A Probabilistic Model to Predict the Time Out of Service for Electronic Devices

Eralda Gjika (Dhamo)
Department of Applied Mathematics
Faculty of Natural Science, UT
eralda.dhamo@fshn.edu.al

Lule Basha (Hallaci)
Department of Applied Mathematics
Faculty of Natural Science, UT
lule.hallaci@fshn.edu.al

Ana Ktona
Department of Informatics
Faculty of Natural Science, UT
ana.ktona@fshn.edu.al

Abstract

Time modeling is one of the challenges attracting the attention of researchers in different fields. The prediction of the working time and out of service of an electronic device is important as it helps to optimize the management of human resource, material resource as well as monetary resources of the company. So, determining a predictive probability model of time will help in minimizing maintenance costs, storage of equipments and increasing efficiency in the service. In this study we have analyzed the waiting time for repair and the out of service time of some electronic equipments. The data are taken from an electronic service center in Tirana which offers service for electronic devices such as: computers, laptops, cellphone, tablets etc. for a period of one year. Probability distributions such as: normal, exponential, weibull, log-normal, gamma and pareto are fitted to the real data using maximum likelihood parameter estimation method. Various graphical and numerical statistical tests are performed to choose the "best" fit to the real data. The chosen probabilistic model helps the company to predict service time and design a maintenance strategy to optimize cost and customer satisfaction at the same time.

1. Introduction

Queue theory is generally considered as an area of operational research because the results are often used when making business decisions about the resources needed to provide a service. The service disciplines are of different types, such as: The First In First Out (FIFO); Last-in-First Out - LIFO; Customers Served Parallel (Processor Sharing - PS); The highest priority client is served first (Priority -P) etc. [Szt16].

In real life there is often a need to estimate the time spent in a queue to benefit a service. The time spent in a queue, which includes the time from the moment of arrival to the system until the time of departure from the system may be constant for any request arriving. In real situations the time spent in the system is not always constant. Since service discipline affect external factors (service provider effectiveness, type of service required, etc.) then this lead to changes in waiting time. Figure 1 shows a situation of arrival times and waiting time in a queue. This pattern is known as "non-regular traffic queue".

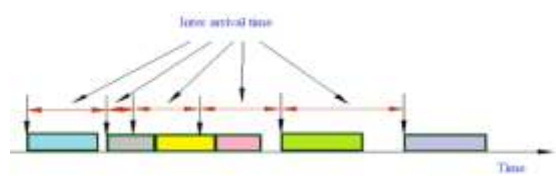


Figure 1: Non-regular traffic queue

In this work we will study the flow of requests arriving at the laboratory of a company which offer service for electronic device. The company has only one laboratory providing the service for electronic devices so it requires the most efficient management of the time.

Service requests arrive at the laboratory following a FIFO discipline. In rare situations the company has the right to offer priority service. Figure 2 illustrates the steps followed by an arrival service request.

Figure 2: Arrival service request scheme

Figure 3 shows the scheme of the system. It shows the waiting time to enter the lab and the time that the device stays in the service center (known as out of service time of the device).

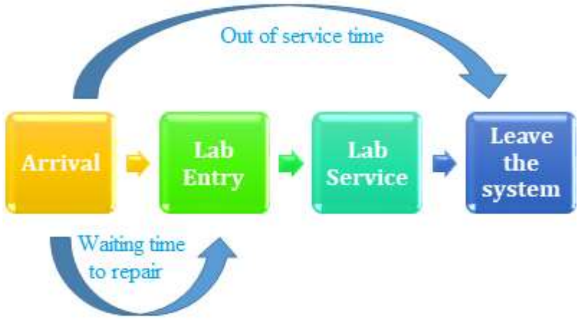


Figure 3: The system scheme of waiting time to repair and out of service of the devices

Waiting time to enter the lab and the service time distribution is one part of the company system [Wan17]. The ability of the servers to serve the queries of their costumers will determine the system performance. The faster the servers, the better the system performance. The company has a line queue system compound of one server.

We have tried to fit a probability model for the waiting time to enter the lab and the out of service time of the electronic equipment's arriving at the service center. Given that the time is a continuous variable we have taken into consideration some continuous distributions such as: *normal; log-normal; gamma; Weibull; exponential and pareto* distribution [Mul15]. The procedure of parameter estimation, evaluation and simulation was done through R software.

2 Data

We have considered two time data: the waiting time from the moment the device enters in the service center until it goes at the lab (known as the waiting time to be repaired) and the out of service time which is the time spent at the service center (from the moment it comes and the moment it goes out of the service center). There are in total 292 observations for the first database (data observed for 1 year) and 330 observations for the second database. The unit time measure is hour.

For each costumer the system has: an ID, the time it enters the service center, the time it enters the lab and the time it goes back to the owner. Unfortunately, the company didn't note the time the device gets out of the lab. This time is registered in the program lately.

We also have to emphasize that there is a lack of information in the secured database, and we have not considered the devices which had missing information in the two databases. Table 1 and Table 2 below give a descriptive statistics for the two databases.

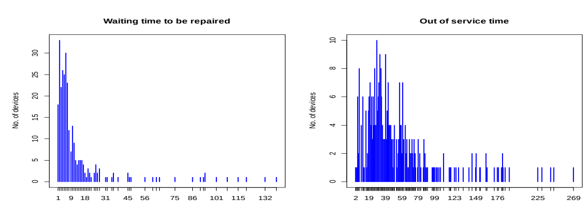
Table 1: Descriptive statistics for waiting time to be repaired data

n	Min	Q1	Me	Mean	sd	Q3	Max
292	1.00	3.75	6.00	13.21	21.75	12.00	139

Table 2: Descriptive statistics for out of service time data

n	Min	Q1	Me	Mean	sd	Q3	Max
330	2.00	26.00	39.50	53.45	45.5	64.00	269

To better understand the behavior of the data we obtain a graphical view of the two dataset. The histogram of the waiting time to be repaired and out of service time of the electronic devices is shown in Figure 4 (a, b) respectively.



(a) (b)

Figure 4: Waiting time to repair and out of service time data

3 Parameter estimation

3.1 Estimation

The probability distributions we have considered are continuous distributions: *normal; log-normal; gamma; Weibull; exponential and pareto* distribution. To estimate the parameter of these distributions we have used the maximum likelihood estimation (MLE) method. We have used the R software for estimation procedure and also for statistical test. The packages used in R are: *AdequacyModel* [Dut08]*; MASS* [Ven10]*; fitdistrplus; actuar*. [Boos12], [Hur13].

Table 3 shows the pdf of the continuous distributions used to fit the data.

Table 3: Probability distribution function of Continuous distributions

Distribution	Probability density function
Exponential	$f(x) = \lambda e^{-\lambda x}, x \geq 0$
Normal	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, x \in R$
Log-Normal	$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}}, x > 0$
Gamma	$f(x) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}, x > 0; \alpha, \beta > 0$
Weibull	$f(x) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-(\frac{x}{\lambda})^k}, x \geq 0; \lambda, k > 0$
Pareto	$f(x) = \frac{ab^a}{x^{a+1}}, x \geq b$

3.2 Model Evaluation

The evaluation procedure for the proposed probability distributions was done through graphical statistical test and numerical test. Among the test we have considered the: QQ-plot, PP-plot, density plot to compare the efficiency of each proposed distributions with the real data.

When fitting continuous distributions, three goodness-of-fit statistics are classically considered to choose the “best” fitted distribution: Cramer-von Mises, Kolmogorov-Smirnov and Anderson-Darling statistics [D’Ag86]. Table 4 below gives the definition and the empirical estimate of the three considered goodness-of-fit statistics. Other accuracy measures used were the Information criteria such as: the Akaike Information Criteria (AIC) and Bayes Information Criteria (BIC).

Table 4: Goodness-of-fit statistics as defined by D’Agostino and Stephens (1986)

General formula	Computational formula
Kolmogorov-Smirnov(KS)	
$\sup \left\| F_n(x) - F(x) \right\|$	$\max(D^+, D^-)$ with $D^+ = \max_{i=1,\dots,n}\left(\frac{i}{n} - F_i\right)$ $D^- = \max_{i=1,\dots,n}\left(F_i - \frac{i-1}{n}\right)$
Cramer-von Mis (CvM)	
$n \int_{-\infty}^{\infty} \left(F_n(x) - F(x)\right)^2 dx$	$\frac{1}{12n} + \sum_{i=1}^{n}\left(F_i - \frac{2i-1}{2n}\right)^2$
Anderson-Darling (AD)	
$n \int_{-\infty}^{\infty} \frac{\left(F_n(x) - F(x)\right)^2}{F(x)(1-F(x))} dx$	$-n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\log(F_i(1-F_{n+1-i}))$

Where, *F(x)* is the fitted cumulative distribution function, $F_n(x)$ is the empirical distribution function and *n* is the number of observations of a continuous variable *X*.

4 Results

4.1 Probability distribution for waiting time to be repaired

In this work we have proposed five probability distributions to fit the waiting time and the out of service time for the devices: the *normal, exponential,* *Weibull*, *gamma*, *lognormal* distribution. These probability distributions are widely used to describe events recurring at random points in time, such as the time between failures of electronic equipment or the time between arrivals at a service center. An important characteristic of the exponential distribution is the "memoryless" property, which means that time has no effect on future outcomes.

To estimate and evaluate the fitting we have used R software as a tool and then we have analyzed the outcomes of the results.

To compare the fitting performance of the probability distributions we have used some accurate graphical tests. The density plot and the CDF plot may be considered as the basic classical goodness-of-fit plots and the QQ plot together with PP plot are complementary but may be very informative in some cases.

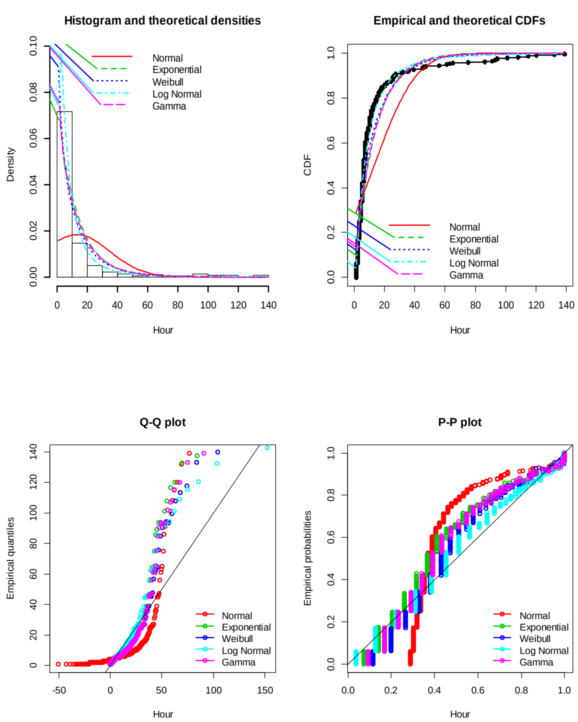


Figure 5: Four goodness-of-fit plots for various distributions fitted to waiting time to repair data

In the waiting time data (in Figure 5), is clearly seen that the normal distribution does not describe at all the empirical data, but the other fitted distribution such as exponential, Weibull, gamma and lognormal describes the left tail of the empirical distribution, especially the lognormal distribution could be preferred for the better description of center of the distribution.

We have used the Kolmogorov-Smirnov statistical test, Cramer-von Mises and Anderson-Darling test to select the "best" fitting distributions for the waiting time and out of service time. Results for the waiting time to repair fitted distributions are shown in Table 5.

Table 5: Goodness of fit statistics for waiting time to repair data

	Kolmogorov-Smirnov	Cramer-von Mises	Anderson-Darling
Normal	0.28	9.27	47.95
Exponential	0.19	2.97	15.37
Weibull	0.13	1.75	10.79
Lognormal	0.09	0.41	2.68
Gamma	0.17	2.41	13.27

	Normal	Exponential	Weibull	Lognormal	Gamma
AIC	2630	2093	2077	1986	2092
BIC	2637	2097	2085	1994	2099

The Anderson-Darling statistic is of special interest when it matters to equally emphasize the tails as well as the main body of a distribution but it should be used with caution when comparing fits of various distributions to the data. By the other side the Cramer-von Mises and Kolmogorov-Smirnov statistics, do not take into account the complexity of the model (i.e., parameter number. To a better decision on fitting model we may consult the information criteria statistics (AIC and BIC).

As it is clearly seen from Table 5 the goodness-of-fit statistics for the waiting time are in favor of the lognormal distribution.

4.2 Probability distribution for out of service time

The empirical histogram for the out of service time is clearly different than the waiting time to be repaired.

We have also taken into consideration the up mention probability distributions for the out of service data and the results of the density plot, CDF plot, QQ plot together with PP plot are shown in Figure 6.

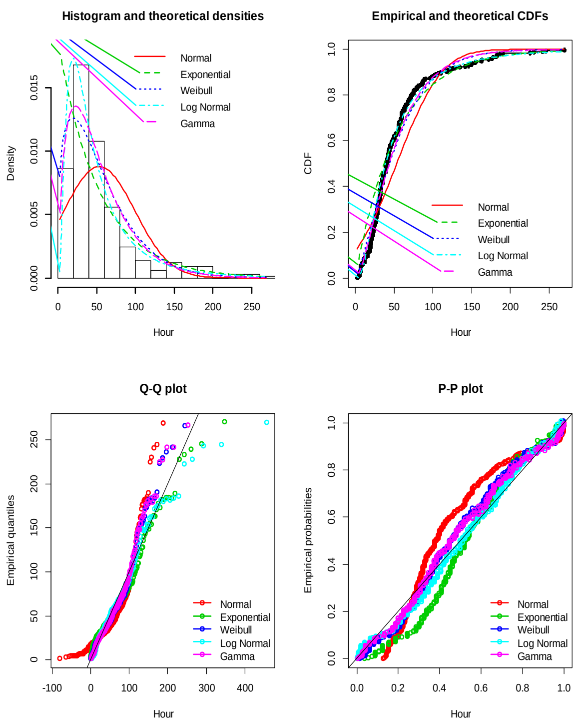


Figure 6: Four goodness-of-fit plots for various distributions fitted to out of service time data

In the out of service time data (in Figure 6), the exponential distribution and the normal distribution do not give an adequate fit, but Weibull, gamma and lognormal distributions describe satisfactory the most of the empirical data. At this point it is difficult to decide on the most preferred distribution so it is a good moment to consult the values of the goodness-of-fit statistics mention above. It seems, form Table 6, that the lognormal distribution and the gamma distributions are comparative with each other to "best" fit the data.

Table 6: Goodness of fit statistics for out of service time

	Kolmogorov-Smirnov	Cramer-von Mises	Anderson-Darling
Normal	0.1672	3.45	19.97
Exponential	0.1677	2.1	12.06
Weibull	0.079	0.7	4.24
Lognormal	0.069	0.28	2
Gamma	0.066	0.46	2.76

	Normal	Exponential	Weibull	Lognormal	Gamma
AIC	3458	3287	3252	3235	3238
BIC	3466	3291	3260	3243	3245

5 Conclusions

From the graphical and numerical tests it was observed that for the waiting time to repair the "best" probability distribution was the lognormal with parameters: *meanlog: 1.91 sdlog: 0.062*. For the out of service time the lognormal distribution with parameters: *meanlog 3.66; sdlog 0.83* and the gamma distribution with parameters: *shape 1.74; rate 0.032* seems to better fit the real data.

From the achieved results we can assert that: approximately 95% of the equipments spend less than 3 hours waiting to receive the service from the lab and 99% of them spend less than 8 hours. Moreover, the time out of service regarding the fitting from the two distributions varies from: 24-28% of the equipment's which spend less than 24 hour and 53-60% of the equipment's spend less than 48 hour; 96-98% of the equipment's obtain the service and return to work within 7 days.

This study was carried out without considering the type of electronic devices arriving at the service center. Of interest would be the categorization of equipment and their study in particular to predict the service time and design a maintenance strategy to optimize cost and customer satisfaction at the same time.

Acknowledgments

The authors want to thank the service center that provided real data for a better evaluation strategy.

References

[Dut08] C. Dutang, V. Goulet, M. Pigeon. "*actuar: An R Package for Actuarial Science*" Journal of Statistical Software, 25(7), 1–37, 2008

[Szt 16] J. Sztrik. *Basic Queueing Theory*. GlobeEdit, OmniScriptum GmbH & Co, KG, Saarbrucken, 2016, ISBN 978-3-639-73471-3.

[Mul15] M. L. Delignette-Muller and C.Dutang. *Fitdistrplus: An R Package for Fitting Distributions*. Journal of Statistical Software, Volume 64, Issue 4, February 2015.

[D'Ag86] R. B. *D'Agostino* and M. A. *Stephens*. *Goodness-of-Fit Techniques*. 1st edition. Dekker, 1986.

[Ven10] W. Venables, B. Ripley BD. *Modern Applied Statistics with S*. 4th edition. Springer Verlag, 2010.

[Wan17] Zh. Wang, M. Zhang, D. Wang, C. Song, M. Liu, J. Li, L. Lou and Zh. Liu. *Failure prediction using machine learning and time series in optical network*. Optics Express 18553, Volume 25, No. 16, August 2017.

[Boos12] D. Boos https://www.rdocumentation.org/packages/Rlab/versions/2.15.1/topics/Gamma

[Hur13] Ch. Hurlin, https://www.univ-orleans.fr/deg/masters/ESA/CH/Chapter2_MLE.pdf