Petr Volf; Aleš Linka

A counting process model of survival of parallel load-sharing system

Kybernetika, Vol. 37 (2001), No. 1, [47]--60

Persistent URL: http://dml.cz/dmlcz/135388

KYBERNETIKA — VOLUME 37 (2001), NUMBER 1, PAGES 47 - 60

A COUNTING PROCESS MODEL OF SURVIVAL OF PARALLEL LOAD-SHARING SYSTEM

PETR VOLF AND ALEŠ LINKA

A system composed from a set of independent and identical parallel units is considered and its resistance (survival) against an increasing load is modelled by a counting process model, in the framework of statistical survival analysis. The objective is to estimate the (nonparametrized) hazard function of the distribution of loads breaking the units of the system (i. e. their breaking strengths), to derive the large sample properties of the estimator, and to propose a goodness-of-fit test. We also examine the relationship between the survival of the system and the survival of its components.

1. INTRODUCTION

In the present contribution we model the reliability (survival) of a system composed from m parallel identical components. The reliability is understand as a resistance of the system against a load (strength, stress) causing its failure, the reliability of the system is derived from the reliability of the components.

We assume that the system is tested by a load increasing from 0 up to the level breaking the system (i. e. all its components) – or up to a given maximal load $S_{\max}$ when the experiment is terminated. Let the testing experiment be relatively fast, so that the time of duration of the stress does not influence the survival. We use more or less the standard survival analysis approach and the counting processes model, however, instead the time, the load per one component is the variable of interest. Simultaneously, we consider a rather simple scheme of re-distribution of the load among the components, namely the Daniels load-sharing model, see Crowder et al [5]. We assume that the breaking strengths of individual components are independent and identically distributed random variables, and that at each moment the load applied to the system is divided equally among the (unbroken) components. The same model has been used, for instance, in Belyaev and Rydén [4]. The global load affecting the system is observed. However, as the break of a component leads to immediate re-distribution of the load to the other components (so that to abrupt increase of the load per each component), the consequence can be an immediate break of several of remaining components. Therefore, in such a case of multiple breaks we observe directly the strength causing the break of the first component only. Moreover, we often are not able to register the order in which the components

broke, but in the case of identical components the information on the order is not important. The statistical analysis will use the values of directly observed (i. e. noncensored) data only. The censoring of the other breaks will be expressed by a properly defined observability indicator process, as it is common in the setting of the counting process model. Thus, we actually omit a part of information. On the other hand, the estimation and the proofs of the large-sample properties are then much more straightforward.

We have to admit that we deal with a rather simple description of a parallel system, that a more accurate model should consider for instance the (irreversible) consequences of shocks caused by the abrupt increase of the load per one component. Some cases, for instance the case of a wire or a textile yarn composed from a set of strands, are even more complicated due the elasticity of the material, or due the mutual dependence of twisted strands. Therefore, the present model is just a step to a more profound investigation of the reliability of systems.

The theme of reliability of a system composed from parallelly organized units has already been studied by a number of authors. In most instances the time to break under a constant stress has been analyzed. The random process approach was used for instance in Daniels [6], who examined the behaviour of maxima of certain Gaussian processes and with their aid he modelled the breaking strengths of a bundle of fibres. As we have already said, the starting point of our analysis is the model of the counting process characterized by a nonparametric hazard function. In the same setting, Belyaev and Rydén [4] proved the uniform consistency and the local asymptotic normality of the Nelson–Aalen estimate of the cumulative hazard function (C.H.F.) characterizing the probability distribution of the breaking strengths of components.

Our objectives are mostly methodological, i. e. we want to collect and propose a set of methods for modelling, computing, testing and simulation of reliability of a parallel system. Quite naturally, certain theoretical problems have to be solved, too. In Section 2 the counting process model of the breaking strengths of the components in a load-sharing system is recalled. In the beginning of Section 3, still following Belyaev and Rydén [4], we present the Nelson–Aalen estimator of C.H.F. of the breaking strength of one component. The main results are concentrated in Sections 3 and 4. After a rather trivial statement of Lemma 1, we offer a modified proof of uniform consistency in Theorem 1. Then the main Theorem 2 proposes the weak convergence of residual process to a Wiener process, on a whole interval. In Section 4 this global asymptotic normality is utilized for the formulation of a goodness-of-fit test. The proofs use the relevant results and theory available in a number of papers and monographs dealing with the counting processes models (e. g. Andersen and Borgan [1], Fleming and Harrington [7], Andersen et al [2]).

The problem how to derive the probability distribution of the breaking strength of the system, if the distribution of breaking strengths of its components is known, is discussed in Section 5. We recall both the computation approach proposed already in Suh et al [10] and the simulation method, and we compare them. In concluding Example 3 we consider also a more general case of two types of units with proportional hazard rates, and we analyze such a situation as a simple case of the Cox's

hazard regression model (Andersen and Gill [3]).

2. THE COUNTING PROCESS OF BREAKS OF COMPONENTS

Let us first consider one component and the random variable U – its breaking strength. We assume that U has a continuous distribution on $[0, \infty)$ with a distribution function $F(u)$, density $f(u)$, hazard function $h(u) = \frac{f(u)}{1-F(u)}$ defined on $u \in [0, S]$ such that $F(S) < 1$. By $H(u) = \int_0^u h(v)\,\mathrm{d}v$ we denote the cumulative hazard function. The 'fate' of a component during the increase of the load affecting it, u, is described by two random processes, by the counting process $N^1(u)$ and the indicator $I^1(u)$. $I^1(u) = 1$ if the load u affecting the component is observed, otherwise $I^1(u) = 0$. Namely, $I^1(u) = 0$ if the component is already broken or if the experiment is terminated. We assume that the trajectories of $I^1(u)$ are left-continuous. As regards $N^1(u)$, $N^1(0) = 0$ and $N^1(u)$ jumps to 1 at the load level u_b causing the observed break of the component (i. e. provided $I^1(u_b) = 1$). Trajectories of $N^1(u)$ are taken as right-continuous. The above description is actually the standard scheme of survival analysis, where the increasing load per component stands instead of time. As we assume a continuous distribution of U, we also consider a continuous scale of u. The difference in comparison with the standard survival analysis scheme consists in that we allow for abrupt jumps-up of the load affecting the component – in these intervals we set $I^1(u) = 0$, too.

2.1. The model of a parallel system

Let us now consider the system composed from m components, let the breaking strengths of components be described by i.i.d. random variables U_j, $j = 1, \ldots, m$, with distribution given by $f(u)$, $F(u)$, $h(u)$, $H(u)$, respectively. The following example illustrates the structure of observed data. Let us imagine that the breaks of components occurred for K 'global' loads affecting the system, $0 < s_1 < s_2 < \ldots < s_K < S_{\max}$, that k_j components broke on level s_j, with $\sum k_j = m$. Therefore, just before the first break the load per each component was $u_1 = s_1/m$, while just before the moment of the second break it was $u_2 = s_2/(m - k_1)$ (naturally affecting only $m - k_1$ remaining components) and, finally, immediately before the moment of the last break the load per each of last k_K components was $u_K = s_K/k_K$. Hence, only K breaks were observed directly, namely these caused by *known* loads per component u_j. Other breaks (if $k_j > 1$) were caused by unknown (unobserved) loads per component from intervals $(u_1, \overline{u}_1 = s_1/(m-k_1+1))$, $(u_2, \overline{u}_2 = s_2/(m-k_1-k_2+1))$, $\ldots$, $(u_K, \overline{u}_K = s_K)$, respectively for $k_1 - 1, k_2 - 1, \ldots, k_K - 1$ components. Our first aim is to analyze the distribution of U_j on $[0, S]$. We assume that the maximal load per system $S_{\max}$ is sufficiently large (e. g. $S_{\max} > S \cdot m$) in order not to terminate experiments too early.

Remark 1. Taking into account the assumption that the probability distribution of U_j is continuous, then (theoretically) there cannot occur two breaks at the same load per component level u. In other words, components break one after another,

not simultaneously (though sometimes we are not able to distinguish their order). The intervals of breaking strengths can be specified even more precisely than $(u_k, \overline{u}_k)$ above. However, as it has been said, our solution will not use the information about interval-censored strengths explicitly, but through a properly defined observability indicator process only. That is why we do not discuss the details of interval censoring here.

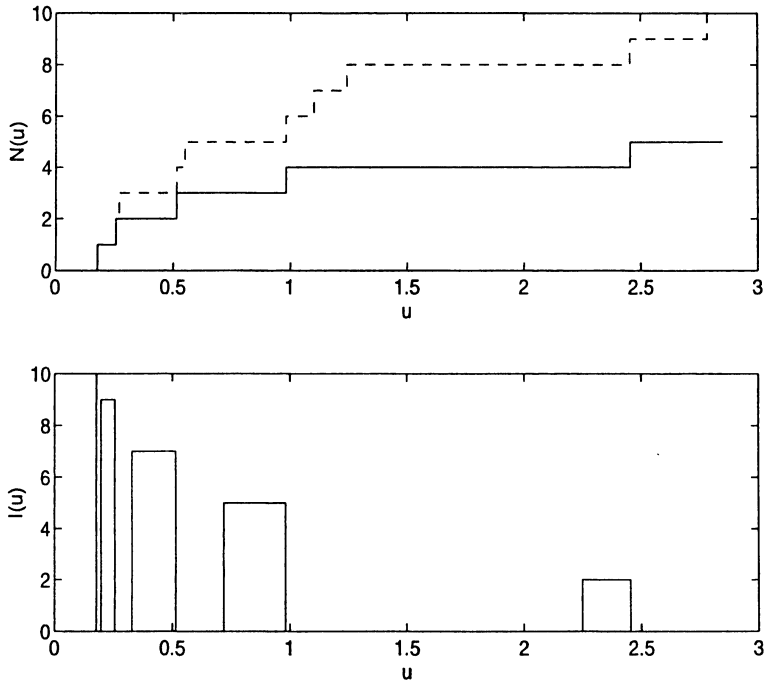

Fig. 1. Counting process $N(u)$ and indicator process $I(u)$.

Thus, the data used for analysis consist of $I(u) = \sum_{j=1}^{m} I_j^1(u)$, $N(u) = \sum_{j=1}^{m} N_j^1(u)$, i. e. the sums of indicators and observed counting processes of the components. By u we again denote the load per (unbroken) component. For the better explanation of the structure of observed data, let us display one example graphically. A system composed from $m = 10$ components has been simulated. Distribution of U_j followed the exponential distribution with the mean one. $K = 5$ successive breaks has been observed. Figure 1 shows the indicator process $I(u)$, the counting process $N(u)$ of observed breaks (full line), and also the underlying counting process of actual breaks of components (dashed line), which, in the real data cases, is not observed.

3. ESTIMATOR OF C.H.F., ASYMPTOTIC PROPERTIES

In the present part the estimator of the cumulative hazard function $H(u)$ (of distribution of the breaking strength of one component) is recalled and the uniform consistency and asymptotic normality of this estimator on interval $[0, S]$ are proven. Let us consider that n identical and independent systems are tested. Denote by U_{ij} the random variables – breaking strengths, by $N_{ij}(u)$, $I_{ij}(u)$ related individual counting and indicator processes for the jth component of the ith system ($j = 1, 2, \ldots, m$, $i = 1, 2, \ldots, n$). Further, denote

$$N_i(u) = \sum_{j=1}^{m} N_{ij}(u), \quad I_i(u) = \sum_{j=1}^{m} I_{ij}(u), \quad N(u) = \sum_{i=1}^{n} N_i(u), \quad I(u) = \sum_{i=1}^{n} I_i(u).$$

Let us first revoke some useful results of martingale theory connected with the counting processes (Andersen et al [2]). Individual counting processes $N_{ij}(u)$ are governed by their intensities, which, in our case, are $\lambda_{ij}(u) = h(u) \cdot I_{ij}(u)$. Cumulative intensities are $L_{ij}(u) = \int_0^u \lambda_{ij}(v)\,\mathrm{d}v$ and corresponding counting processes can be decomposed to the compensator and martingale, $N_{ij}(u) = L_{ij}(u) + M_{ij}(u)$. $M_{ij}(u)$ are local square-integrable martingales, with zero mean, mutually orthogonal and with variance process $\langle M_{ij} \rangle(u) = L_{ij}(u)$. Here $\langle\ \rangle$ denotes actually the process of conditional variance conditioned by the nondecreasing sequence of σ-algebras – the filtration – containing the *observed* history of the process. More precisely, $\sigma(u)$ is a σ-algebra constructed above the trajectories of $N_{ij}(v), I_{ij}(v), v < u$. For more details, see again Andersen et al [2], Fleming and Harrington [7]. The martingale-compensator decomposition, together with the law of large numbers and the central limit theorem applied to martingales, are the basis for the derivation of large-sample (asymptotic) properties of the estimator.

3.1. Nelson–Aalen estimator of C.H.F.

The most common estimator of the cumulative hazard function is the Nelson–Aalen one

$$\hat{H}_n(u) = \int_0^u \frac{1[I(v) > 0]}{I(v)}\,\mathrm{d}N(v),$$

where we set $0/0 = 0$. It is seen that the ability of the estimator to approximate well the 'true' $H(u)$ depends on the indicator process, i. e. on the observability of the counting processes for all values of strength u in the interval of interest $[0, S]$.

Let us assume that the number of tested systems, n, tends to infinity. Then it is also desirable that at each point $u \in [0, S]$ the number of observed unbroken components is of order n. Lemma 1 shows that such a property follows from the initial simple assumptions of identical, independent and continuous distribution of U_{ij} together with the assumption that $F(S) < 1$.

Lemma 1. There exists, with probability 1, a limit

$$\lim_{n\to\infty} \frac{I(u)}{n} = r(u)$$

for $u \in [0, S]$. Moreover, this limit is uniform w.r. to $u \in [0, S]$ and $r(u) \geq \varepsilon$ for some $\varepsilon > 0$.

Proof.

1. For each fixed u, $I_i(u)$ are *i.i.d.* random variables, with values from $\{0, 1, 2 \ldots, m\}$. Hence, the law of large numbers yields the almost sure convergence

$$\frac{1}{n} \sum_{i=1}^{n} I_i(u) \to r(u) = E\{I_1(u)\}.$$

2. From assumptions that random variables U_{ij} are *i.i.d.* and that $F(S) < 1$ it follows that for each $u \leq S$ there is a positive probability $[1 - F(u)]^m$ that all components of the system survive u. Therefore, $r(u) \geq m \cdot [1 - F(u)]^m \geq m \cdot [1 - F(S)]^m > 0$. The last expression can be used as ε in Lemma 1.

3. It remains to prove the uniformity of convergence. Random functions $I_i(u)$ are mutually independent, with the same distribution. They have maximally $2m$ finite jumps (m down and up). We shall use the results collected in Hoffmann–Jørgensen [8] (Volume II, Parts 9.13 to 9.17). We can imagine that each trajectory of $I_i(u)$ is given by a bounded, piecewise-constant nonrandom function $b(\mathbf{v}, \mathbf{k}; u)$, where $\mathbf{v}, \mathbf{k} = (v_1, v_2, \ldots, v_m, k_1, k_2, \ldots, k_m)$ are the realizations of random vector $\mathbf{V}, \mathbf{K} = (V_1, K_1, V_2, K_2, \ldots, V_m, K_m)$, K_j are levels of $I_i(u)$ and V_j are its points of jumps down. Again, for different i-s these vectors are *i.i.d.*, $\mathbf{K}$ is bounded, $\mathbf{V}$ has continuous distribution. Hence, the proof of Lemma 1 follows from the boundedness in the mean and from a.s. piecewise continuity of functions $b(\cdot; u)$ with respect to the distribution of $\mathbf{V}, \mathbf{K}$ (cf. again Hoffmann–Jørgensen [8], Theorem 9.17 on uniform convergence). □

From the uniform convergence and from the boundedness of jumps of $I(u)$ it also follows that the limit function $r(u)$ is continuous on $[0, S]$.

Remark 2. The statement of Lemma 1 implies that for every $\delta > 0$, for sufficiently large $n > n_\delta$, $Pr\{I(u) \geq \delta$ on the whole $[0, S]\} = 1$. Hence, with probability one it also holds that $1[I(u) = 0] = 0$ on $[0, S]$, and also $\sqrt{n} \int_0^S 1[I(u) = 0] \mathrm{d}H(u) = 0$. Such a property corresponds to one of conditions required in Andersen et al [2] (Theorems IV.1.1 and IV.1.2) for the consistency and asymptotic normality of the Nelson–Aalen estimator.

3.2. Asymptotic properties

1. Consistency. The uniform consistency of $\hat{H}_n$ has already been proved in Belyaev and Rydén [4]. We shall prove the same result with the aid of Lemma

1. Let us examine residuals $\hat{H}_n(u) - H(u)$ for $u \in [0, S]$. Denote $M_i = \sum_{j=1}^{m} M_{ij}$, $M = \sum_{i=1}^{n} M_i$. From the martingale-compensator decomposition we obtain

$$\begin{aligned}
\hat{H}_n(u) - H(u) &= \int_0^n \frac{\sum\sum \mathrm{d}N_{ij}(v)}{I(v)} 1[I(v) > 0] - H(u) \\
&= \int_0^u \frac{\sum\sum \mathrm{d}M_{ij}(v)}{I(v)} 1[I(v) > 0] + \int_0^u \frac{\sum\sum \mathrm{d}L_{ij}(v)}{I(v)} 1[I(v) > 0] - H(u) \\
&= \int_0^u \frac{\mathrm{d}M(v)}{I(v)} 1[I(v) > 0] + \int_0^u \frac{\sum\sum h(v) I_{ij}(v)\,\mathrm{d}v}{I(v)} 1[I(v) > 0] - H(u) \\
&= \frac{1}{n} \int_0^u \frac{\mathrm{d}M(v)}{I(v)/n} 1[I(v) > 0] - \int_0^u 1[I(v) = 0] \mathrm{d}H(v). \qquad (1)
\end{aligned}$$

Theorem 1. $\hat{H}_n(u)$ is an *a.s.*-consistent estimate of $H(u)$ on $[0, S]$. Moreover, this consistency is uniform w.r. to $u \in [0, S]$, i. e. $\sup_{u \in [0,S]} |\hat{H}_n(u) - H(u)| \to 0$ *a.s.*

Proof. Processes $\frac{1}{n} \int_0^u \mathrm{d}M(v) = \frac{1}{n} \sum_{i=1}^{n} M_i(u)$ have zero mean, M_i are mutually independent. Moreover, as $M_i(u) = N_i(u) - \int_0^u h(s) I_i(s)\,\mathrm{d}s$, they are uniformly bounded on $[0, S]$. Therefore, at fixed u, $\frac{1}{n} \int_0^u \mathrm{d}M(s) \to 0$ *a.s.* (it follows from the law of large numbers).

Uniform convergence $\sup_{u \in [0,S]} \frac{1}{n} \int_0^u \mathrm{d}M(s) \to 0$ *a.s.* can be proved similarly as in the preceding Lemma 1. We can represent $N_i(u)$ by a piecewise constant function $c(\mathbf{v}; u)$ which has maximally m steps $+1$ at points $v_1, v_2, \ldots, v_m$, and $\int_0^u h(s) I_i(s)\,\mathrm{d}s$ can be represented by a continuous and bounded function $\int_0^u h(s) b(\mathbf{v}, \mathbf{k}; s)\,\mathrm{d}s$, where b is a function defined in the proof of Lemma 1, $\mathbf{v}, \mathbf{k}$ are *i.i.d.* realizations of $\mathbf{V}, \mathbf{K}$ (also the same as in the proof of Lemma 1). Then, Theorem 9.17 of Hoffmann–Jørgensen [8] can again be applied to the proof of the *a.s.* uniform convergence $\frac{1}{n} \int_0^u \mathrm{d}M(s) \to 0$, on $[0, S]$.

From this and further from Lemma 1 and Remark 2 the statement of Theorem 1 follows immediately. □

2. Asymptotic distribution. Let us now analyze the behaviour of the process $\sqrt{n}(\hat{H}_n(u) - H(u))$ on $[0, S]$, for $n \to \infty$. Similarly as in (1), we obtain

$$\sqrt{n}(\hat{H}_n(u) - H(u)) = \sqrt{n} \int_0^u \frac{1[I(v) > 0]}{I(v)}\,\mathrm{d}M(v) - \sqrt{n} \int_0^u 1[I(v) = 0]\,\mathrm{d}H(v). \quad (2)$$

Taking into consideration the uniform convergence of $I(v)/n$ given in Lemma 1, the statement of Remark 2, and the boundedness of jumps of $\mathrm{d}M(v)$ (jumps are less or equal to m), we immediately obtain the following theorem specifying the asymptotic distribution of residual process.

Theorem 2. Random process $\sqrt{n}(\hat{H}_n(u) - H(u))$ converges weakly on $[0, S]$ to a Gauss random process with independent increments, zero mean and with variance

function

$$w(u) = \int_0^u \frac{\mathrm{d}H(v)}{r(v)}.$$

In other words, the process is asymptotically distributed as $W(w(u))$, where $W(.)$ is a Wiener process.

P r o o f . The proof follows directly either from the central limit theorem for martingales (e. g. Andersen et al [2], part II.5.) or from Theorem 3.2. of Andersen and Borgan [1]. It remains to show the convergence of the variance process and to compute the exact form of its limit. The convergence follows from our Lemma 1 and from the boundedness of both $H(u)$ and $r(u)$ on $[0, S]$, namely

$$\begin{aligned} \text{as var} &\sim \operatorname{var}\{\sqrt{n} \int_0^u \frac{\mathrm{d}M(v)}{I(v)}\} = n\,E \int_0^u \frac{\langle \mathrm{d}M \rangle (v)}{I^2(v)} \\ &= n\,E \int_0^u \frac{\mathrm{d}L(v)}{I^2(v)} = E \int_0^u \frac{h(v)\,\mathrm{d}v}{I(v)/n} \to \int_0^u \frac{\mathrm{d}H(v)}{r(v)}. \end{aligned}$$

□

4. GOODNESS–OF–FIT TEST

Let a hypothetical model be given by a cumulative hazard function $H^0(u)$. We want to decide whether the data correspond to it. The data are represented by observed trajectories of $N_i(u)$, $I_i(u)$, $i = 1, \ldots, n$. The tests are quite naturally based on the comparison of $\hat{H}_n(u)$ with expected $H^0(u)$.

Graphical test. Let us order all observed strengths breaking the components into one nondecreasing sequence u_k, $k = 1, \ldots, K$. For the graphical comparison, we plot the values

$$L(u_k) = \int_0^{u_k} \mathrm{d}H^0(v) I(v)$$

against $N(u_k) = k$ on the abscissa. If the model holds the residual process $L(u) - N(u)$ is a martingale. Then it is expected that the curve $L(u_k)$ will be close to the line $y(k) = k$. An opposite case (e. g. expanding distance of both curves) indicates that the model $H^0(u)$ does not correspond to the data. Approximate critical bounds for such a comparison can be derived e. g. from the following numerical procedure.

Numerical test. Numerical test is based on asymptotic distribution. From the result of Theorem 2 it follows that the process

$$D_n(u) = \sqrt{n}(\hat{H}_n(u) - H^0(u))\,/(1 + w(u))$$

is (if the model holds) asymptotically distributed as a Brownian bridge process $\mathcal{B}(\tau(u))$, where $\tau(u) = w(u)/(1 + w(u))$, $u \in [0, S]$. Hence, a test of Kolmogorov–Smirnov type can be used. From the theory of Brownian bridge it follows, for

instance, that if $d \geq 0$,

$$P\left(\max_u D_n(u) \geq d\right) = P\left(\min_u D_n(u) \leq -d\right) \approx \exp(-2d^2)$$

approximately. So that the value $\exp(-2d^2)$, where d is observed $\max_k |D_n(u_k)|$, is an approximate p-value for the test of hypothesis of the goodness-of-fit against a proper one-sided alternative. Simultaneously it holds that the critical value of the two-sided test, on level α, namely the value $d(\alpha)$ fulfilling

$$P\left\{\sup_u |D_n(u)| > d(\alpha)\right\} = \alpha,$$

can be approximated by $d(\alpha) = \sqrt{\ln(\frac{2}{\alpha})\frac{1}{2}}$. It follows that the approximate $(1-\alpha)$ confidence region for 'true' $H(u)$ is the band $\hat{H}(u) \pm d(\alpha)(1+w(u))/\sqrt{n}$. A more precise critical values can be obtained from the relevant results on the Brownian bridge process and on its probability of crossing a given level. An example of the test is provided in Section 6, Example 2.

5. DISTRIBUTION OF BREAKING STRENGTH OF SYSTEM

Let us now assume that we know the characteristics of breaking strengths distribution of individual components (e. g. the distribution function $F(u)$) and our aim is to compute the reliability for the whole system composed from m such components. Though such a problem has already been considered elsewhere, for instance in Suh et al [10], we think that it is useful to recall this approach in order to complete the set of methods presented in the paper. More precisely, let the probability that the system will not survive the (global) load s be given by the distribution function $F_R(s) = P(R < s)$, where R is the random variable describing the breaking strength of the system. If we denote by $U_{(1)} < U_{(2)} < \ldots < U_{(m)}$ the order statistics created from the random strengths breaking individual components of the system, $U_1, U_2, \ldots, U_m$, then evidently

$$F_R(s) = P(R < s) = P\left\{\cap_{k=1}^{m}\left[U_{(k)} < \frac{s}{m-k+1}\right]\right\},$$

which can be computed from the joint distribution of order statistics $U_{(1)}, \ldots, U_{(m)}$. Though such a distribution is well-known (see e. g. Rao [9], Chapter 3.6.), the computation of the joint distribution function is not easy. In our case a sequential computation yields that

$$F_R(s) = m!\, A_m(s),$$

where $A_0(s) = 1$ and

$$A_k(s) = \sum_{j=1}^{k} \frac{(-1)^{j-1}}{j!} A_{k-j}(s)\, F^j(\frac{s}{k}). \tag{3}$$

Another simple and universal approach to the evaluation of distribution of random variable R consists in the simulation. The following example illustrates and compares both methods.

6. EXAMPLES

Example 1

Let us consider a system composed from $m = 10$ components and assume that the breaking strength of each component (i.e. random variable U_{ij}) has the standard exponential distribution (i.e. with $EU_{ij} = 1$). We simulated the breaks of $n = 200$ such systems. The results observed for one of them are already in Figure 1. Naturally, the global load under which the system broke was observed, too. We thus obtained a sample of $n = 200$ independent realizations r_i of random variable R – the breaking strength of the system. The empirical distribution function $\hat{F}_R(s) = \frac{1}{n}\sum 1[r_i < s]$ constructed from this sample is displayed in Figure 2a and compared with $F_R(s)$ computed from (3). Other empirical characteristics can be easily derived, too. For instance, the estimate of cumulative hazard function can be obtained either as $\hat{H}_R(s) = -\ln(1-\hat{F}_R(s))$ (see Figure 2b) or directly from the ordered sample: Let (i) be the order of r_i in $r_1, r_2, \ldots, r_n$, then the standard Nelson–Aalen estimator is $\tilde{H}_R(s) = \sum_{i=1}^{n} \frac{1[r_i \le s]}{n-(i)+1}$.

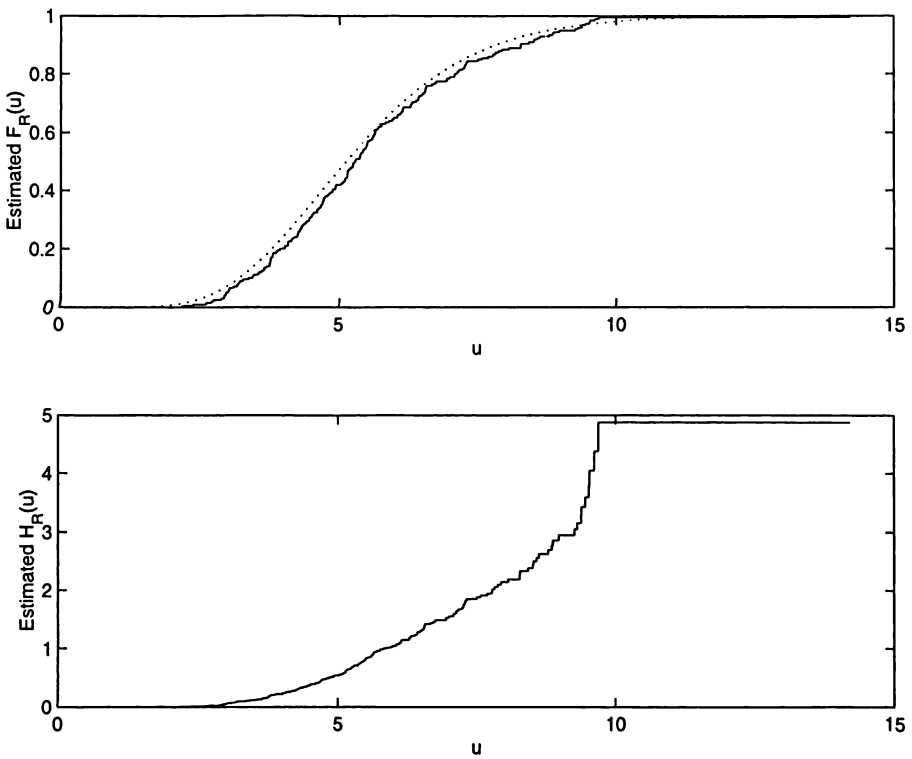

Fig. 2. Estimated $\hat{F}_R(s)$ and $\hat{H}_R(s)$ of distribution of random variable R, compared with dotted $F_R(s)$ computed from (3).

Example 2

In this example of the numerical goodness-of fit test we use the data simulated in Example 1. We intend to test the hypothesis $\mathbf{H_0}$ that the data really correspond to the standard exponential distribution, on interval $u \in [0, 5]$. Therefore, we should compute $\hat{H}_n(u)$, estimate $w(u)$ by

$$\hat{w}_n(u) = \int_0^u \frac{n \cdot \mathrm{d}\hat{H}_n(v)}{I(v)},$$

and find the maximum of $|D_n(u)|$, assuming that the hypothetical C.H.F. of standard exponential distribution is $H^0(u) = u$.

The maximal and minimal observed values of $D_n(u)$ on $[0, 5]$ were $d^+ = 0.0158$, $d^- = -0.0711$. We then computed approximate critical value for the test level $\alpha = 5\,\%$, $d(\alpha) = 1.3581$. As it was considerably greater than $d = \max(d^+, -d^-) = 0.0711$, the hypothesis $\mathbf{H_0}$ was not rejected (on approximately 5 % level of test significance). Estimated $\hat{H}_n(u)$ together with hypothetical $H^0(u) = u$ and approximate 95 % confidence bands are displayed in Figure 3.

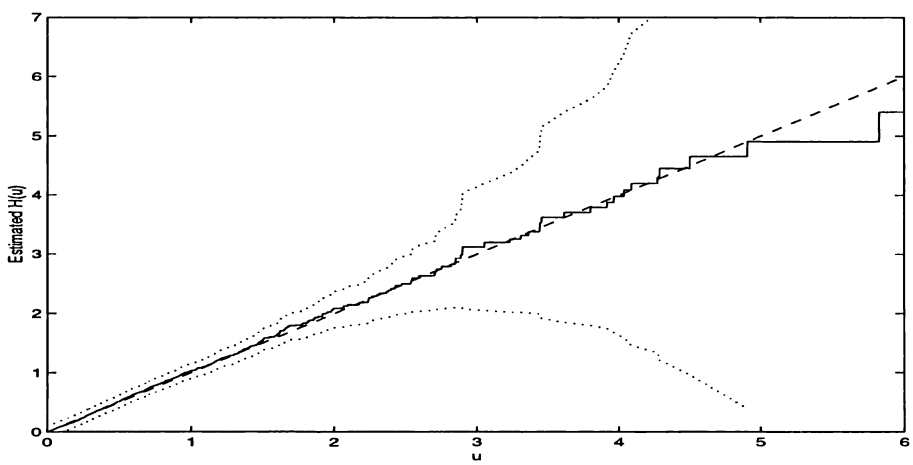

Fig. 3. Cumulative hazard function for individual components: estimated $\hat{H}_n(u)$ (full line), hypothetical $H^0(u) = u$ (dashed), and asymptotic 95 % confidence bands for $H(u)$ (dotted).

Example 3 – A system with non-identical units

Let us now consider a simple case of components with different breaking strengths, namely the case of the *proportional hazard model* with only two types of components. Let a standard type have the hazard function $h_0(s)$, while the other type has the hazard function $h_1(s) = c \cdot h_0(s)$, $c \geq 0$. Equivalently, the situation can be described via the hazard functions of components $h(s) = h_0(s) \exp(bX)$, where $b = \ln c$ and X is a random variable with $p = P(X = 1), 1 - p = P(X = 0)$. In the follow up

we shall assume that the configuration of x's is known, i.e. we are able to match a certain value x to each observed broken or censored component. Then the case can be regarded as a simple version of Cox's regression model (with only two levels of regressor). The objective of statistical data analysis is to estimate parameter b and function $h_0(s)$, respectively its cumulative version $H_0(u) = \int_0^u h_0(s)\,\mathrm{d}s$, on [0,S]. It is well known (cf. Andersen and Gill [3]) that such an estimation problem is solvable consistently, moreover with estimates possessing the property of asymptotic normality.

Assumptions. We assume that $H_0(S) < \infty$ and that $1 > p > 0$. These assumptions actually suffice for the validity of conditions (given in Andersen and Gill [3]) ensuring the desirable large sample properties of estimates.

Let us denote

$$I^1(s) = \sum_i \sum_j I_{ij}(s)\mathbf{1}[X_{ij} = 1], \quad I^0(s) = \sum_i \sum_j I_{ij}(s)\mathbf{1}[X_{ij} = 0],$$

let $N^1(s)$ and $N^0(s)$ be defined in a similar way. Then a variant of Lemma 1 (with a quite analogical proof) holds:

Lemma 2. There exist, with probability one, uniform limits

$$r_1(s) = \lim_{n\to\infty} \frac{I^1(s)}{n}, \quad r_0(s) = \lim_{n\to\infty} \frac{I^0(s)}{n},$$

which are positive, bounded, and also bounded away from zero on $[0, S]$.

Estimation. In the framework of the proportional hazard model, the estimation has two stages. First, the parameter b of proportionality is estimated from relevant partial likelihood (which actually can be derived from the full likelihood, because it is a 'profile' likelihood of b). Its logarithm, after some simplification, reads

$$\ln L_p(b) = \int_0^S b\,\mathrm{d}N^1(s) - \int_0^S ln\{I^0(s) + \exp(b)I^1(s)\}\,\mathrm{d}N(s). \tag{4}$$

Optimal $\hat{b}$ (the maximizer of (4)) is obtained from the solution of equation $\mathrm{d}\ln L_p/\mathrm{d}b = 0$, via the Newton–Raphson algorithm (or via another iterative procedure). In such a simple case considered here the solution is unique and, as the second derivative of (4) is negative, the maximum of $\ln L_p(b)$ can be reached practically from arbitrary starting value of the iteration procedure. In practical examples, the Newton–Raphson algorithm converged as a rule in less than 10 steps. The next stage consists in the estimation of the cumulative baseline hazard function, by the Newton–Aalen type estimator (in the regression context called the Breslow–Crowley one):

$$\hat{H}_0(u) = \int_0^u \frac{1[I^0(s) + I^1(s) > 0]}{I^0(s) + \exp(\hat{b})I^1(s)}\,\mathrm{d}N(s). \tag{5}$$

As we have already said, the large sample properties (consistency and asymptotic normality) follow from the results derived for a more general case of Cox's model.

Numerical example. We generated a sample of 200 'systems', each composed from 10 components with the breaking strengths given by exponential distributions: 5 standard components had the mean 1, 5 stronger components had the mean 2. So that $H_0(u) = u$ and the proportionality parameter $c = \exp(b) = 0.5$. After a fast and short iteration we obtained the estimate of $b = -0.7072$, i. e. of $c = 0.4930$, with approximate 95 % confidence interval, based on the asymptotic normality of estimate of b, (0.4358, 0.5577). Estimated cumulative baseline hazard function in Figure 4 shows a linear trend with the slope close to one. Approximate 95 % confidence bands for $H_0(u)$ computed in accordance with the results of Andersen and Gill [3] are displayed by dotted lines.

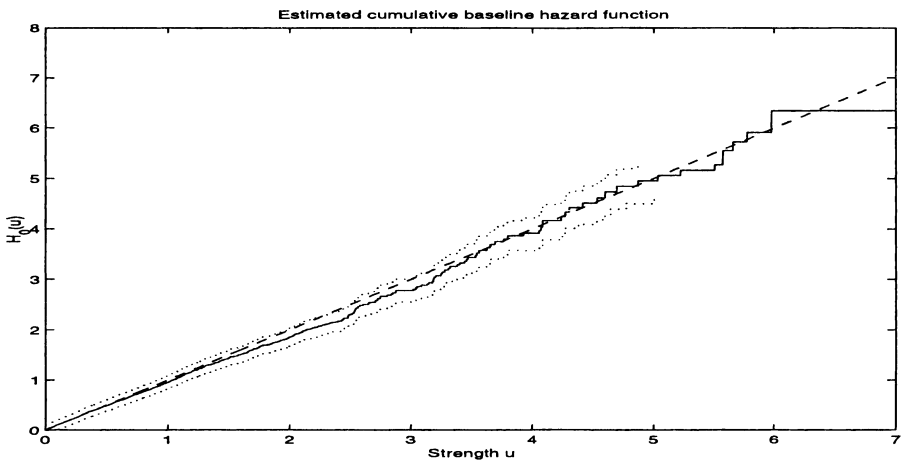

Fig. 4. Final estimate of cumulative baseline hazard function, with asymptotic confidence bands (dotted) and $H_0(u) = u$ (dashed).

7. CONCLUSION

We presented a set of procedures for the probabilistic modelling and statistical analysis of breaking strengths in a system of parallel components. The data were treated as the lifetime data, with the increasing load per one component as the leading variable. We studied the asymptotic behaviour of the estimator of the cumulative hazard function describing the breaking strength distribution. In particular, asymptotic normality of the estimator on the whole interval was proved and on this basis the goodness-of-fit test was proposed. Such a test can for instance be useful for the assessing the agreement of observed breaking strength data with the expected resistance of the system (e. g. with the resistance guaranteed by the producer).

In the concluding example we also showed how the approach used in the present work could be generalized to a system of units with nonequal reliability. Another generalization can consist in models considering simultaneously the load, the time and/or the cumulated load, in the framework of the hazard regression model.

ACKNOWLEDGEMENTS

We express our gratitude to P. Rydén from the University of Umea, Sweden, for his criticism of the initial version of the paper and for his comments to our work.
The research has been supported by Grant 402/98/0472 of the Grant Agency of the Czech Republic and by Grant VŠ 97084 of the Ministry of Education of the Czech Republic.

(Received October 11, 1999.)

REFERENCES

[1] P. K. Andersen and O. Borgan: Counting process models for life history data: A review. Scand. J. Statist. *12* (1985), 97–158.
[2] P. K. Andersen, O. Borgan, R. D. Gill, and N. Keiding: Statistical Models Based on Counting Processes. Springer, New York 1993.
[3] P. K. Andersen and R. D. Gill: Cox's regression model for counting processes: A large sample study. Ann. Statist. *10* (1982), 1100–1120.
[4] Yu. K. Belyaev and P. Rydén: Non-parametric estimators of the distribution of tensile strengths for wires. Research Report 1997-15, University of Umea 1997.
[5] M. J. Crowder, A. C. Kimber, R. L. Smith, and T. L. Sweeting: Statistical Analysis of Reliability Data. Chapman and Hall, London 1991.
[6] H. E. Daniels: The maximum of a Gaussian process whose mean path has a maximum, with an application to the strength of bundles of fibres. Adv. in Appl. Probab. *21* (1991), 315–333.
[7] T. R. Fleming and D. P. Harrington: Counting Processes and Survival Analysis. Wiley, New York 1991.
[8] J. Hoffmann–Jørgensen: Probability with a View Towards Statistics, Volume II. Chapman and Hall, London 1994.
[9] C. R. Rao: Linear Statistical Inference and Its Application. Wiley, New York 1965.
[10] M. W. Suh, B. B. Bhattacharyya, and A. Grandage: On the distribution and moments of the strength of a bundle of filaments. J. Appl. Probab. *7* (1970), 712–720.
[11] P. Volf: Analysis of generalized residuals in hazard regression models. Kybernetika *32* (1996), 501–510.

Petr Volf, CSc., Institute of Information Theory and Automation – Academy of Sciences of the Czech Republic, Pod vodárenskou věží 4, 182 08 Praha 8. Czech Republic.
e-mail: volf@utia.cas.cz

RNDr. Aleš Linka, CSc., Department of Textile Materials, Technical University Liberec, Hálkova 6, 461 17 Liberec. Czech Republic.
e-mail: ales.linka@vslib.cz