MODELY SMĚSÍ RIZIKOVÝCH FUNKCÍ S APLIKACÍ V ANALÝZE SPOLEHLIVOSTI

MODELS OF MIXTURES OF HAZARD RATES WITH APPLICATION TO RELIABILITY ANALYSIS

Petr Volf

Adresa: ÚTIA AV ČR, Pod vodárenskou věží 4, 182 00 Praha 8

E-mail: volf@utia.cas.cz

Abstrakt: V příspěvku jsou studovány modely pro analýzu životnosti sestavené jako kombinace několika jednoduchých rizikových funkcí. Užitečnost takových modelů je ukázána na příkladech s umělými i reálnými daty.

Klíčová slova: riziková funkce, Weibullovo rozdělení, analýza spolehlivosti.

Abstract: The objective of this contribution is to present and explore some methods of construction and estimation of models based on mixtures of hazard rates. The goal is to demonstrate their usefulness and applicability with the aid of both artificial and real examples.

Keywords: hazard rate, Weibull distribution, reliability analysis.

1. Introduction, models based on mixtures

Distributions of probability constructed as mixtures of two or more simple distributions are used frequently in many fields of application, in cases when no simple model describes the data sufficiently. Such mixtures are based on the convex combination of probability densities or distribution functions, in order to ensure that the final function is also a density (or distribution function). The most popular model is composed from a set of normal densities. The simplest example is the following:

$$f(x) = p_1 \cdot f_1(x) + p_2 \cdot f_2(x). \tag{1}$$

Interpretation (used also in cluster analysis) could be such that with probability p_1 an item belongs to first group having distribution with density f_1, with probability p_2 to the second group. It is also the way how data representing mixture (1) could be generated. Further, probability densities (and normal densities in particular) are used also as components for the construction of regression models, in the same way as regression splines. For instance a combination (now it need not be convex) of gaussians (in this context called also "radial basis functions") is used to create a curve or surface.

In the statistical reliability analysis, when modeling the time to failure, one of often used characteristics is the hazard rate (HR) $h(t)$ or its integrated version, the cumulative hazard rate (CHR) $H(t) = \int_0^t h(z)\,\mathrm{d}z$. Let us denote $F(t)$ the distribution function and $S(t) = 1 - F(t)$ the survival or reliability function, then $H(t) = -\ln(S(t))$. The use of hazard rate as one of basic characteristics suggests that the mixture models in reliability could be based as well on mixtures of hazard rates instead on mixtures of distributions.

The first step, namely the linear failure rate model $h(t) = a + bt$ was proposed already in Kodlin [5], polynomial failure rate models have then been studied in Bain [1], advanced computation methods also in Pandey et al. [7]. Nonlinear (and no polynomial) models $h(t) = a + bt^\gamma$, $a, b, \gamma > 0$ and methods of estimation have, naturally, also been examined, for instance in Salem [9]. Some systematic recent investigation is also due Tien Thanh Thach from the Technical University in Ostrava, see for instance Briš and Thach [3]. Essentially, there is no problem to extend the methods to deal with more than two components. A question arises, however, whether and when it is reasonable, improving the fit of model to data significantly. In fact, both examples used in the present paper are of such a kind, in both it is shown that two components of additive hazard rate do not suffice.

The next section introduces hazard mixture models in more details. Then, a comparison of two examples illustrates the difference among mixtures of distributions and of hazard rates. Section 4 solves an artificial example leading to the bath-tube shaped hazard rate, and, finally, Sections 5 and 6 solve a real data case, except a plain sum of hazard rate considering also an incremental model, i.e. a variant of model with change points.

2. Mixtures of hazard rates

In practical survival data analysis, rather frequently the hazard rate during the lifetime (not only of a technical device, but also of a biological object) may have a "bath-tube" shape, decreasing in the first short lifetime period, then being approximately constant for the most of lifetime, finally increasing as the object is ageing. This could be a reason for considering a mixture of three simple hazard rates (see an example in Figure 2 below). Notice that the mixture of hazard rates need not be convex, as hazard rate is not standardized (in the sense as the density function is, i.e. that the integral from density function equals one). Notice also a special property of Weibull distribution: When its hazard rate is multiplied by a constant, we obtain the hazard rate of another Weibull distribution. Recall that the Weibull cumulative hazard rate can have two forms,

$$H(x) = a\,x^{\beta} \quad \text{or} \quad H(x) = \left(\frac{x}{\alpha}\right)^{\beta}.$$

Hence

$$c\,H(x) = c\,a\,x^{\beta} = a^*\,x^{\beta} \quad \text{or} \quad H(x) = \left(\frac{x}{\alpha^*}\right)^{\beta}, \quad \alpha^* = \alpha/c^{1/\beta}. \tag{2}$$

I shall use mostly the second notation, as it corresponds to the notation in Matlab (and in Excel, too). From above it is seen that the mixture of two Weibull hazard rates is equivalent to a simple sum $h = h_1 + h_2$ of other Weibull hazard rates, parameters of would-be combination are not identifiable.

However, other distribution types have not such property, hence it has a sense to consider a model – combination (possibly convex) of hazard rates. What is an interpretation of such a model? The sum of hazard rates $h = \sum h_j$ corresponds to the hazard rate of a serial system composed from independent items with h_j, the survival time of the system then to minimum of survival times of components. Multiplication $h^* = c \cdot h$ then corresponds to a proportional change of hazard rate (it is actually the first step to the "proportional hazard regression model" or to frailty models), then $S^* = S^c$.

Immediately a question arises when one should prefer the model based on density function or on hazard rate. Naturally, each approach has its advantages and also specific tools, model fitting methods and procedures. In general, in both cases the model estimation often uses iterative methods of optimization of a criterion based for instance on the maximal likelihood, on Bayes estimation (nowadays often connected with the Markov chain Monte Carlo method, MCMC), or on some other distance. For instance, one of traditional methods how to fit the Weibull distribution is based on the least squares and linear regression. Namely, let theoretical Weibull cumulative hazard rate be $H(t) = a \cdot t^{\beta}$, data T_i, $i = 1, \ldots, N$, estimated CHR $\widehat{H}(t)$. Then, after further logarithmization, we obtain the relation

$$\ln \widehat{H}(T_i) \sim \ln(a) + \beta \cdot \ln(T_i).$$

Or, when dealing with additive hazard rate composed from two Weibull hazard rates, we can use the following:

$$\widehat{H}(T_i) \sim a_1\,T_i^{\beta_1} + a_2\,T_i^{\beta_2},$$

which is linear at least w.r. to a_j-s. The solution is found with the aid of the least squares method, eventually weighted by the asymptotic variance of $\widehat{H}$. Other criteria can use distances of Kolmogorov-Smirnov and variants (Cramér-Von Mises, Anderson-Darling) measuring departures of parametric models from nonparametric estimates of $S(t)$ or $H(t)$.

3. Two examples of mixture models

The first example on Figure 1 shows a case of mixture of two probability densities. Such a model is applicable for instance in the analysis of unemployment duration, as, roughly, there are two types of people: one with certain qualification and willingness to find a new employment, so called "movers", here represented by shorter distribution with density f_1. The other type, called "stayers", however, have problems with finding a new job, their staying time in unemployment is represented by density f_2. Hence, considering these two groups together and their proportions, the density of common distribution of unemployment time is given by a mixture $f = p \cdot f_1 + (1-p) \cdot f_2$. In Figure 1 f_1 corresponds to Weibull distribution with $\alpha_1 = 50$, $\beta_1 = 1.5$, f_2 to Weibull with $\alpha_2 = 150$, $\beta_2 = 3$, $p = 0.5$.

The second example displayed in Figure 2 shows an instance of the hazard rate having "bath-tube" shape. It was constructed by mixing a decreasing HR of Weibull distribution with parameters $\alpha_1 = 1$, $\beta_1 = 0.5$, constant HR (i.e. of

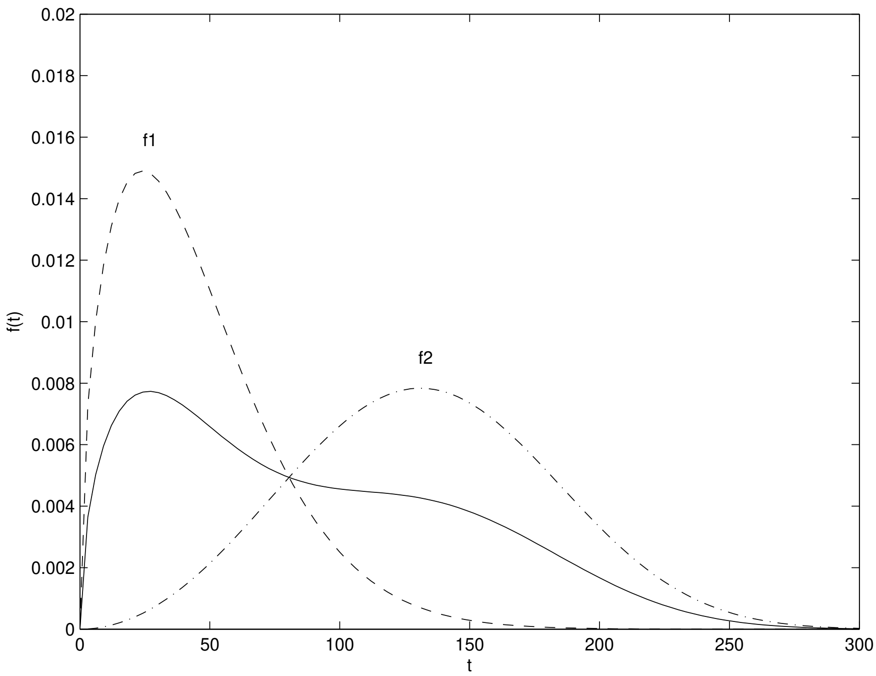

Figure 1: Example of mixture of 2 densities representing "movers", f_1, and "stayers", f_2, resulting $f = (f_1 + f_2)/2$ (thick curve).

exponential distribution, $\alpha_2 = 30$, $\beta_2 = 1$) and an increasing HR of Weibull distribution with $\alpha_3 = 150$, $\beta_3 = 5$. The final hazard rate was obtained as $h = p_1\, h_1 + p_2\, h_2 + p_3\, h_3$ with $p_1 = 0.2$, $p2 = 0.5$, $p_3 = 0.3$. Results are displayed in Figure 2. As it has already been said, the model is equivalent to a simple additive one $h = h_1^* + h_2^* + h_3^*$, namely with $\alpha_1^* = 25$, $\alpha_2^* = 60$, $\alpha_3^* = 190.8$ and β-s the same as above.

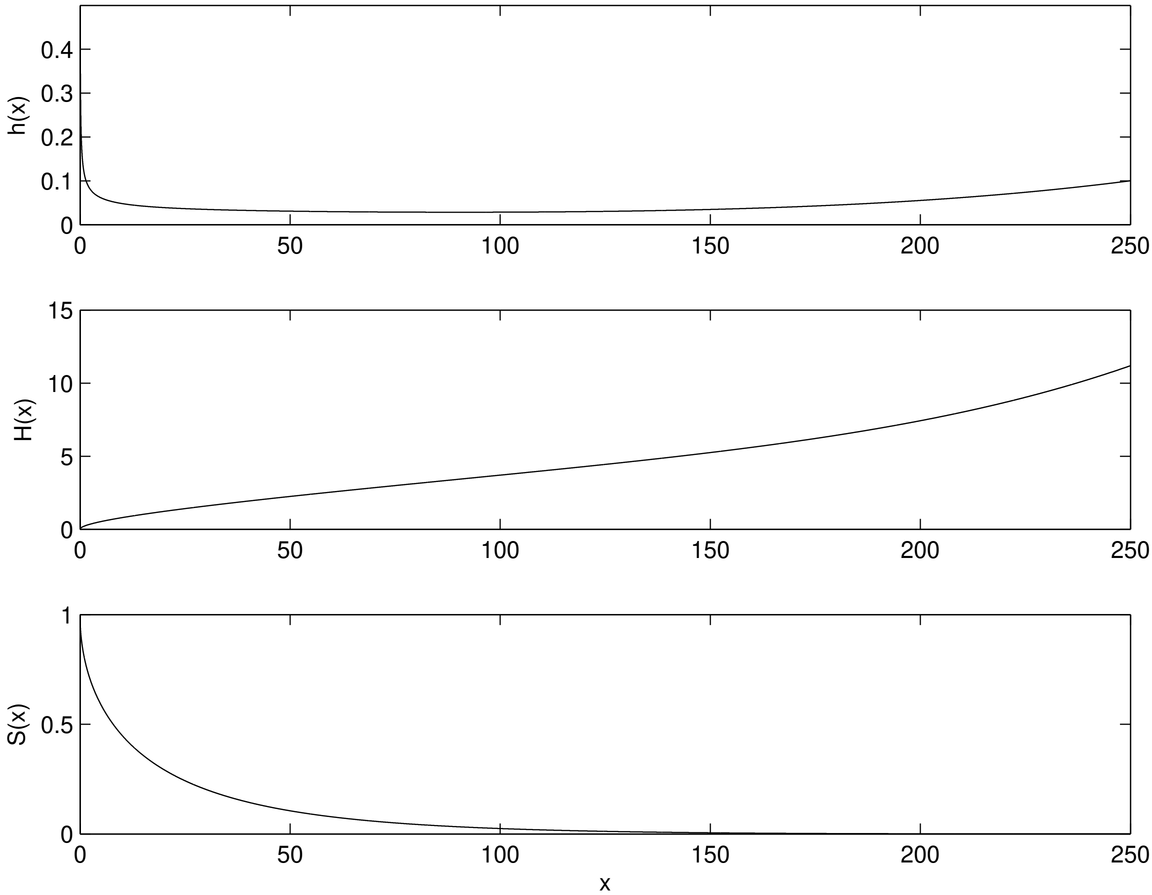

Figure 2: Example of "bath-tubed" hazard rate shape, a mixture of 3 hazard rates.

4. Artificial data example

The data X_i, $i = 1, \ldots, N$, $N = 500$, were generated from the model displayed on Figure 2. Hence, it was expected that its HR is given by the sum of two Weibull and one exponential components. Figure 3 shows nonparametric estimates of the CHR and survival function. The task is to estimate parametric model. From several methods listed above the maximum likelihood estimate (MLE) was chosen. It is possible, for given parameters, to

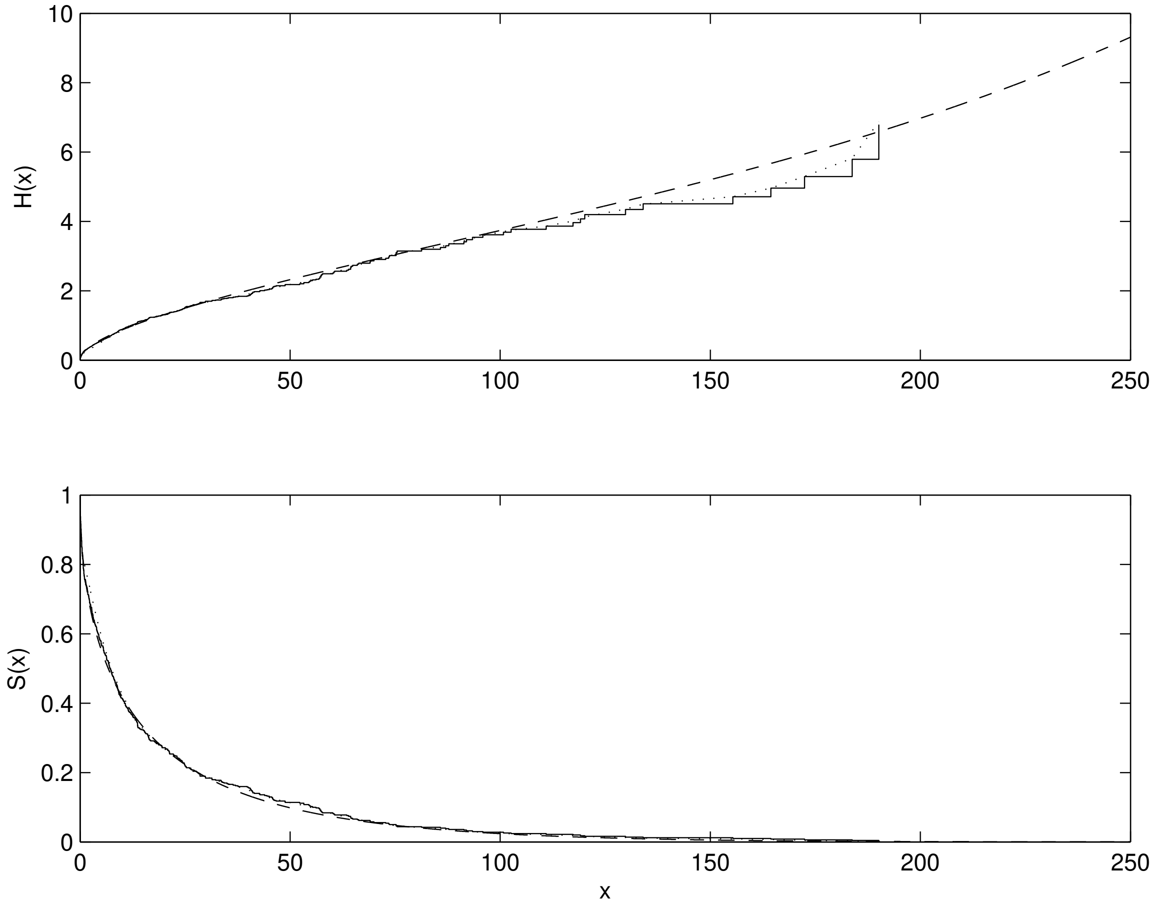

Figure 3: Stepwise: empirical CHR and survival function, dotted are their kernel-smoothed versions, dashed curves are based on the MLE.

evaluate both the log-likelihood as well as its derivatives. On the other hand, parameters maximizing the log-likelihood are found just by a random search (which is here computationally easier than for instance an iteration using the derivatives repeatedly, like the Newton-Raphson algorithm). Therefore the result is just approximate, as well the estimates of borders of 90% confidence intervals. Dashed curves in Figure 3 correspond to the model using the MLE of parameters, their values are listed in Table 1.

5. Real data example

The data used in this part are "Aircraft windshield failure data" taken from Ruhi [8], they were analyzed also in several other papers, e.g. in Blischke et al. [2]. They concern to the "damage or delamination of the nonstructural outer ply or failure of its heating system". These failures do not cause a severe damage to the aircraft but lead to replacement of the windshield. The

	α_1	α_2	α_3	β_1	β_3
MLE	18.0308	78.1834	197.4708	0.5047	3.6122
LCI	10.8970	42.4971	148.4805	0.4549	2.2910
UCI	25.1646	113.8697	246.4611	0.5545	4.9334

Table 1: ML estimates, LCI and UCI are the lower and upper bounds of asymptotic 90% confidence intervals.

	α_1	α_2	β_1	β_2
MLE	39.3812	3.5834	0.9571	2.9336
LCI	38.0081	2.9995	0.5560	1.9097
PCI	40.7543	4.1673	1.3582	3.9575

Table 2: ML estimates, LCI and UCI are again the lower and upper bounds of asymptotic 90% confidence intervals.

data consist of 153 observations, from them 88 are times to failures, the remaining 65 are censored times when no failure occurred during the time of observation. It means that we deal with the random right-censoring scheme. Hence the PLE (Product Limit Estimator) of Kaplan and Meier will be used as nonparametric estimator of survival function, while cumulative hazard rate will be estimated by the Nelson-Aalen Estimator (NAE), see e.g. Kalbfleisch and Prentice [4]. The unit of measurement was 1000 hours.

Both nonparametric estimates are displayed in Figure 4 (stepwise functions), together with estimates obtained from them by kernel smoothing. Ruhi [8] as well as other authors (references see in Ruhi) tried to describe the data by models mixing two plain distribution densities. Success of modelling was assessed by several criteria, as the likelihood or the Kolmogorov-Smirnov distance of model survival function from the nonparametric PLE. Naturally, these two criteria lead to different results, though both are asymptotically consistent. For comparison, I decided to use a model based on mixture of hazard rates with its parameters estimated by the maximum likelihood method. Figure 5 shows the result (dashed curve), namely the model constructed just by sum of two Weibull hazard rates. Such a model has 4 parameters. Similarly as above, the best solution was approached by a random search, then asymptotic confidence intervals were computed from the second derivative

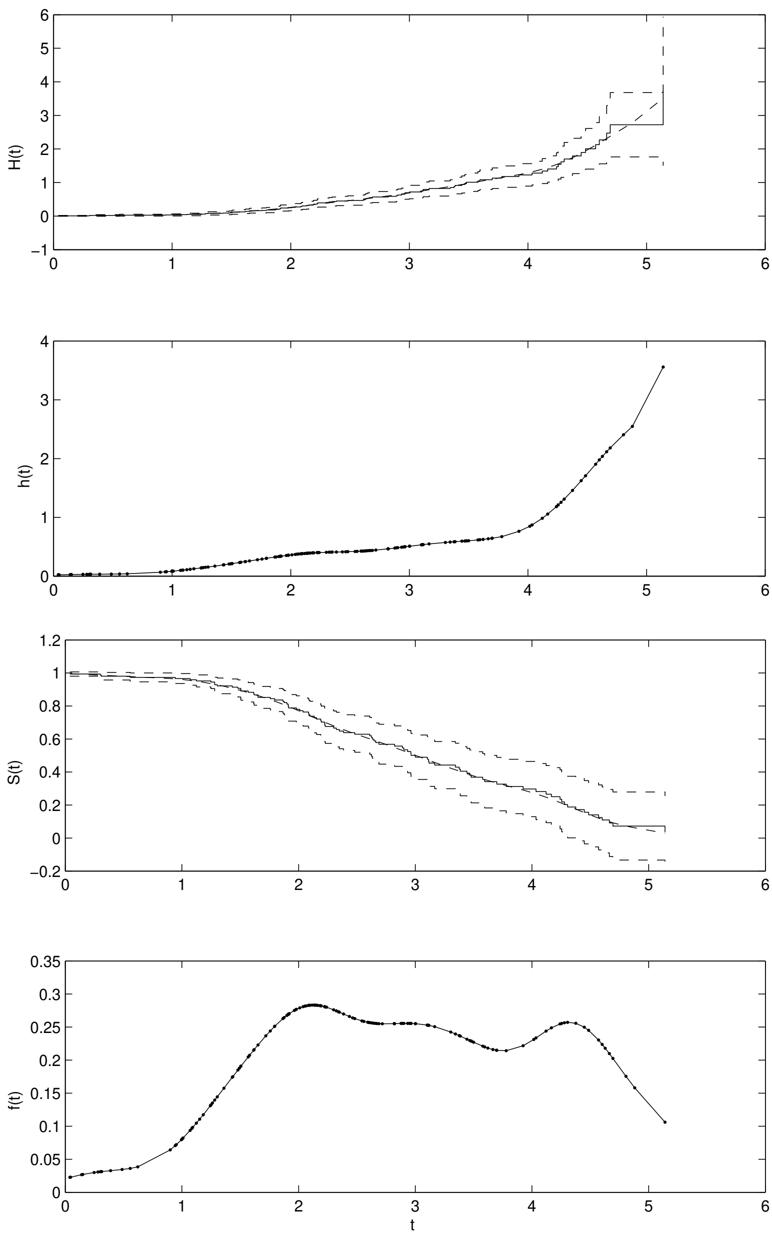

Figure 4: From above: the NAE with 95% CI-s, kernel-smoothed estimate of HR, then the PLE with 95% CI-s, kernel-smoothed estimate of density, data of Ruhi (2015).

of the log-likelihood. Figure shows that the fit should be improved yet. For comparison, achieved maximal value of the log-likelihood was -170.14, while the best result of Ruhi [8] was -176.7 obtained by the mixture of densities of Weibull and Normal distributions having 4 parameters of components plus one of mixture.

Estimated parameters of the sum of two Weibull hazard rates are in Table 2. It is seen that the first distribution can be taken as the exponential one,

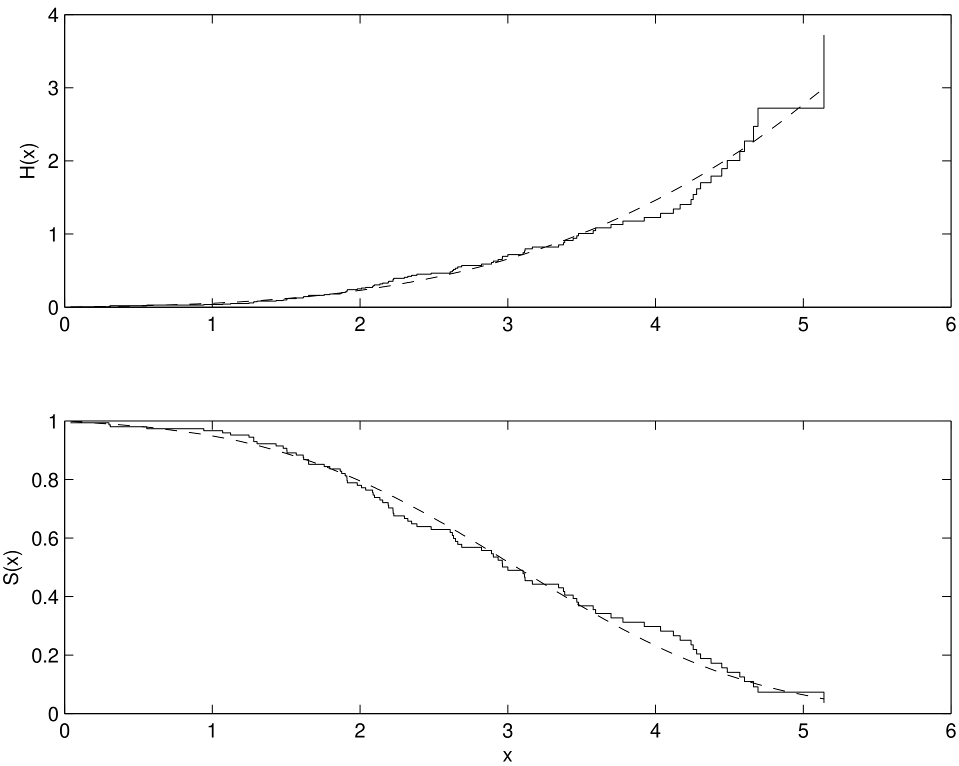

Figure 5: CHR above, with stepwise NAE, survival functions are below, with stepwise PLE. Dotted are kernel-smoothed nonparametric estimates, dashed curves correspond to mixed hazard rates model obtained by the MLE.

i.e. with $\beta_1 = 1$. Then the MLE of other parameters is comparable, as well as maximal log-likelihood. Now the model has just 3 parameters estimated as $\alpha_1 = 38.5612$, $\alpha_2 = 3.5717$, $\beta_2 = 2.9590$, with maximal log-likelihood -170.16.

Blischke [2] as well as Ruhi [8] considered also the Akaike information criterion (AIC) for comparison of different models performance. Let us recall that AIC$= 2k - 2L$, where k is the number of model parameters and L equals achieved maximum of the log-likelihood, the model with smaller AIC is regarded as being better. Hence, in the case of our model with $\beta_1 = 1$ AIC $= 346.3$, which is less than the AIC of all models considered in Blischke or Ruhi. The model has the form of the "nonlinear failure rate model" with $h(t) = a + bt^\gamma$, where $a = 1/\alpha_1 = 0.0259$, $b = \beta_2/(\alpha_2^{\beta_2}) = 0.0684$, $\gamma = \beta_2 - 1 = 1.9590$.

As it has been said, there are also other model fit methods available. For instance the use of the MCMC methods can improve the random search results, it also offers Bayes credibility intervals for parameters. Naturally, one may consider mixtures with also other hazard rates types, not limiting

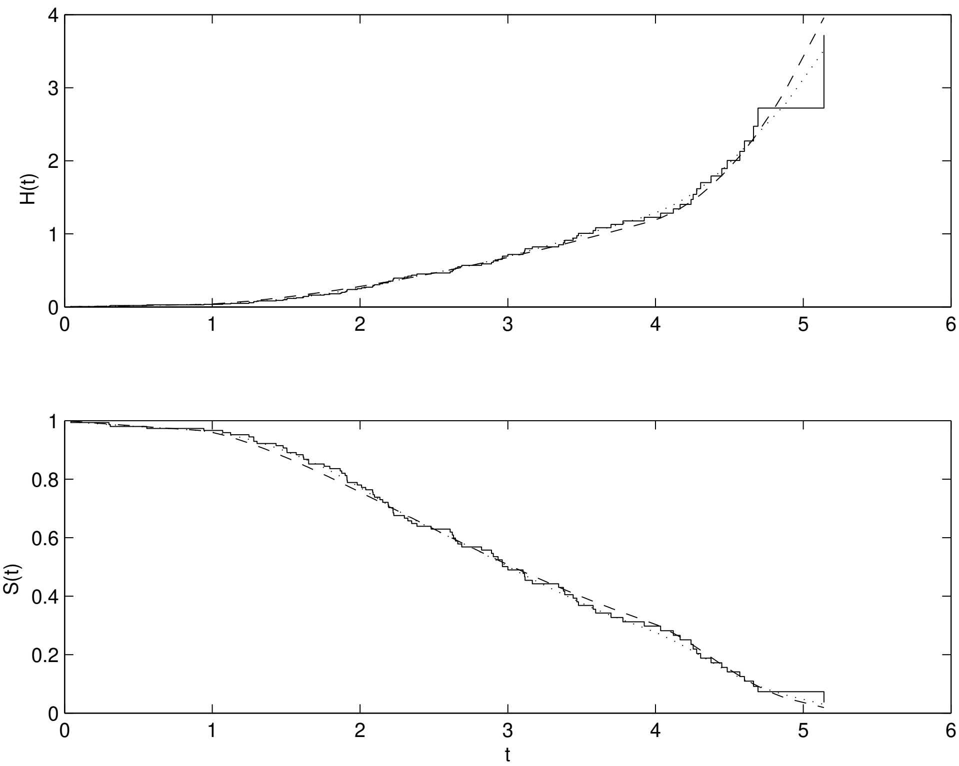

Figure 6: CHR above, with stepwise NAE, survival functions are below, with stepwise PLE. Dotted are kernel-smoothed nonparametric estimates, dashed curves corresponding to change-point model obtained by the MCMC.

himself just to Weibull, or to use more than two components. In fact, kernel estimate of density functions in Figure 4 below suggests that 3 component could be optimal. That is why the same data were analyzed also in the next section, with, it seems, a better result.

6. Change-point models for hazard rates

The task of change point detection belongs among quite well developed statistical techniques. There exist also results dealing with changes of hazard rates. While corresponding asymptotic theory connected with the MLE is rather complicated, cf. Nguyen et al. [6], practical analysis can be based on rather simple methods detecting a region where the residuals (i.e. reasonably defined deviation of data from actual model) are crossing a given border. Our problem could be viewed also as an incremental construction of a signal model, which means that in detected point of change a new component is added to current model. This is possible when the change leads to the increase of hazard.

Param.	α_1	α_2	α_3	β_1	β_2	β_3	T_{ch1}	T_{ch2}
Estim.	31.384	2.996	0.834	0.957	1.555	2.053	0.907	3.952
LCrI	20.792	2.400	0.760	0.762	1.304	2.118	0.754	3.075
UCrI	39.476	3.338	1.891	1.196	1.925	4.745	1.235	4.226
$\beta_1 = 1$	37.705	2.883	0.877	1.000	1.811	2.273	0.835	3.914

Table 3: Estimates of incremental model parameters and change-points, with LCrI and UCrI the lower and upper bounds of sample-based 90% credibility intervals. The last row – model with $\beta_1 = 1$.

The example shown in Figure 6 deals still with the data from Ruhi [8]. Figure displays again stepwise estimates (thick) and dotted smoothed estimates, the NAE of the CHR above and the PLE of survival function below. Further, dashed curves show the best "two change-points" model achieved. Namely, the model of hazard rate was constructed from one Weibull hazard rate starting from $t = 0$ and two Weibull hazard rates added at two times of change, T_{ch1}, T_{ch2}, which also had to be found optimally. Then $h(t) = h_1(t) + h_2(t - T_{ch1}) \cdot I[t > T_{ch1}] + h_3(t - T_{ch2}) \cdot I[t > T_{ch2}]$, $I[.]$ denotes the indicator function. Thus, the model had 8 parameters. The method of solution used the Metropolis MCMC algorithm generating a representation of Bayes posterior distribution of parameters, while their prior distributions were chosen to be independent uniform, in reasonable intervals. Therefore the posterior distribution was proportional to the likelihood. Together 50 000 iterations of the algorithm were performed, results were taken from last 20 000. Table 3 contains the "modes" of posterior distribution, i.e. the values maximizing the likelihood, and empirical quantiles representing the Bayes 90% credibility intervals. Achieved maximum of the log-likelihood was -165.5.

Similarly as above it is seen that the first component is close to exponential. When we fixed $\beta_1 = 1$, the results were comparable, they are in the last row of Table 3. The max of log-likelihood was -165.6. Model had just 7 parameters, with AIC = 345.2, which was even smaller than the best (regarding the AIC) result in the preceding section.

7. Concluding remarks

The problem of construction of hazard rate from additive components can be also interpreted as the problem of a regression model fitting the nonparametric estimate (e.g. smoothed from the NAE). To preserve some interpretation, the model should be constructed from a small set of given parametric function and kept non-negative.

In fact, the case studied here was simplified, just sums of Weibull hazard rates were considered. A next problem to explore is to use (convex) mixtures of hazard rates of other distributions and to study whether it is possible to estimate both parameters of distributions and coefficients of mixture and whether this task is unambiguous. Or, variantly, the flexibility of mixture models (based either on hazard rates or on probability densities) could be examined after the time is transformed (e.g. to the logarithmic scale).

Acknowledgements

The research was supported by the grant No. 18-02739S of the Grant Agency of the Czech Republic.

References

[1] Bain L. J.: Analysis for the linear failure-rate life-testing distribution. *Technometrics,* 16, 551–559, 1974.

[2] Blischke W. R., Karim M. R., Murthy D. P.: *Warranty data collection and analysis.* Springer, 2011.

[3] Briš R., Thach T. T.: Bayesian approach to estimate the mixture of failure rate model. *Proceedings of the 1st International Conference on Applied Mathematics in Engineering and Reliability,* CRC Press, 9–17, 2016.

[4] Kalbfleisch J. D., Prentice R. L.: *The Statistical Analysis of Failure Time Data.* Wiley, 2002.

[5] Kodlin D.: A new response time distribution. *Biometrics,* 2, 227–239, 1967.

[6] Nguyen H. T., Rogers G. S., Walker E. A.: Estimation in change-point hazard rate models. *Biometrika* 71, 299–304, 1984.

[7] Pandey A., Singh A., Zimmer W. J.: Bayes estimation of the linear hazard-rate model. *IEEE Transactions on Reliability,* 42, 636–640, 1993.

[8] Ruhi S.: Application of mixture models for analyzing reliability data: A case study. *Open Access Library Journal,* 2: e1815, 2015. `http://dx.doi.org/10.4236/oalib.1101815`

[9] Salem S. A.: Bayesian estimation of a non-linear failure rate from censored samples type II. *Microelectronics and Reliability*, 32, 1385–1388, 1992.

Obsah

Informační bulletin České statistické společnosti vychází čtyřikrát do roka v českém vydání. Příležitostně i mimořádné české a anglické číslo. Vydavatelem je Česká statistická společnost, IČ 00550795, adresa společnosti je Na padesátém 81, 100 82 Praha 10. Evidenční číslo registrace vedené Ministerstvem kultury ČR dle zákona č. 46/2000 Sb. je E 21214.

The Information Bulletin of the Czech Statistical Society is published quarterly.
The contributions in the journal are published in English, Czech and Slovak languages.

Předseda společnosti: RNDr. Marek Malý, CSc., Státní zdravotní ústav, Šrobárova 48, Praha 10, 100 42, e-mail: `mmaly@szu.cz`.

Redakce: prof. RNDr. Gejza Dohnal, CSc. (šéfredaktor), prof. RNDr. Jaromír Antoch, CSc., doc. Ing. Jozef Chajdiak, CSc., doc. RNDr. Zdeněk Karpíšek, CSc., RNDr. Marek Malý, CSc., doc. RNDr. Jiří Michálek, CSc., prof. Ing. Jiří Militký, CSc., doc. Ing. Iveta Stankovičová, PhD., doc. Ing. Josef Tvrdík, CSc., Mgr. Ondřej Vencálek, Ph.D.

Redaktor časopisu: Mgr. Ondřej Vencálek, Ph.D., `ondrej.vencalek@upol.cz`.
Informace pro autory jsou na stránkách společnosti, `http://www.statspol.cz/`.

DOI: 10.5300/IB, http://dx.doi.org/10.5300/IB
ISSN 1210–8022 (Print), ISSN 1804–8617 (Online)

Toto číslo bylo vytištěno s laskavou podporou Českého statistického úřadu.