Automation of Life Data Analysis Processes

Jonathan Krämer¹, Gregor Thiele², Theresa Johanni² and Jörg Krüger¹

¹ Department for Industrial automation at the Institute for Machine Tools and Factory Management, Technical University of Berlin
² Department for Process automation and Robotics, Fraunhofer Institute for Production Systems and Design Technology

E-mail: jonathan.kraemer@gmx.de

Abstract. Life data analysis is an effective statistical process to gain information on the reliability of technical facilities. The Weibull analysis is a widely-used method for modeling the probabilities of different functional failures. For its application in modern manufacturing and maintenance processes, an automation of this analysis is expedient. An exemplary process of data collection and fitting of a parametric model with continuous time scale is analyzed in this article. Its automation with different statistic software tools is described and comparatively evaluated. With the ambition of minimizing manual effort in the process, a highly automated solution was developed in this research. We describe this self-implemented solution in “R” and justify that this procedure automatizes the process more than the examined commercial implementations. The here presented work proves to be a practical real-time assistance, right now applied in aircraft maintenance.

1. Introduction

1.1. Life data analysis and Weibull distribution

A technical device fulfills its demanded functionality only for a certain time period, from its launching date to its end-of-life (EOL). For mounted parts the EOL is often caused by a malfunction and comes with the substitution of the component. Therefore, failure data is collected and analyzed to investigate the reasons and estimate the time of failure for identical or similar constructions. Life data analysis is the statistical field which provides the methods to model the failure data and gain insight on the behavior of technical facilities.

To get a more concrete picture of the kind of data this work is tackling, one can imagine this everyday world example: Lets say an overhaul worker works in aviation industry. To be able to better predict the EOL of all the technical devices he is responsible for, it is natural to collect some data and try to gain insights on this basis. So, if for example an engine fails and the worker identifies destroying corrosion on the engine blades, he makes a note to his failure data repository with the engine’s specification and its individual life period (measurable e.g. in hours or hot-cold-cycles) and the corrosion as specific failure cause. All this collected data has to be analyzed by means of statistical methods. If done manually, this analysis is a time-consuming task. Especially if the workperson is not that well trained in these methods, it would be advantageous to have this procedure as automatized as possible. As a result, the workperson is able to reliably predict the EOL of engines due to corrosion and replace them in advanced at a
suitable moment without endangering the function of the product during employment, meaning e.g. an aircraft during its flight.

The above-mentioned statistical methods build the scientific field of life data analysis. Here, the Weibull distribution is often used to model failure data. In 1951, Wallodi Weibull promised that this distribution serves for a wide variety of data [13]. In fact, it proves to be state-of-the-art for life data analysis on all kind of data from different industrial sectors. Especially, it is capable to approximate all typical failure frequencies of a technical facility during its whole utilization time, representing different modes of failure of a technical facility. Moreover, it serves to mimic other commonly used distributions like the exponential distribution. For more information on the Weibull distribution see the references given in Sec. 1.3.

1.2. Outline of this article
After the above given motivation and the short introduction to the theoretical background of this work, Sec. 1.3 summarizes the state-of-the-art of science and technology this research is based on. Sec. 2 describes the examined life data analysis process, its requirements and outputs. The considered software tools for the automation are introduced. Sec. 3 covers the realization of the life data analysis using different software solutions. Sec. 4 gives the conclusion and an outlook.

1.3. Related Work
Our understanding of life data analysis is based on [3], [8] and [10]. For more engineering orientated work, see [1]. The usage of software for life data analysis is described in [12], [4], [11] and [2]. One specific solution to automate a life data analysis, stepping in the same direction as this work, is mentioned in [9]. This solution motivated the alternative of self-implementing the process in the programming language “R” using the package [5] as discussed in Sec. 3. More information on this self-written implementation and its evaluation can be found in [6].

2. Concept
This section presents the industrial process that was automated in the course of this survey. We present features of the data that was provided for analysis, expound the requirements of the specific process and formulate the steps of its automation and its expected outputs. For more detailed information on the calculations that are part of these steps, see the listed literature for the process of life data analysis mentioned in Sec. 1.3.

2.1. Structure of the given process
In this work, we discuss an overhaul process of the aviation industry. Especially in the maintaining of aircraft, the reliability of all system components is of paramount importance. Plenty of technical entities are measured frequently, resulting in a comparatively rich data pool. Still, the hereon based analysis process does not constitute a fully variable but custom-designed compilation of the methods that are summarized under the name of life data analysis. Also, it is expected to create a specific output: a parametric model with continuous time scale that fits the available set of data points.

The life data analysis starts with sorting the data and calculating the median ranks. On this basis, the probability of failure for each data point is calculated. Subsequently, for each cumulative distribution function (CDF) we have to configure a probability paper (coordinate system) in which the respective distribution forms a straight line. Even this simple assistance by predefined probability papers helps automating tasks in overhaul processes.

After defining a suitable (still empty) plot, the data points used for the regression are drawn in the graph of the distribution followed by the regression line respectively. Fitting can be performed using two different methods which should be available in the software tool. The
median rank regression has a graphical reference as it tries to minimize the residuals \( x \) on \( y \) or \( y \) on \( x \). Alternatively, one can estimate the maximum likelihood function to find the most likely parameters for the distribution. After this calculation of the distribution parameters, the distribution can be plotted into the probability plot. For a deeper understanding of the fitted distribution, we add confidence intervals for point estimations as well as confidence bounds across the whole distribution. The calculation method of those intervals depends on the used method for the model estimation. Preferred methods for interval estimation are maximum likelihood bounds, fisher matrix bounds and beta-binomial bounds. In this specific application, extrapolation of the resulting model serves the prediction of future effects. On the basis of this model, different output values can be derived, as discussed in the following Sec. 2.2.

2.2. Structure of the result
The procedure of fitting the model to the data is described above. Here, we list the results and outputs that are needed for the user interface. The data and distribution should be visualized in a probability plot in an interactive way. This allows the user to retrieve the exact data and distribution parameters. Moreover, the region of confidence needs to be visible in the plot. For more detailed information about the model and the failure behavior, special values like the distribution parameters, the goodness-of-fit indicators and the important B-lifes should be accessible as well. For a user with a special interest in the value of a certain failure time or percentage of failure, a corresponding calculator is advantageous. To get a perspective of the model fit a comparison of all implemented distributions is useful.

2.3. Process requirements
Oriented on the considered use-case of the aviation industry with the described process structure we deduce a list of requirements that the automated solution must fulfill. They are divided into the categories technical/statistical and automatizing. Referring to these requirements, we assess the software tools according to their adequacy for implementation in the aimed life data analysis. Concluding from experience, the default distribution to fit the data should be the Weibull distribution. Still, different distributions that are additionally relevant in the practice of life time analysis should be available for modeling as well.

3. Practical realization
3.1. Software
Available software solutions for automating life data analysis are listed in [7]. This extensive table particularly provides the evaluation of the requirements given in Tab. 1 for each candidate. Like that, it helps with the decision which program is most suited to automate a given process. The results of the automation of the process with the most promising solutions for our case are laid out in the following.

3.2. Evaluation and Interpretation
Here, the automation of the given industrial process described in Sec. 2 with the different suitable software solutions of the previous Sec. 3.1 is presented. Four different solutions are promising to fit the requirements of our process. The commercially available tools “Minitab”, “Weibull++” and “Relyence Weibull” and a self-written implementation in “R”. For a detailed description of the implementations of the process with these solutions, see [6] or the related work in Sec. 1.3. The applications and results of the automation of the process are expounded and compared as displayed in Tab. 2. For all four solutions advantages and disadvantages are gathered for these aspects: data input, possibilities for individual configurations and availability of the process results. In the given table each row represents a software solution and the columns are the sections of interest.
Table 1. Requirements

| Category | Criterion |
|----------|-----------|
| Automating | graphical user interface |
| | structured results that are easy to understand |
| | user-friendly but extensive configurability |
| | maximized automation |
| Technical | database communication |
| | probability plot with confidence region |
| | processing of censored data: left, right and interval censored |
| | modelling the data with different distributions: exponential, normal, log-normal, logistic, log-logistic, smallest extreme value and Weibull distribution |
| | methods for estimating distribution parameters: median rank regression and maximum likelihood estimation |
| | methods for estimating confidence intervals: beta-binomial bounds, likelihood-ratio bounds, fisher-matrix bounds |
| | estimation quality of the different distributions based on goodness-of-fit parameters |

Table 2. Interpretation

| data input | manipulation | process output |
|------------|--------------|----------------|
| Minitab | + Database connectivity | + many options manipulatable | + detailed information |
| | − manual entry | − no further option for automation | − no calculator available |
| | + request for censoring of data | + options changeable when looking at results | + used methods and options |
| | − no import option | − not well described | + easy export of results |
| | Weibull++ | − import only from software from Relyence | − plot and values representable only separately |
| | − manual entry | + options changeable when looking at results | − default plot w/o confidence region |
| | Relyence Weibull | + import from .csv-file | + interactable plot |
| | + database connection programmable | + options changeable at all time | + overview over all available distributions |
| | − no direct manual entry | + default options given | + information over used methods and options |
| | R-solution | − import only from software from Relyence | + plot is main output |
| | − manual entry | + options changeable when looking at results | + plot is main output |
| | + import from .csv-file | + default options given | − no export option |

4. Conclusion and Outlook
This last section presents a conclusion of the results of this work. It also lists interesting research questions that came up during this work but were not pursued.

4.1. Contribution
From all discussed tools, the self-implemented solution in “R” reaches the highest degree of automation. This was achieved by a suitable preparation of the results which makes them
understandable even for inexperienced users and by the generic design appropriate for many tasks in overhaul work.

A manual data import or an option to export results is not given, but just because this was not part of our specific requirements. Such interfaces can be added easily. This solution allows to perform the entire procedure of life data analysis by only one click. The automation is especially striking with regards to time-consuming tasks as the analysis of the data differentiated according to specific failure causes as described in our everyday example in Sec. 1. The software is designed user-friendly and easy to operate also for users with little statistical background, i.e. shop floor staff. Because of the focus on overhaul tasks, our solution is not designed for statisticians who may require more flexibility in the analysis. Despite of the extensive functionality, the results still need to be checked and interpreted manually.

4.2. Further Potentials

To further validate our results, it would be interesting to implement the different solutions in real overhaul plants and evaluate the impact of this innovation as well as the feedback on the user experience with the different software tools. For every use-case, the interested engineers need to evaluate whether it is more efficient to utilize commercial toolboxes or to implement own approaches as demonstrates in this article.

References

[1] Dr. R. B. Abernethy. The New Weibull Handbook. Dr. R. B. Abernethy, 1993. ISBN: 0-9653062-3-2.
[2] R. S. Amberti, D. Kenett and S. Zacks. Modern Industrial Statistics. Wiley Academic, 2014. ISBN: 978-1-118-45606-4.
[3] L. A. Escobar and W. Q. Meeker. Statistical Methods for Reliability Data. Wiley-Interscience, 1998. ISBN: 978-0-471-14328-4.
[4] S. M. Gupta and K. K. Pochampally. Reliability Analysis with Minitab. Taylor & Francis Ltd., 2016. ISBN: 978-1-4987-3767-8.
[5] T.-G. Hensel. cran/weibulltools. 2019. URL: https://github.com/cran/weibulltools (visited on 05/24/2020).
[6] J. Krämer. “Automatisierte Lebensdatenanalyse mittels der Weibull-Verteilung zur Instandhaltung von Triebwerken”. Technische Universität Berlin, 2020. DOI: https://dx.doi.org/10.14279/depositonce-10996.
[7] J. Krämer. Software solutions for life data analysis. DOI: https://dx.doi.org/10.14279/depositonce-11002.
[8] J. F. Lawless. Statistical Models and Methods for Lifetime Data. John Wiley & Sons, 2002. ISBN: 978-0471372158.
[9] K. McLain and R. Warren. “Automated Reliability Life Data Analysis of Missiles in Storage and Flight”. In: Annual Proceedings on Reliability and Maintainability Symposium. IEEE, 1990. DOI: 10.1109/ARMS.1990.68006.
[10] W. B. Nelson. Applied Life Data Analysis. John Wiley & Sons, 1982. ISBN: 9780471094586.
[11] K. Stephens. Reliability data analysis with Excel and Minitab. ASQ Quality Press, 2012. ISBN: 978-0-87389-831-7.
[12] A. Tobias and D. C. Trindade. Applied Reliability. Taylor & Francis Inc, 2009. ISBN: 978-1584884668.
[13] W. Weibull. “A Statistical Distribution Function of Wide Applicability”. In: ASME Journal of Applied Mechanics. Transactions of the American Society of Mechanical Engineers 18 (1951), pp. 293–297.