Proceedings of the 2018 Winter Simulation Conference
M. Rabe, A.A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, eds.

THE DIFFERENT TYPE OF DATA IN REALIBILITY AND THE INFLUENCE IN SIMULATION TECHNIQUES

Daniel Gaspar

Mechanical Engineering Department
FEUP-University of Porto
Rua Dr. Roberto Frias
Porto, 4200-465, PORTUGAL

ABSTRACT

The world of simulation involve the generation of several independent simulated data sets. These generated data sets must also be completely independent for the different scenarios considered, such as, in the presence of censored data. Designing high-quality simulations that reflect the complex models seen in practice, such as in failures prognostic studies, sometimes is an hard process. In our PhD work, we intend to contribute in the way of design and programming algorithms that generate correctly, robust and non-skewed censored data. The methodology and procedure can be a useful tool in the field of simulation of reliability of equipment and components. On the other hand, the purpose of this presentation in PhD colloquium, is to expose the first part of PhD, that develop a methodology and a procedure based in algorithms made in an open-source software – **R:** a programming language and free software environment for statistical computing.

1 INTRODUCTION

Simulation studies should be designed with rigor, similar to the study of real data, since they should represent the results of real events. Simulating data sets requires an assumed distribution for the data and full specification of the required parameters.

Today there is a lot of simulation studies in reliability censor data using Weibull and other distributions, unfortunately, in very few published simulation studies, on reliability, sufficient detail is provided to assess the integrity of the study design or to allow readers to fully understand all the processes required when designing their own simulation study, and explain how they have done the simulation and how generate the random number.

Random numbers are the essential basis of the simulation. Usually, all the randomness involved in the model is obtained from a random number generator, that produces a succession of values that are supposed to be realizations of a sequence of independent and identically distributed (i.i.d.) random variables, L'Ecuyer (2012).

These random numbers are then conveniently transformed to simulate the different probability distributions that are required in the model. In general, the validity of the transformation methods depends strongly on the hypothesis that the starting values are realizations of random variables i.i.d. U(0,1).

To decide whether a given sequence is "truely" random, or independent and identically distributed, it is necessary to use non-parametric tests for randomness. Combining the decisions made by each test, improves the confidence on the randomness of a given sequence.

The data is considered complete when it is known the exact time of each system failure. In many cases the data contain uncertainties, i.e., it is not known the exact moment when the failure occurred. The data containing such uncertainty as to when the event occurred are regarded as incomplete or partial. Incomplete data give only part of the information about the failure time of the units under review, truncated Gijbels (2010). However, this information should not be ignored or treated as failure. In the absence of such data, it would not be possible to make good estimation parameters and after that make a proper analysis.

2 METHODOLOGY

For each model of censored random data generation, in the first phase, we developed the censored data generation algorithm, then we apply our methodology, which starts with random number generation, based on uniform distribution, then conversion to the distribution chosen, as well as the parameters defined. To random number generating, it´s use the RNG "Mersenne-Twister", from Matsumoto and Nishimura (1998). In the next step, is chose a *n* dimension of sample (1000) to do for each simulation, then apply the five randomness tests and for each one we extracted the p-value (for the chosen confidence level - 0.05 and 0.1), after that, is made this simulation cycle for *m* times (1000) and calculate the average of the p-values of each distribution. This routine are made for different shape parameters (five) of the five statistical distribution and for six sample sizes. For each of them, we use two level of confidence and different percentage of censorship (5, 10, 20 and 30%); we use the following scheme to do a classification in a range, from 1 to 3 in which we have the value 3 - which means good and accepted, the value 2 - reasonable and accepted and finally the value 1 - bad and not acceptable. The percentages defined were: for 3 of (100-90%); for 2 (90-75%) and for 1 (75-0%).

For this study and analysis, we use the distributions that are most used in reliability and mantainability: - Weibull, normal, gamma, log-normal and exponential. In each of them we had to calculate the censorship time for which we would have a reliability of 5, 10, 20 or 30%. The methodology and a procedure based in algorithms made in a language, open-source - **R** a programming language and free software environment for statistical computing and graphics that is supported by the R Foundation for Statistical Computing, R Development Core Team (2017).

3 RESULTS AND CONCLUSIONS

The presentation permit to emphasize, the importance to test the simulation algorithms and show the influence of parameters of distribution or the parameters of model simulations (like that number of samples or simulations in one cycle) in the randomness of data generation.

The hypothesis test to apply to the specific generator censor data, have to be select very carefully in order to have good results and to optimize the simulation time.

Ideally, when testing a model in reliability for simulation, there are some difficulties to choose the good RNG in combine with the algorithm to generate the censor data, because the variable of practical interest can have a bad interference between the RNG and the rest of algorithm of simulation model; the results of the simulation model can be biased or not “truely” random.

The methodology that we use for simulations of the reliability, can be applied in situations of data in real time, like equipment autonomous and with artificially intelligent.

REFERENCES

Gijbels, I. 2010. “Censored Data”. *Wires Computational Statistics* 2(2):178 – 188.

L’Ecuyer, P. 2012. “Random Number Generation.” In *Handbook of Computational Statistics: Concepts and Methods*, 2rd ed., edited by J. E. Gentle, W. Hardle and Y. Mori, 35-71. Berlin, Germany: Springer-Verlag.

Matsumoto, M. and T. Nishimura. 1998. “Mersenne Twister: A 623-dimensionally Uniform Pseudo-random Number Generator”. *ACM Transactions on Modeling and Computer Simulation* 8:3-30.

R Development Core Team. 2017. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. https://www.r-project.org/, accessed October 2018.