ESReDA Conference 2011
41st ESREDA Seminar
5-6 october 2011, La Rochelle, France

Session No.2, <Paper No.>

Change-point estimation in a statistical reliability model

Jean-François DUPUY
INSA de Rennes/Institut de Recherche Mathématique de Rennes, France
Jean-Francois.Dupuy@insa-rennes.fr

Abstract

The Weibull distribution is one of the most widely used lifetime distributions in reliability engineering. This distribution is characterized by a shape parameter and a scale parameter, and its flexibility allows to model various patterns of the hazard of failure. It is usually assumed that the shape and scale parameters are constant over the duration of the test or experiment. In many settings however, such as in software reliability engineering, unexpected events (e.g., presence of a frailty in the operating system caused by an assembly quality problem, modification of the operating constraints applied to the system,...) may result in an abrupt change in the hazard rate. The off-line investigation of such a change is crucial for designing subsequent maintenance policies. Such investigation includes the estimation of the unknown change-point time and of the unknown hazard rates before and after the change-point. This communication gives an overview of the recent literature devoted to this problem. In particular, we focus on the maximum likelihood principle, which can be used to estimate both the change-point and the parameters of the reliability model before and after the change-point. In the case of the Weibull model for right-censored failure times, we evaluate the numerical properties of the maximum likelihood estimations via simulations.

Keywords – Weibull distribution, change-point, maximum likelihood, simulations

Introduction

Motivated by engineering applications, Weibull [1] has proposed a distribution that has become one of the most popular statistical models in reliability. Its instantaneous hazard function h(t) (with $t \geq 0$) has the form

$$h(t) = a\,b\,t^{a-1}$$

where a and b are unknown positive parameters, respectively called the shape and scale parameters. The corresponding reliability function is

$$S(t) = \exp(-b\,t^{a})$$

This distribution is flexible enough to accommodate increasing (when $a \geq 1$), decreasing (when $a < 1$), or constant ($a = 1$) hazard rate functions. Bagdonavicius and Nikulin [2] and Meeker and Escobar [3] provide detailed expositions of the various shapes of the hazard, survival, and density functions of the Weibull law. A comprehensive perspective on Weibull models is given by Murthy *et al.* [4].
The Weibull distribution has been widely used so far in reliability to describe the failure of various materials. A non-exhaustive list includes applications to the life of

steel [1], to pitting corrosion in pipes [5], to failure of carbon fibre composites [6], to automobile warranty data [7]. Many other examples and references can be found in [8]. Several generalizations and extensions of the traditional two-parameters Weibull distribution have also been proposed. We refer the interested reader to [4], [9], and to the numerous references therein for detailed treatments and expositions.
In the above references, it is usually assumed that the shape and scale parameters are constant over the duration of the test or experiment. In many settings however, such as in software reliability engineering, unexpected events (e.g., presence of a frailty in the operating system caused by an assembly quality problem, modification of the operating constraints applied to the system,...) may result in an abrupt change in the hazard rate (see, among others, [10], [11], [12], [13], [14], and the references therein). The off-line investigation of such a change is crucial for designing subsequent maintenance policies. Such investigation includes the estimation of the unknown change-point time and of the unknown hazard rates before and after the change.
However, most of the available literature on change-point in parametric failure time models has focused on the special case of the exponential distribution. In particular, the following references: [10], [11], [15], and [16] have considered the problem of estimation in the change-point model:

$$h(t)=b+c\cdot 1(t>\tau)$$

where 1(A) is the indicator function which is equal to 1 if the event A is met (and to 0 otherwise), and τ is an unknown positive change-point time (some other references can be found in [17]).
In this communication, we investigate the problem of estimating the Weibull distribution, when a change occurs in the shape and scale parameters at an unknown change-point time τ . An off-line maximum likelihood-based procedure is proposed and investigated numerically.
The rest of the communication is organized as follows. In Section 1, we describe the change-point model under consideration and we develop an estimation procedure for its parameters. Section 2 reports the results of a simulation study investigating the numerical properties of the proposed estimators. A discussion concludes the communication.

1. A Weibull model with change-point at an unknown time

We consider the following Weibull model, where a change occurs both in the shape and scale parameters, at an unknown change-point time τ :

$$h(t)=a_1 b_1 t^{a_1-1} 1(t\leq\tau)+a_2 b_2 t^{a_2-1} 1(t>\tau) \qquad (1)$$

Note that before τ , the hazard function h(t) is characterized by the shape and scale parameters a_1 and b_1 , and that after τ , the hazard function h(t) is characterized by the shape and scale parameters a_2 and b_2 , with $a_1 \neq a_2$ and $b_1 \neq b_2$.
We assume that we observe n independent times T_i (i=1,...,n), where $T_i=\min(X_i, C_i)$, X_i denotes the failure time of interest, and C_i is a right-censoring time assumed to be independent of X_i and non-informative. We denote by $D_i=1(T_i=X_i)$ the indicator function which is 1 if the failure time has been observed, and 0 if censoring has occurred.

Using the following relationships between the hazard rate h, the reliability function S, and the density function f:

$$S(t)=\exp(-\int_0^t h(u)\,du) \text{ and } f(t)=h(t)\cdot S(t)$$

we can calculate the likelihood of ($a_{1,}a_{2,}b_{1,}b_{2,}\tau$) given the observed sample of failure times (T_i, D_i), i=1,...,n. This likelihood has the form

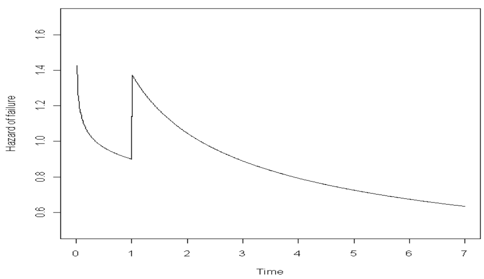
$$L_n(a_{1,}b_{1,}a_{2,}b_{2,}\tau)=\prod_{i=1}^n \left\{a_1 b_1 T_i^{a_1-1}\right\}^{1(T_i\le\tau)D_i}\left\{a_2 b_2 T_i^{a_2-1}\right\}^{1(T_i>\tau)D_i}$$
$$\cdot\exp(-b_1 T_i^{a_1} 1(T_i\le\tau)-b_1\tau^{a_1}1(T_i>\tau))$$
$$\cdot\exp(-b_2[T_i^{a_2}-\tau^{a_2}]1(T_i>\tau)).$$

Estimations of $(a_{1,}a_{2,}b_{1,}b_{2,}\tau)$ can be obtained using the following procedure: for any fixed value of τ , let $(\hat{a}_1(\tau),\hat{b}_1(\tau),\hat{a}_2(\tau),\hat{b}_2(\tau))$ be the value of $(a_{1,}a_{2,}b_{1,}b_2)$ which maximizes $L_n(a_{1,}b_{1,}a_{2,}b_{2,}\tau)$. Then, the unknown change-point time τ is estimated by the value $\hat{\tau}$ which maximizes $L_n(\hat{a}_1(\tau),\hat{b}_1(\tau),\hat{a}_2(\tau),\hat{b}_2(\tau))$ over a fine grid of the expected range of τ **.** At the final stage of the procedure, $(a_{1,}a_{2,}b_{1,}b_2)$ are estimated by $(\hat{a}_1(\hat{\tau}),\hat{b}_1(\hat{\tau}),\hat{a}_2(\hat{\tau}),\hat{b}_2(\hat{\tau}))$ **.**This kind of estimate was investigated in details (both theoretically and numerically) in the case of the exponential model with a change-point (see, for example, [10], [11], [15], [16]). In the sequel, we investigate the properties of this estimate in the case of the Weibull distribution.

2. A simulation study

We conduct a simulation study to investigate the numerical properties of the proposed estimates of the shape and scale parameters of the Weibull model before the change $(a_{1,}b_1)$, after the change $(a_{2,}b_2)$, and of the change-point τ . Precisely, we simulate 1000 samples of n right-censored failure times from the model (1). Using the procedure described in the previous section, we obtain, for each simulated sample, an estimate of $(a_{1,}a_{2,}b_{1,}b_2)$ and τ . From these 1000 estimates, we obtain the average absolute bias, standard deviation and root mean square error (RMSE) of the estimates. The following values: $(a_1=0.9\,,a_2=0.6\,,b_1=1,b_2=2.3\,,\tau=1)$ are used in the simulations, resulting in the hazard function pictured on Figure 1. For these values, the average percentage of event times lying after the change-point is about 37%.

Figure 1. Simulated hazard function for the Weibull model with change-point.



The censoring times are generated from the exponential distribution, with parameter chosen to yield a given censoring percentage. Moderate (30%) and heavy censoring (50%) are successively considered. The sample size n is successively chosen equal to 200, 500, and 1000. The free statistical software R was used for the simulations. The results are given in Table 1 to 6 below, for the various combinations of the design parameters.

Table 1. Simulation results for n=200 and censoring percentage = 30%.

	Absolute bias	Standard deviation	RMSE
a_1	0.0166	0.0909	0.0920
b_1	0.0353	0.1184	0.1235
a_2	1.0231	0.4065	0.7362
b_2	0.1133	0.6415	0.6922
τ	0.0472	0.0612	0.0773

Table 2. Simulation results for n=200 and censoring percentage = 50%.

	Absolute bias	Standard deviation	RMSE
a_1	0.0299	0.0934	0.0971
b_1	0.0412	0.1366	0.1426
a_2	1.4326	0.5886	1.0416
b_2	0.1949	0.6084	0.7555
τ	0.0872	0.0978	0.1310

Table 3. Simulation results for n=500 and censoring percentage = 30%.

	Absolute bias	Standard deviation	RMSE
a_1	0.0177	0.0662	0.0681
b_1	0.0225	0.0850	0.0878
a_2	0.8949	0.3366	0.6336
b_2	0.1025	0.6383	0.6801
τ	0.0309	0.0439	0.0537

Table 4. Simulation results for n=500 and censoring percentage = 50%.

	Absolute bias	Standard deviation	RMSE
a_1	0.0246	0.0626	0.0664
b_1	0.0245	0.0897	0.0929
a_2	1.3014	0.4056	0.8798

b_2	0.1831	0.6339	0.7607
τ	0.0714	0.0829	0.1094

Table 5. Simulation results for n=1000 and censoring percentage = 30%.

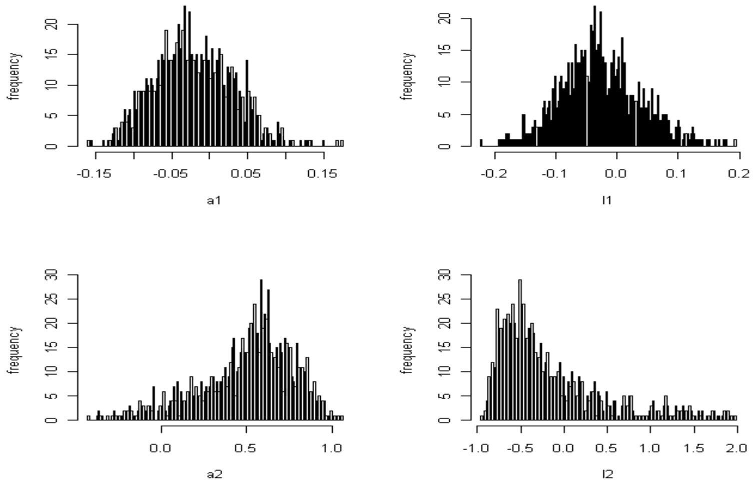
	Absolute bias	Standard deviation	RMSE
a_1	0.0189	0.0538	0.0564
b_1	0.0241	0.0731	0.0769
a_2	0.8490	0.3057	0.5940
b_2	0.0690	0.6556	0.6742
τ	0.0216	0.0294	0.0364

Table 6. Simulation results for n=1000 and censoring percentage = 50%.

	Absolute bias	Standard deviation	RMSE
a_1	0.0241	0.0504	0.0548
b_1	0.0210	0.0688	0.0720
a_2	1.2850	0.3348	0.8405
b_2	0.1883	0.5898	0.7315
τ	0.0514	0.0631	0.0814

From these results, it appears that the proposed estimation method performs quite well, even when the sample size is relatively small. We note that the estimate of the change-point time τ is quite sensitive to the censoring percentage. Precisely, for a given sample size, the accuracy of the estimate (as measured by the standard deviation and RMSE) substantially decrease when the censoring fraction increases from 30% to 50%. The estimate of the shape parameter after the change-point (namely a_2) seems also to be quite sensitive to the censoring proportion, with a bias that can be substantially large when the censoring percentage is about 50%.
The point estimation of the shape and scale parameters in the change-point Weibull model provides some useful preliminary informations about the underlying model. A further step in the analysis is to obtain confidence intervals and to derive tests of hypothesis. This requires distributional results about the estimates of the various parameters (or at least large-sample distributional results).
We can gain insight into the large-sample distributions of the proposed estimates by representing the histograms of the estimated values when n is large. Taking n=2000, and again 1000 simulated samples, we obtain the following histograms for $(\hat{a}_1(\hat{\tau}), \hat{b}_1(\hat{\tau}), \hat{a}_2(\hat{\tau}), \hat{b}_2(\hat{\tau}))$:

Figure 2. Histograms of the $(\hat{a}_1(\hat{\tau}), \hat{b}_1(\hat{\tau}), \hat{a}_2(\hat{\tau}), \hat{b}_2(\hat{\tau}))$ for n=2000.



From these histograms, it appears that the large-sample distributions of $\hat{a}_1(\hat{\tau})$ and $\hat{b}_1(\hat{\tau})$ look like gaussian distributions. The distributions of $\hat{a}_2(\hat{\tau})$ and $\hat{b}_2(\hat{\tau})$ seem to be skewed, however.

Summary & Conclusion

In this communication, we have considered a change-point Weibull model for failure time data. We have proposed an estimation method for the parameters in this model (scale and shape parameters before and after change, and the unknown change-point time), based on the maximum likelihood principle. Our (yet limited) simulation study indicates that provided that the sample-size is large enough, the resulting estimates of the shape and scale parameters and of the change-point time provide reasonable approximations of the unknown values. Our simulation study also indicates that the change-point estimate is quite sensitive to the censoring percentage. Finally, from our simulations, it appears that the large-sample distributions of the estimates of the shape and scale parameters before change can be reasonably estimated by a gaussian law. This may allow the construction of confidence intervals and tests of hypothesis, and constitutes a topic for future research.
Although it is more flexible than the usual Weibull model, the change-point Weibull model assumes that the Weibull distribution is appropriate both before and after the change. This hypothesis may itself not be satisfied in real world situations. Therefore, we may consider generalizing our approach to the case where the hazard of failure before and after the change is modeled using more flexible functions (such as spline functions, see, for example, Krivtsov *et al.* [18]). This is also the topic for future research.

References

[1] Weibull, W. A statistical distribution of wide applicability. Journal of Applied Mechanics, 18, 293-297, **1951.**
[2] Bagdonavicius, V., Nikulin, M. Accelerated Life Models. Modeling and Statistical Analysis**.** Chapman & Hall, **2002**.

[3] Meeker, W.Q., Escobar, L.A. Statistical Methods for Reliability Data**.** Wiley & Sons, **1998**.
[4] Murthy, D.N.P., Xie, M., Jiang, R. Weibull Models**.** Wiley & Sons, **2004**.
[5] Sheikh, A., Boah, J.K., Hansen, D.A. Statistical modelling of pitting corrosion and pipeline reliability. Corrosion, 46, 190-196, **1990**.
[6] Durham, S.D., Padgett, W.J. Cumulative damage model for system failure with application to carbon fibers and composites. Technometrics, 39, 34-44, **1997**.
[7] Attardi, L., Guida, M., Pulcini, G. A mixed-Weibull regression model for the analysis of automotive warranty data. Reliability Engineering & System Safety, 87, 265–273, **2005**.
[8] Bebbington, M., Lai, C.-D., Zitikis, R. A flexible Weibull extension. Reliability Engineering & System Safety, 92, 719-726, **2007**.
[9] Pham, H. (editor). Springer Handbook of Engineering Statistics. **2006**.
[10] Chang, I.-S., Chen, C.-H., Hsiung, C.A. Estimation in change-point hazard rate models with random censorship. In: Carlstein, E., Muller, H.-G., Siegmund, D. (Eds.), Change-point Problems, Institute of Mathematical Statistics Lecture Notes - Monograph Series, vol. 23, 78–92, **1994**.
[11] Muller, H.-G., Wang, J.-L. Change-point models for hazard functions. In: Carlstein, E., Muller, H.-G., Siegmund, D. (Eds.), Change-point Problems, Institute of Mathematical Statistics Lecture Notes - Monograph Series, vol. 23, 224–241, **1994**.
[12] Wu, C.Q., Zhao, L.C., Wu, Y.H. Estimation in change-point hazard function models. Statistics and Probability Letters, 63, 41–48, **2003**.
[13] Dupuy, J.-F. Estimation in a change-point hazard regression model. Statistics and Probability Letters, 76, 182-190, **2006**.
[14] Karasoy, D., Kadilar, C. A new Bayes estimate of the change point in the hazard function. Computational Statistics and Data Analysis, 51, 2993-3001, **2007**.
[15] Matthews, D.E., Farewell, V.T., Pyke, R. Asymptotic score-statistic processes and tests for constant hazard against a change-point alternative. Annals of Statistics, 13, 583-591, **1985**.
[16] Abdel-Aty, Y., Ferger, D. Estimation of the Jump-Point in a Hazard Function. Economic Quality Control, 18, 251-261, **2003**.
[17] Dupuy, J.-F. A survival model with change-point in both hazard and regression parameters. In: Nikulin, M., Commenges, D., Huber, C. (Eds.), Probability, Statistics and Modeling in Public Health, Springer, **2006**.
[18] Krivtsov, V.V., Kolmanovsky, I.V., Davis, T.P. A constrained quadratic spline as a model for the cumulative hazard function. International Journal of Reliability and Safety, 2, 170-178, **2008**.