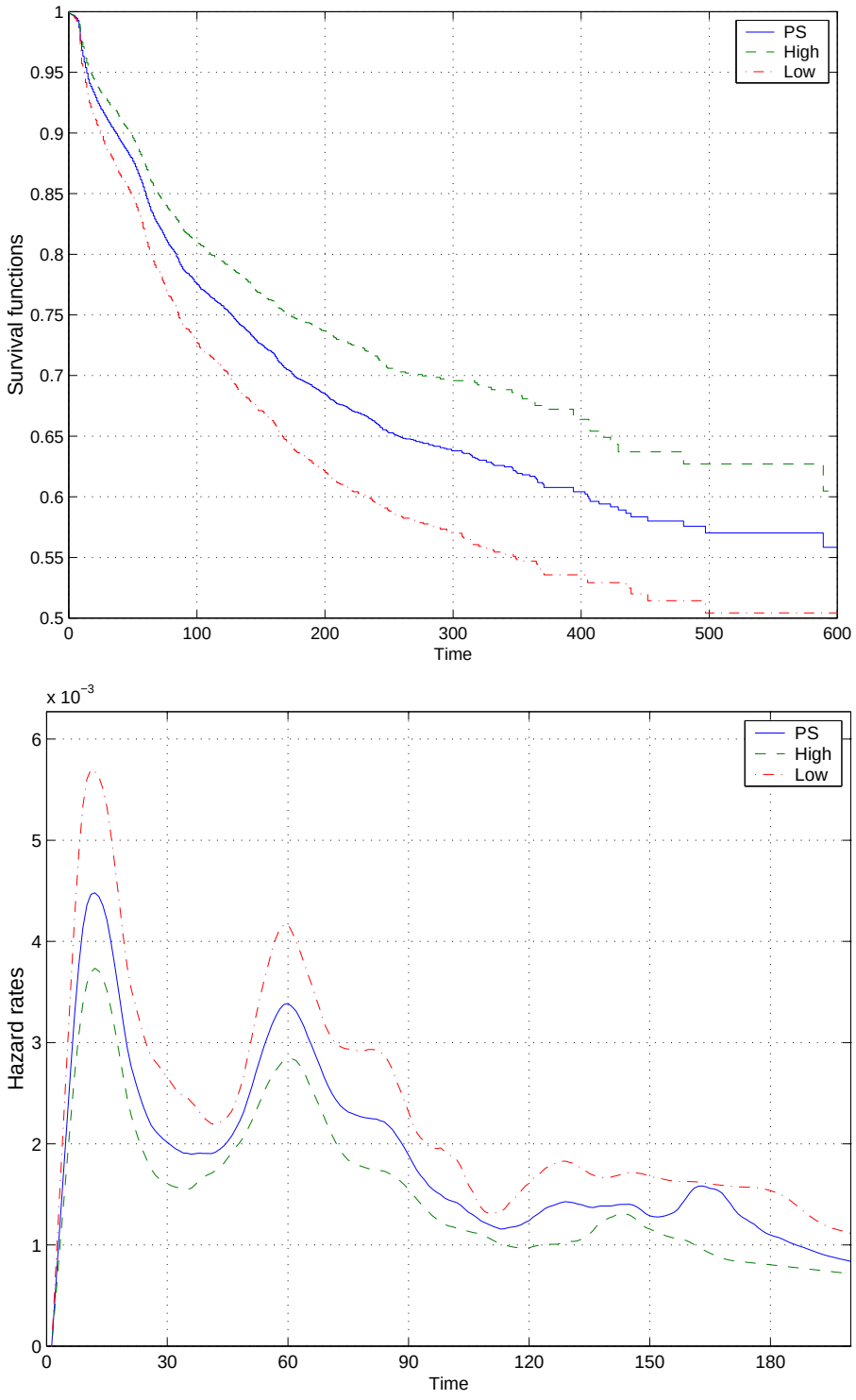
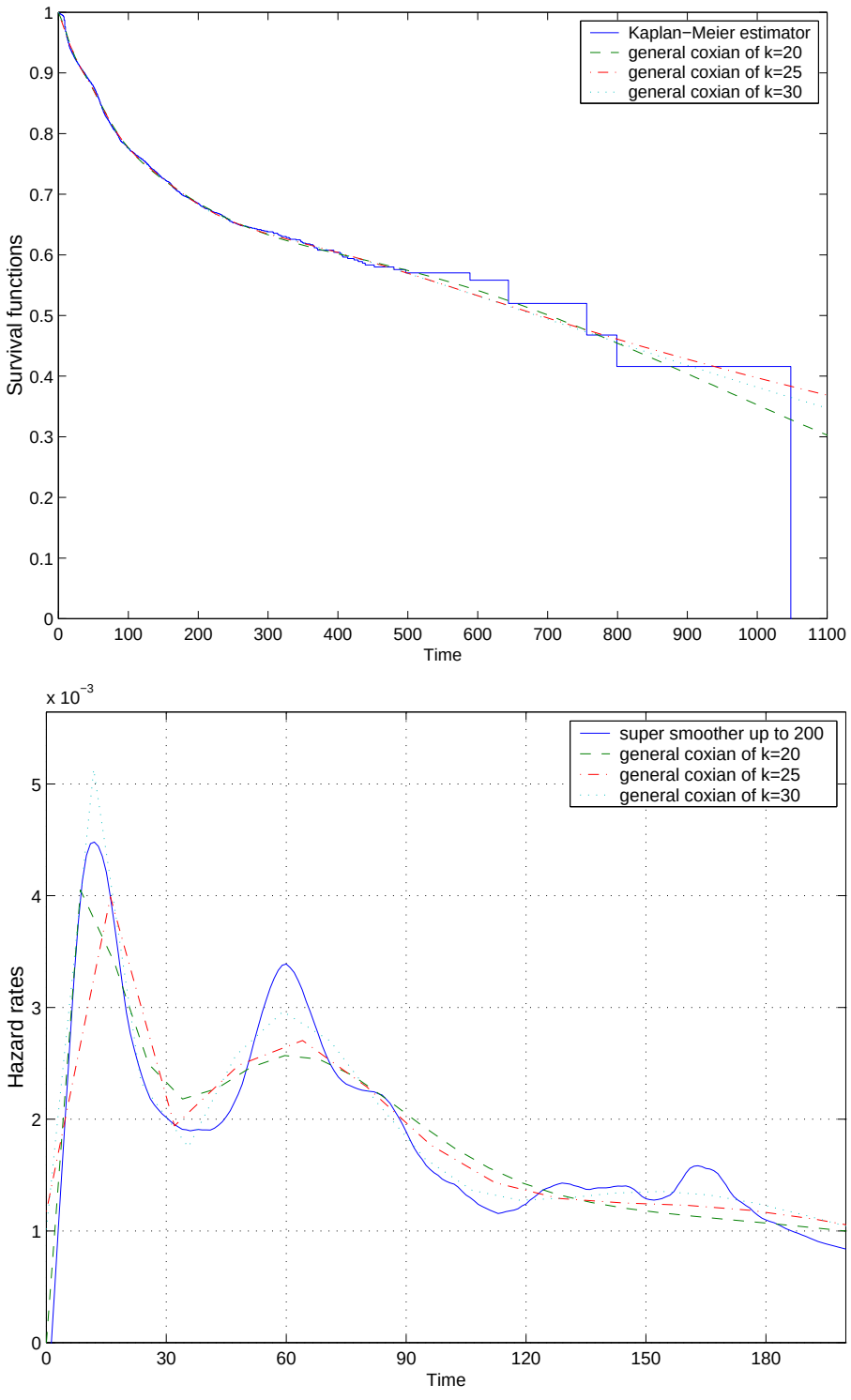


Figure 8.29: Phase-type fits of a general Coxian structure of order 20, 25, 30 to Patience - PS, December.



The fitted PH-models of general Coxian structure of order $k = 20, 25, 30$ all fit into a simultaneous confidence band ($\pm\frac{1}{\sqrt{n}}$) around empirical CDF in time interval $[10, 500]$. According to the hazard rates (Fig. 8.29), the most appropriate PH-model that fits patience is the PH-distribution of general Coxian structure of order $k = 30$.

The number of phases for estimation of patience is very large. In general, it is hard to induce rapid changes of the hazard rate, and it requires very high k-dimensions and a lot of "fast" states. This is especially so if the changes take place in the small time-interval, such as in plot of hazard rates, the two peaks at 15 and 60 seconds, while the overall time interval of patience is between 1 and 1048 seconds. Thus, the distribution of patience requires an approximation by PH-distributions of high order.

Figure 8.30 presents the derived structure of the fitted PH-distribution of general Coxian structure of order 30. The time from state 1 to state 4, until

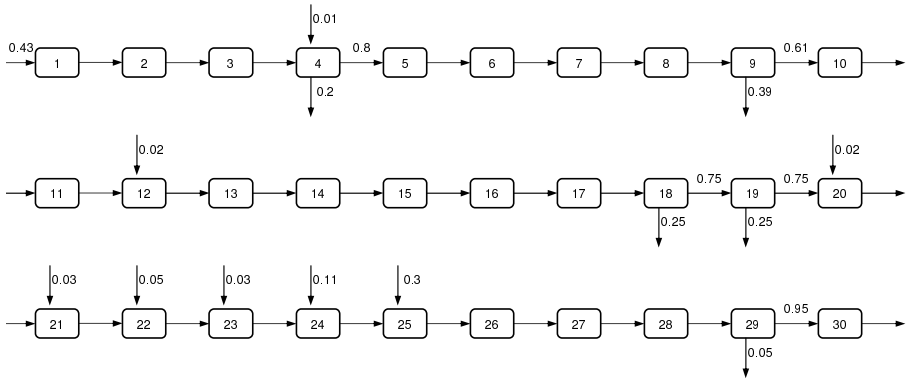


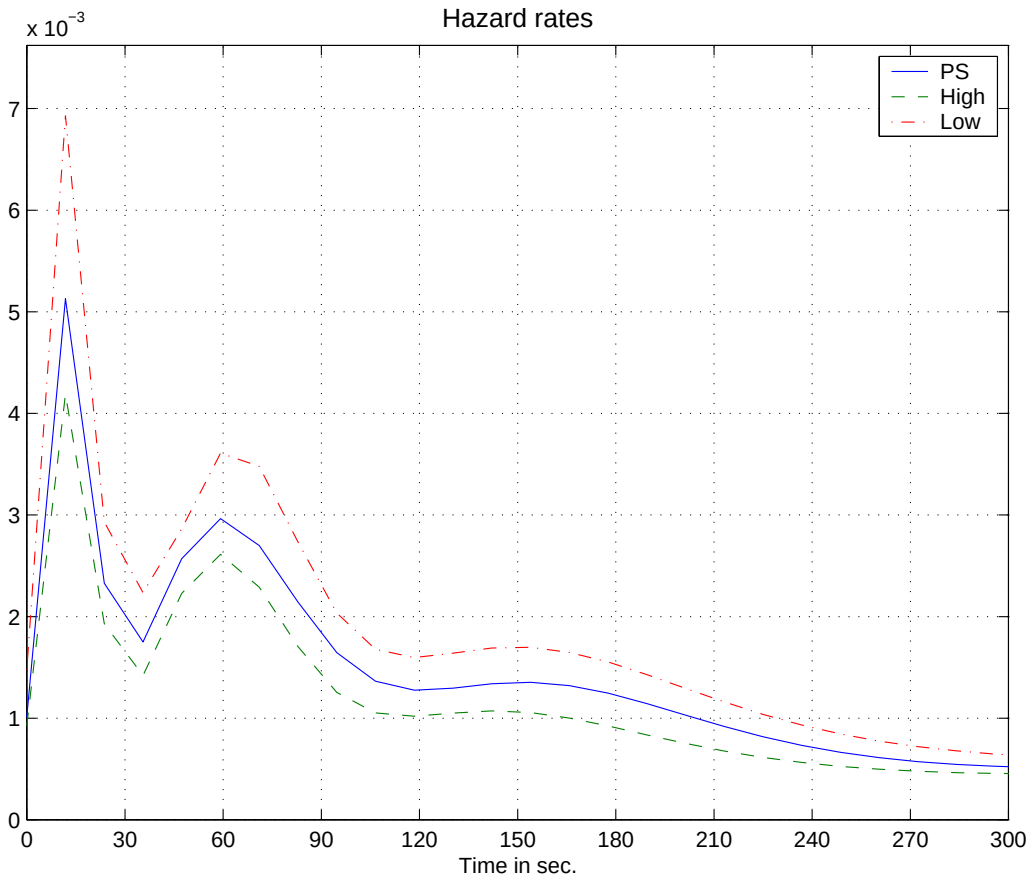
Figure 8.30: Patience - PS, December. The derived structure by fitting the general Coxian one of order $k = 30$.

first time of customer abandoning, is 15 seconds. The time to next abandonment happens from state 9, and the overall time till state 9 is about 68 seconds. The third time to abandonment happens at states 18 and 19, when the waiting time till state 19 is 180 seconds. The fourth time to abandonment happens at state 29, preceding the last state, with 834 seconds from state 1 to state 29. The last state, with the larger spending time - 705 seconds, is the state from which the underlying Markov process jumps to the absorption state with probability 1. Indeed, the derived PH-model describes the underlying process of waiting time in the queue and improves our understanding of the hazard rate, given in Figure 8.27 above.

Figure 8.31 presents the hazard rates of fitted general Coxian structure of

order 30 to the patience of PS type, PS - High and Low priorities. The pattern

Figure 8.31: Patience - PS, by priorities, December. There are hazard rates of fitted general Coxian structure of order $k = 30$.

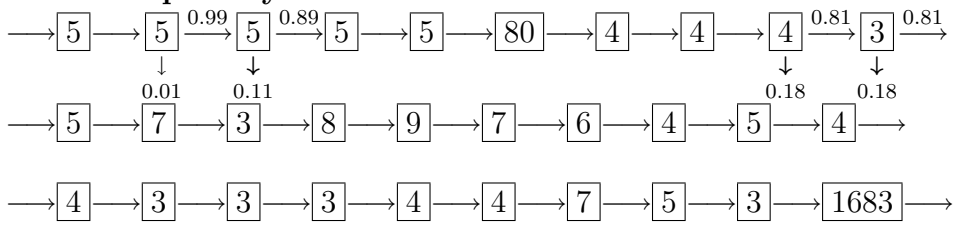


of stochastic ordering presented at Figure 8.31 can be seen more obviously by fitting Coxian structure of order 30 to the customers of PS type, PS - High and Low priorities than General Coxian structure of the same order. Figure 8.32 presents the derived structures of the fitted PH-distribution of Coxian structure of order 30, for customers of PS type, PS - High and Low priorities.

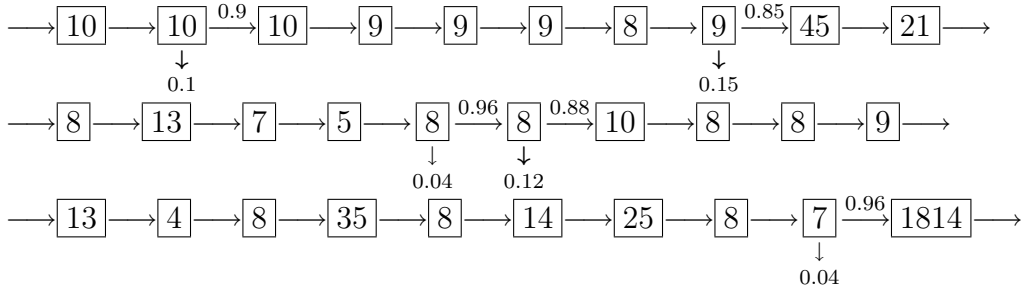
It emerges that from third phase there is abandonment of customers of High and Low priorities with probability 0.07 and 0.11, after 18 and 15 seconds, correspondingly, and from phase number 9, with probability 0.14 and 0.18, after 73 and 117 seconds, correspondingly. Customers of PS type abandon from second phase, after 20 seconds, with probability 0.1 and from

Figure 8.32: Patience - PS, December. The derived structures by fitting the Coxian one of order $k = 30$, by priorities.

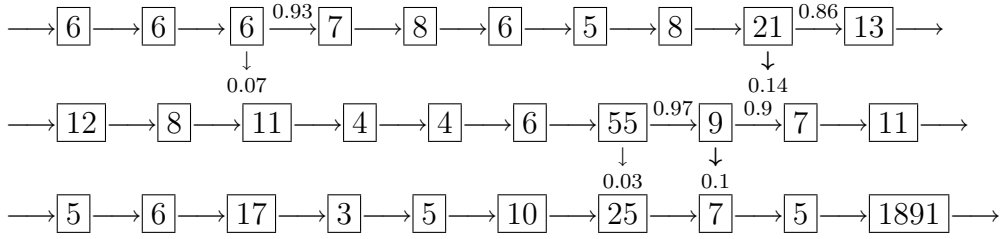
PS - Low priority:



PS:



PS - High priority:



phase number 8, after 74 seconds with probability 0.15. So the probability of abandonment happens in stochastic ordering. For example, after 15 - 20 seconds the customers of Low priority abandon with higher probability, 0.11, then there are the customers of Low and High priority together, with probability of abandonment 0.1, and thereafter the customers of High priority only, with probability of abandonment 0.07. According to the graphs of hazard rates, the fits of Coxian structure of order 30 are not very good as fits of General Coxian structure of the same order. Therefore, for Patience data the fit of General Coxian structure of order 30 is selected.

Chapter 9

Comparison between Log-normal and PH-type distributions

The service times are Log-normal distributed, according to Mandelbaum et al. [20]. Therefore, it is of interest to compare Phase-type with Log-normal distribution by minimizing of the distance of their density functions or Laplace transforms. Let us define this precisely.

9.1 Objective

The goal is

$$\min_{\boldsymbol{q},\boldsymbol{R}} \int_0^\infty (f_{PH}(y) - f_{LN}(y))^2 dy = \tag{9.1}$$
$$= \min_{\boldsymbol{q},\boldsymbol{R}} \int_0^\infty \left(\boldsymbol{q} \cdot \exp\{\boldsymbol{R}y\} \cdot \boldsymbol{r} - \frac{1}{\sigma y \sqrt{2\pi}} \exp\left\{ \frac{-(\ln y - \mu)^2}{2\sigma^2} \right\} \right)^2 dy,$$

for any specific order k of PH-distribution.

In general, there is no analytical way to derive such ($\boldsymbol{q}$,$\boldsymbol{R}$), the parameters of PH-distribution, which minimize the distance between the two distributions above. However, for any specific values of parameters of Log-normal distribution, (μ,σ), and for any specific order k of PH-distribution, we can derive numerically the optimal parameters of PH-distribution, ($\boldsymbol{q}$,$\boldsymbol{R}$), which minimize the integrand of quadratic difference of two densities: PH-type and Log-normal distributions. For this purpose, we use two methods of optimization, the `fmincon` function for constrained nonlinear minimization in Matlab

and the approximation of a theoretical density by a phase-type density, using EMpht-program (section 6.4).

Method of moments is implemented to compare Log-normal distribution with PH-distribution of specific order, k.

9.2 Method of Moments

The rth moment of Log-normal distribution [9] with parameters (μ, σ) is:

$$\mu_r = E(Y^r) = \exp\left\{r\mu + \frac{1}{2}r^2\sigma^2\right\}. \tag{9.2}$$

Consequently, the first moment, the variance and the coefficient of variation are:

$$E(Y) = e^{\mu+\frac{\sigma^2}{2}} \tag{9.3}$$

$$Var(Y) = e^{2\mu+\sigma^2}\left(e^{\sigma^2}-1\right) \tag{9.4}$$

$$CV(Y) = \sqrt{e^{\sigma^2}-1}. \tag{9.5}$$

The first and second moments of PH-distribution with parameters ($\boldsymbol{q}$,$\boldsymbol{R}$), according to formula (5.5) (section 5.2) are:

$$\mu_1 = -\boldsymbol{q}\boldsymbol{R}^{-1}\mathbf{1} \tag{9.6}$$

$$\mu_2 = 2\boldsymbol{q}\boldsymbol{R}^{-2}\mathbf{1}. \tag{9.7}$$

By comparing first moment and coefficient of variation of Log-normal and Phase-type distributions:

$$\mu = \ln(\mu_1) - \frac{\sigma^2}{2} \tag{9.8}$$

$$\sigma^2 = \ln\left(\frac{\mu_2}{\mu_1^2}\right). \tag{9.9}$$

Next, we introduce some examples of PH-distributions of specified structure are compared with Log-normal distribution.

9.2.1 Hyperexponential structure of order k

The density, first two moments and the square of the coefficient of variation of Hyperexponential distribution with k phases, exponentially distributed

with parameters λ_i, $i = 1, \ldots, k$, presented in figure 5.2 (p. 24), are:

$$f_Y(y) = \sum_{i=1}^{k} q_i \lambda_i e^{-\lambda_i y}, \quad y > 0, \quad \sum_{i=1}^{k} q_i = 1 \tag{9.10}$$

$$E(Y) = \sum_{i=1}^{k} \frac{q_i}{\lambda_i} \tag{9.11}$$

$$E(Y^2) = 2\sum_{i=1}^{k} \frac{q_i}{\lambda_i^2} \tag{9.12}$$

$$\vdots \tag{9.13}$$

$$E(Y^k) = k! \sum_{i=1}^{k} \frac{q_i}{\lambda_i^k}$$

$$CV^2(Y) = \frac{2\sum_{i=1}^{k} \frac{q_i}{\lambda_i^2}}{\left(\sum_{i=1}^{k} \frac{q_i}{\lambda_i}\right)^2} - 1 > 1 \tag{9.14}$$

The coefficient of variation of Hyperexponential distribution is greater than one. That is, for $\lambda_i > 0, \sum q_i = 1, i = 1, \ldots, k$,

$$\sum_{i=1}^{k} \frac{q_i}{\lambda_i^2} \geq \left(\sum_{i=1}^{k} \frac{q_i}{\lambda_i}\right)^2 \tag{9.15}$$

must holds. This is follows from the following. Let X be random variable defined by

$$P(X = \frac{1}{\lambda_i}) = \begin{cases} q_i, & i = 1, \ldots, k, \\ 0, & \text{otherwise.} \end{cases}$$

Then, the inequality (9.15) is implied by the relation $E(X^2) \geq (E(X))^2$.

By comparing first moment and coefficient of variation of Log-normal and Phase-type distributions:

$$\mu = \ln\left(\sum_{i=1}^{k} \frac{q_i}{\lambda_i}\right) - \frac{\sigma^2}{2} \tag{9.16}$$

$$\sigma^2 = \ln\left[\frac{2\sum_{i=1}^{k} \frac{q_i}{\lambda_i^2}}{\left(\sum_{i=1}^{k} \frac{q_i}{\lambda_i}\right)^2}\right] \tag{9.17}$$

9.2.2 Generalized Erlang structure of order k

There is Generalized Erlang distribution with k phases, exponentially distributed with parameters λ_i, $i = 1, \ldots, k$:

$$\longrightarrow \boxed{\lambda_1} \longrightarrow \boxed{\lambda_2} \longrightarrow \cdots \longrightarrow \boxed{\lambda_k} \longrightarrow$$

The first two moments and the square of the coefficient of variation are:

$$E(Y) = \sum_{i=1}^{k} \frac{1}{\lambda_i} \tag{9.18}$$

$$E(Y^2) = \sum_{i=1}^{k} \frac{1}{\lambda_i^2} + \left(\sum_{i=1}^{k} \frac{1}{\lambda_i} \right)^2 \tag{9.19}$$

$$CV^2(Y) = \frac{\displaystyle\sum_{i=1}^{k} \frac{1}{\lambda_i^2}}{\left(\displaystyle\sum_{i=1}^{k} \frac{1}{\lambda_i} \right)^2} < 1 \tag{9.20}$$

The coefficient of variation of Generalized Erlang distribution is smaller than one.

Then, by comparing first moment and coefficient of variation of Log-normal and Phase-type distributions:

$$\mu = \ln \left(\sum_{i=1}^{k} \frac{1}{\lambda_i} \right) - \frac{\sigma^2}{2} \tag{9.21}$$

$$\sigma^2 = \ln \left[\frac{\displaystyle\sum_{i=1}^{k} \frac{1}{\lambda_i^2}}{\left(\displaystyle\sum_{i=1}^{k} \frac{1}{\lambda_i} \right)^2} + 1 \right] \tag{9.22}$$

9.2.3 Log-normal Model for Call-Center Service-Times

Let Y be a random variable, denoting service time such that $Y \sim$ Log-normal(μ, σ^2). Then, $X = \ln(Y)$ is normally distributed with mean μ and variance σ^2. The parameters of Log-normal distribution, estimated according

to maximum likelihood estimation method, are as follows:

$$\hat{\mu} = \frac{\sum_{i=1}^{n} \ln y_i}{n} \tag{9.23}$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} (\ln y_i - \hat{\mu})^2}{n-1} \tag{9.24}$$

The estimators of the expectation, the standard deviation and coefficient of variation of service time are computed according to formulas (9.4) - (9.5) with estimated μ and σ above.

Overall Service time - December is approximately Log-normal($\mu = 4.8, \sigma = 1.03$), see Figure 9.1 (p. 91).

Figure 9.2 (p. 92) shows an approximation of Log-normal distribution with $\mu = 4.8$ and $\sigma = 1.03$ by general structure of order $k = 3$. This Log-normal distribution has an expectation about 207 seconds, standard-deviation about 284, and consequently, coefficient of variation of about 1.37. In EMpht-program the distribution is truncated at 2000. As truncation point gets larger, the better is the phase-type fit at the tail.

Figure 9.3 (p. 92) presents the derived fitted phase-type structures of order $k = 3$, by different truncation points. By truncation the distribution at point larger than 2000, we derive that the holding time at third phase (with the larger time in Figure 9.3) is larger, and the absorption probability is larger too. Consequently, the fit is better, because it coincides with the hazard rate of Log-normal distribution in the tail too. There is a similarity between these structures and the structure derived by fitting PH-distribution of the same order to the data - Service times, December (see Fig. 8.5 a), p. 46).

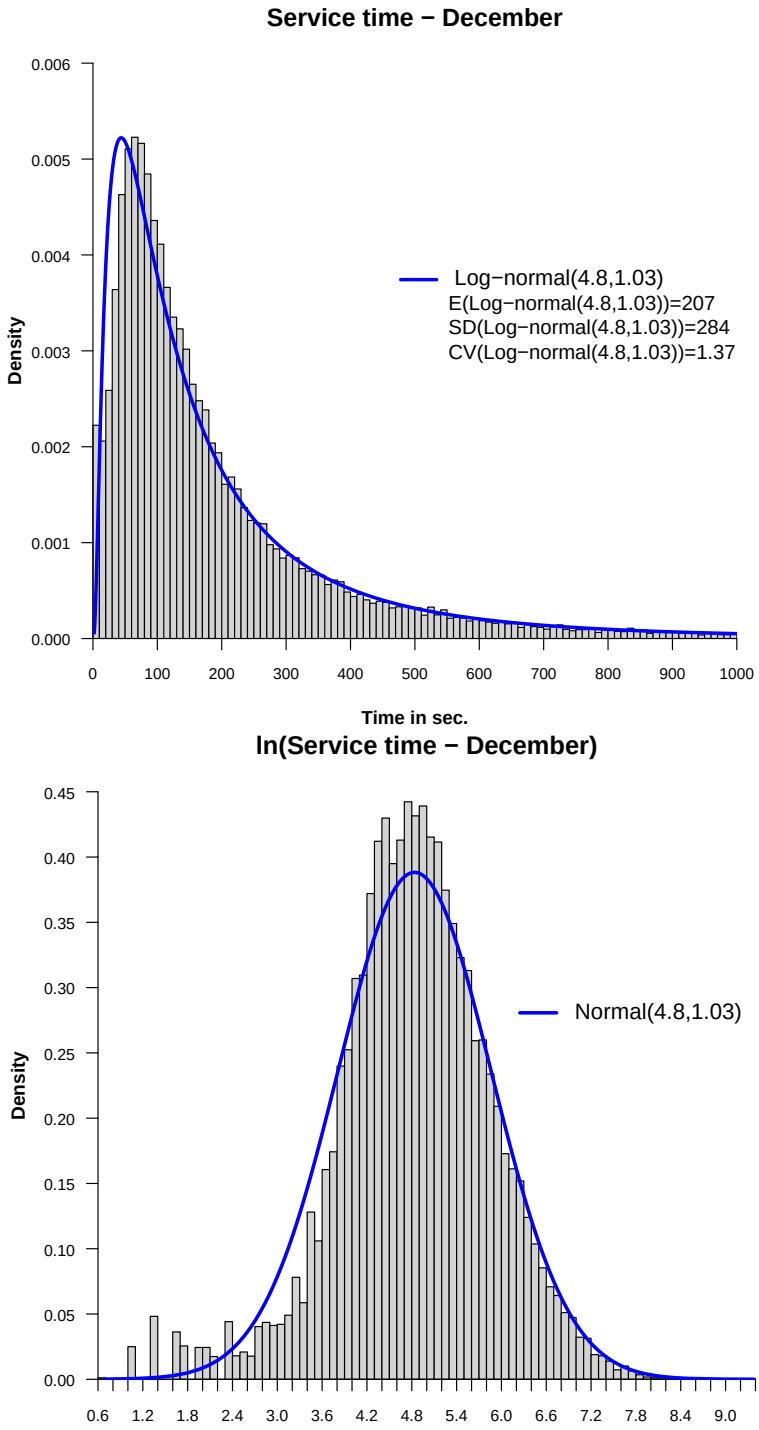


Figure 9.1: In the top plot, histogram of service time versus Log-normal density with parameters $\mu = 4.8$ and $\sigma = 1.03$. In the bottom plot, histogram of ln(service time) versus Normal density.

Figure 9.2: Fitted PH-distribution of order $k = 3$ (dashed line) together with Log-normal distribution with parameters $\mu = 4.8$ and $\sigma = 1.03$, $E(LN) = 207$, $SD(LN) = 284$, $CV(LN) = 1.37$ (solid line).

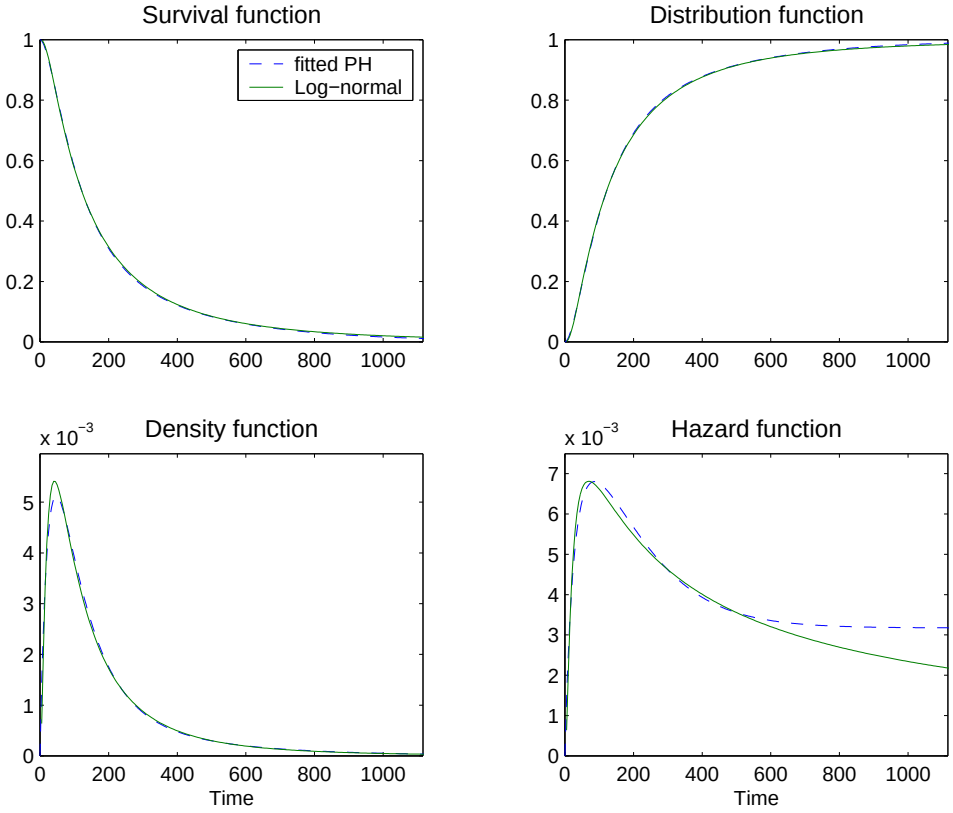
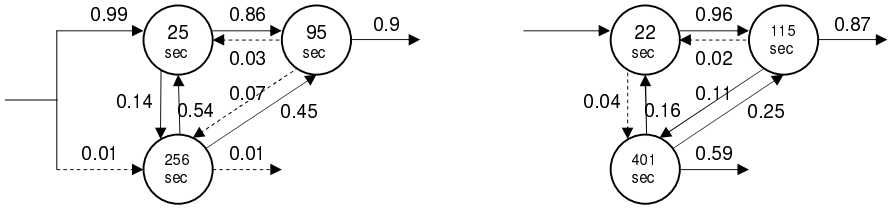


Figure 9.3: Fitted PH-structure of order $k = 3$ for Log-normal distribution with parameters $\mu = 4.8$ and $\sigma = 1.03$.



Truncation point = 2000
Log-likelihood = -6.205290
Fitted mean = 198
Fitted standard-deviation = 230
Fitted coefficient of variation = 1.16

Truncation point = 5000
Log-likelihood = -6.250258
Fitted mean = 206
Fitted standard-deviation = 261
Fitted coefficient of variation = 1.27

9.3 Optimization methods

9.3.1 Constrained optimization, using Matlab

Using Matlab, the optimal parameters of PH-distribution of order k, ($\boldsymbol{q}$,$\boldsymbol{R}$), which minimize the distance between Log-normal and Phase-type densities, are obtained by:

$$\min_{\boldsymbol{q},\boldsymbol{R}} \int_0^{100} \left(\boldsymbol{q} \cdot \exp\{\boldsymbol{R}y\} \cdot \boldsymbol{r} - \frac{1}{\sigma y \sqrt{2\pi}} \exp\left\{ \frac{-(\ln y - \mu)^2}{2\sigma^2} \right\} \right)^2 dy, \quad (9.25)$$

for known μ and σ.

$$Distance = \int_0^{100} \left(\boldsymbol{q} \cdot \exp\{\boldsymbol{R}y\} \cdot \boldsymbol{r} - \frac{1}{\sigma y \sqrt{2\pi}} \exp\left\{ \frac{-(\ln y - \mu)^2}{2\sigma^2} \right\} \right)^2 dy, \quad (9.26)$$

for known (μ,σ) and optimal ($\boldsymbol{q}, \boldsymbol{R}$).

9.3.2 Minimizing the information divergence, using EMpht

Using the EMpht-program, an approximation of Log-normal distribution by a PH-distribution is done by minimizing the information divergence (the Kullback-Leibler information) (see section 6.4), that is:

$$\max_{\boldsymbol{q},\boldsymbol{R}} \int_0^{100} \log\left(\boldsymbol{q} \cdot \exp\{\boldsymbol{R}y\} \cdot \boldsymbol{r}\right) \cdot \left(\frac{1}{\sigma y \sqrt{2\pi}} \exp\left\{ \frac{-(\ln y - \mu)^2}{2\sigma^2} \right\} \right) dy, \quad (9.27)$$

for known μ and σ.

9.3.3 Comparison of the two optimization methods

Table 9.1 presents the results of optimization methods above for three cases: Log-normal($\mu = 1, \sigma = 0.5$), Log-normal($\mu = 1, \sigma = 1$) and Log-normal($\mu = 0, \sigma = 1$).

The last row in each table above, ($k = 5^*$), presents the results of constrained optimization in Matlab. There are obtained by using the derived EMpht-results for $k = 5$ as an initial point to Matlab.

It seems that the EMpht-program gives better results than Matlab. Besides, the optimization, using Matlab, is very time-consuming.

Table 9.1: Comparison between two optimization methods.

$\mu = 1$,	Distance		CV(LN)=0.53		E(LN)=3.08		SD(LN)=1.63	
$\sigma = 0.5$	Matlab	EMpht	Matlab	EMpht	Matlab	EMpht	Matlab	EMpht
$k = 2$	0.0321	0.0335	0.71	0.71	3.32	3.08	2.35	2.18
$k = 3$	0.0098	0.0099	0.58	0.58	3.04	3.08	1.76	1.78
$k = 4$	0.0023	0.0028	0.51	0.54	2.94	3.08	1.49	1.65
$k = 5$	0.0022	0.0006	0.51	**0.53**	2.95	**3.08**	1.51	**1.63**
$k = 5^*$	0.0004		0.53		3.02		1.59	

$\mu = 1$,	Distance		CV(LN)=1.31		E(LN)=4.48		SD(LN)=5.87	
$\sigma = 1$	Matlab	EMpht	Matlab	EMpht	Matlab	EMpht	Matlab	EMpht
$k = 2$	0.0009	0.0173	0.88	1.29	3.57	4.46	3.15	5.74
$k = 3$	0.0009	0.0006	0.89	1.22	3.58	4.46	3.15	5.46
$k = 4$	0.0013	0.0004	0.91	1.26	3.67	4.46	3.32	5.62
$k = 5$	0.0007	0.0004	0.91	1.26	3.69	4.46	3.37	5.64
$k = 5^*$	0.0004		1.20		4.35		5.23	

$\mu = 0$,	Distance		CV(LN)=1.31		E(LN)=1.65		SD(LN)=2.16	
$\sigma = 1$	Matlab	EMpht	Matlab	EMpht	Matlab	EMpht	Matlab	EMpht
$k = 2$	0.0437	0.0469	1.00	1.31	1.63	1.65	1.63	2.16
$k = 3$	0.0437	0.0019	1.00	1.24	1.63	1.65	1.63	2.04
$k = 4$	0.0011	0.0013	1.07	1.29	1.53	1.65	1.64	2.12
$k = 5$	0.0011	0.0013	1.05	**1.31**	1.52	**1.65**	1.58	**2.15**
$k = 5^*$	0.0011		1.16		1.58		1.83	

Conclusions

In this research we have estimated the distribution of service time and patience using the empirical call-center data of one of Israel's banks. The main steps in this research were:

- Phase-type distributions of different order and structure were used to fit empirical data as well as other theoretical distributions.
- The parameters of Phase-type distributions were computed via the EM-algorithm, using the EMpht-program.
- The empirical survival, density and hazard functions were plotted versus the fitted functions to examine visually the qualitative differences.
- The simultaneous confidence interval for empirical CDF was used as a heuristic stopping rule for adding phases of the fitted PH-distribution.
- We used the Kolmogorov-Smirnov and Anderson-Darling goodness-of-fit tests to evaluate quantitative aspects of the produced fits.
- We have compared Phase-type with Log-normal distributions using the method of moments.
- The optimal parameters of the PH-distribution were numerically found for the specific parameters of Log-normal distribution and for the given order of PH-distribution by two methods. We succeeded to approximate the Log-normal distribution by Phase-type distribution of order $k = 3$.

PH-distributions of order $k = 2, 3, 4, 5, 6$ were used to fit the service durations of call-center data, for different priorities and service-types. According to statistic D, which is more robust to the size of the sample than statistics D^* and A^2, the general structure of order $k = 3$ already provides a reasonable fit to the overall service time, for December (see Table 8.2, p.50). Moreover, the fitted Coxian structure of the same order has the same log-likelihood function and, therefore, its fitted density, survival and hazard functions coincide

with the fitted corresponding functions of the general phase-type structure. In view of the fact that PH-distributions have non-unique representation, it is difficult to give a physical interpretation to the phases. According to Figure 8.9, in every structure of specific order, there exists the phase with a longer length time, which plausibly corresponds to the customers with a longer service time. Figures 8.10, 8.16 demonstrate histograms of service time - December, by priorities and four main service types. We marked the peaks that imply a higher percentage of customers that depart from the service at corresponding times. From Figures 8.11, 8.17 we note the stochastic ordering between the priorities and the service types.

The PH-model that provides a perfect fit to the patience is the PH-model of order $k = 30$ of general Coxian structure. As can be seen from Figure 8.25, which shows the hazard rates, there are two peaks around 15 and 60 seconds. These peaks take place within a small time-interval, while the overall time-interval is [1,1048]. Therefore, it requires very high k-dimensional fit of PH-distribution. Figure 8.29 presents the hazard rates of the fitted PH-model of order 30 of general Coxian structure, for HIGH and LOW priorities. The pattern of stochastic ordering can be noted once again. It follows that HIGH priority customers are more patient. The derived Coxian structures of order 30 in Figure 8.30 demonstrate it too.

Table 9.1 presents several measures of proximity of PH-distribution to Log-normal one. Namely, for specific parametric values of Log-normal distribution, (μ,σ), and for any given order k of PH-distribution, we have derived the optimal parameters of PH-distribution, ($\boldsymbol{q}$,$\boldsymbol{R}$). Then we have calculated the distance, the mean, the standard deviation and the coefficient of variation of PH-distribution with optimal parameters as above. We have compared the performances of two optimization methods: the minimization of the integral of quadratic difference of corresponding densities via constrained nonlinear minimization in Matlab and the minimization of information divergence, using EMpht-program. The results obtained by EMpht-program are better than that of Matlab.

Figure 8.32 (in the top plot) shows histogram of overall service time for December versus Log-normal density with parameters $\mu = 4.8$ and $\sigma = 1.03$ that are derived by maximum likelihood estimation. In the bottom plot, the histogram of ln(service time) versus Normal density is presented. Using EMpht-program, we succeed to approximate this Log-normal distribution by the Phase-type distribution of order $k = 3$, see Figure 9.1. As can be seen from Figure 9.2, the larger the upper truncation point is, which is specified together with parameters of the theoretical density to be fitted, the better the PH-distribution fits the theoretical one.

The analysis of the data from the call center of "Anonymous Bank" contributes to understanding of the underlying processes describing the service and the behavior of customers by modelling their patience. The results of our analysis can be used to optimize the call center efficiency as well as the customer-service quality.

Possible directions of future research:

- It is important to develop advanced models for patience distribution that explain the two sharp peaks of the hazard rate. For example, one could consider a mixture of PH-distribution with a small number of phases and two distributions with a small variance that is "responsible" for the peaks.
- It is of interest to analyze the data from other call centers in order to compare its functionals and appropriate mathematical models to our findings and models, originated in the "Empirical Analysis of a Call Center" by Mandelbaum et al. [20] and herein.
- As discovered in this research, the PH-distribution of order $k = 3$ already provides a reasonable fit to the service time. It is recommended to investigate the physical interpretation of the phases of service, which requires consultation with the managers at "Anonymous Bank", or perhaps even a field study.

Bibliography

[1] O.O. Aalen. Modelling heterogeneity in survival analysis by the compound poisson distribution. *The Annals of Applied Probability*, 2:951 – 972, 1992.

[2] O.O. Aalen. On phase-type distributions in survival analysis. *Scandinavian Journal of Statistics*, 22:447 – 463, 1995.

[3] O.O. Aalen and H.K. Gjessing. Understanding the shape of the hazard rate: A process point of view. *Statistical Science*, 16:1 – 22, 2001.

[4] S. Asmussen. *Applied Probability and Queues*. John Wiley & Sons, 1987.

[5] S. Asmussen, O. Nerman, and M. Olsson. Fitting phase-type distributions via the EM algorithm. *Scandinavian Journal of Statistics*, 23:419 – 441, 1996.

[6] G.R. Bitran and S. Dasu. Analysis of the $\Sigma Ph_i/Ph/1$ queue. *Operations Research*, 42:158 – 174, 1994.

[7] E. Chlebus. Empirical validation of call holding time distribution in cellular communications systems. 1997.

[8] D.R. Cox and D. Oakes. *Analysis of Survival Data*. Chapman and Hall, 1984.

[9] E.L. Crow and K. Shimizu. *Lognormal Distributions*. Marcel Dekker, 1988.

[10] R.B. D'Agostino and M.A. Stephens. *Goodness-of-fit techniques*. Marcel Dekker, 19.

[11] O. Häggström, S. Asmussen, and O. Nerman. EMPHT – a program for fitting phase-type distributions. Technical report, Department of Mathematics, Chalmers University of Technology, Göteborg, 1992.

[12] S. Kang and R.F. Serfozo. Parallel-processing times: Extreme values of phase-type and mixed random variables.

[13] D.G. Kleinbaum. *Survival Analysis*. Springer, 1996.

[14] E.T. Lee. *Statistical Methods for Survival Data Analysis*. John Wiley & Sons, 2rd edition, 1992.

[15] E.L. Lehmann. *Theory of point estimation*. New Jork: Wiley, 1983.

[16] Call Center Magazine. `http://www.callcentermagazine.com`, WEB site.

[17] A. Mandelbaum. *Service Engineering of Stochastic Networks. Background, with a focus on Tele-Services*. Technion, Israel Institute of Technology.

[18] A. Mandelbaum. *Call Centers. Research Bibliography with Abstracts*. Technion, Israel Institute of Technology, 2rd edition, September 2001.

[19] A. Mandelbaum. Hazard rate functions. Technical report, April 2001.

[20] A. Mandelbaum, A. Sakov, and S. Zeltyn. Empirical analysis of a call center. Technical report, Technion, Israel Institute of Technology, 2000.

[21] R. Nelson. *Probability, Stochastic processes, and Queueing theory*. New-Jork, Springer, 1995.

[22] M.F. Neuts. *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach*. The Johns Hopkins University Press, Baltimore, 1981.

[23] C.A. O'Cinneide. On non-uniqueness of representations of phase-type distributions. *Commun.Statist.-Stochastic Models*, 5:247 – 259, 1989.

[24] C.A. O'Cinneide. Characterization of phase-type distributions. *Commun.Statist.-Stochastic Models*, 6:1 – 57, 1990.

[25] M. Olsson. Estimation of phase-type distributions from censored data. *Scandinavian Journal of Statistics*, 23:443 – 460, 1996.

[26] M. Olsson. The EMpht-programme. Technical report, Department of Mathematics,Chalmers University of Technology, and Göteborg University, 1998.

[27] V.K. Rohatgi. *An Introduction to Probability Theory and Mathematical Statistics*. John Willey & Sons, 1976.

[28] B. Silverman. *Density estimation*. Chapman and Hall, 1986.

[29] Call Center Statistics. `http://www.callcenternews.com/resources/statistics.shtml`, WEB site.

[30] M.A. Tanner. *Tools for Statistical Inference*. Springer, 3rd edition, 1996.

[31] W. Venables and B. Ripley. *Modern applied statistics with S-plus*. Springer, 3rd edition, 1999.

[32] E. Zohar, A. Mandelbaum, and N. Shimkin. Adaptive behavor of impatient customers in tele-queues: Theory and empirical support. 2000.