

ON STATISTICAL METHODS FOR SURVIVAL DATA ANALYSIS

JAN HURT

ABSTRACT. In the first part the basic probability and statistical concepts needed in survival data analysis are reviewed. The second part is devoted to parametric statistical inference, particularly to Weibull model. In the third part, recent nonparametric results are discussed. The fourth part deals with some aspects of using selected statistical packages for survival data processing. Some of new trends and the state of art are mentioned in the fifth part.

I. INTRODUCTION

In what follows we will suppose that X is a *lifetime or survival time*, that is, X is a non-negative¹ random variable possessing an absolutely continuous *distribution function* F_X , *survival function* $R_X = 1 - F_X$, *Lebesgue density* f_X , and the corresponding *hazard rate function* $r_X(x) = f_X(x)/R_X(x)$ defined for all x such that $R_X(x) > 0$. Furthermore, let $\Lambda_X(x) = \int_0^x r_X(t)dt$ denote the *cumulative hazard rate*². We just mention two important relationships concerning the hazard rate:

$$r_X(x) = \lim_{\Delta \rightarrow 0} \frac{1}{\Delta} P(x \leq X < x + \Delta | X \geq x)$$

$$R_X(x) = \exp\left\{-\int_0^x r_X(t)dt\right\} = \exp\{-\Lambda_X(x)\}.$$

The purpose of this paper is to present selected statistical methods for *censored samples*. This is a common situation in survival analysis problems. We will use the model of random censorship where the data are censored from the right. This type of censorship is often met in many applications, especially in clinical research or in life testing of complex technical systems. Together with the survival time X we will consider another nonnegative random variable T called *time censor*. Under the random censorship setup we can only observe the pair

$$W = \min(X, T) \quad I = \{X \leq T\} = I\{W = X\}.$$

It means that we observe either death, i. e., the survival time X , or we know that an individual has survived at least time T . Also an important piece of information is in I - we know *what* has happened first. In the next we will suppose that the distribution of T is *absolutely continuous* and that

X and T are independent.

Suppose that X_1, \dots, X_n are *true* (hypothetical, not all observable) survival times of n individuals censored by independent identically distributed (i.i.d.) random variables T_1, \dots, T_n from the right. The experiment thus results in observing n pairs

$$(I.1) \quad (W_1, I_1), \dots, (W_n, I_n)$$

¹ The only exception will be Example II.1.5 dealing with the extreme value distribution.

² In many applications the cumulative hazard rate becomes now more popular and more often used than the hazard rate.

where $W_j = \min(X_j, T_j)$, $I_j = I\{X_j \leq T_j\}$, $j = 1, \dots, n$. (I.1) is a (complete) two-dimensional random sample. Let us use the symbols F, R, f, r introduced above, with corresponding indices also for forthcoming random variables. Due to the independence of X and T we have

$$(I.2) \quad R_W(w) = 1 - F_W(w) = F_X(w)F_T(w)$$

Obviously, $\sum_{j=1}^n I_j$ represents the number of uncensored observations and $nEI_1 = nP(X < T)$ is the expected number of uncensored observations.

It is not difficult to show that the joint distribution of (W_1, I_1) possesses the density

$$(I.3) \quad h(w, i) = [f_X(w)R_T(w)]^i [f_T(w)R_X(w)]^{1-i}, \quad i = 0, 1, \quad w \geq 0$$

with respect to Lebesgue \times counting product measure.

In parametric inference a natural question arises about the distribution of the time censor T . There are good both theoretical and practical reasons to adopt the KOZIOL AND GREEN model of random censorship (Koziol and Green, 1976) under which it is assumed that there is a nonnegative constant γ such that

$$(I.4) \quad \boxed{R_T = R_X^\gamma}$$

Formally, $\gamma = 0$ corresponds to the case of no censoring - we observe all survival times. Let us denote

$$p = P(X < T),$$

the expected ratio of uncensored observations.

Remark I.1. Under the Koziol-Green model the density (I.3) becomes

$$h(w, i) = f_X(w)R_X^\gamma \gamma^{1-i}, \quad i = 0, 1, \quad w \geq 0.$$

Theorem I.1. If $p \in (0, 1)$ we have

$$p = \frac{1}{1 + \gamma}, \quad R_W = R_X^{1/p}, \quad r_W = \frac{1}{p}r_X, \quad \Lambda_W = \frac{1}{p}\Lambda_X,$$

$$f_W = \frac{1}{p}R_X^\gamma f_X, \quad r_T = \gamma r_X, \quad \Lambda_T = \gamma \Lambda_X.$$

Proof. See (Herbst, 1992a), e.g. \square

Note that the family of distributions generated by hazard rates

$$(I.5) \quad r(\cdot; \varrho) = \varrho r_X(\cdot), \quad \varrho > 0$$

is sometimes called *Lehmann or proportional hazards family* based on r_X (Cox and Oakes, 1984).

Theorem I.2 (Characterization of the Koziol-Green model). *There is a $\gamma > 0$ such that (I.4) holds iff W_j and I_j are independent, $j = 1, \dots, n$.*

Proof. Despite the result can be found in (Chen, Hollander and Langberg, 1982), the correct proof comes from (Herbst, 1992a). In loc. cit. the proof is given even without assuming F_X and F_T absolutely continuous. \square

Another characterization of the Koziol-Green model can be also found in (Herbst, 1992a):

Theorem I.3. *There is a $\gamma > 0$ such that (I.4) holds iff the conditional distributions $\mathcal{L}(W_j | I_j = 0)$ and $\mathcal{L}(W_j | I_j = 1)$ are identical.*

II. PARAMETRIC INFERENCE

With the exception of the best unbiased (=minimum variance unbiased) estimator (BUE) of the hazard rate of an exponential distribution under Koziol-Green model of random censorship, BUE's are not known for survival characteristics. Therefore, the parametric inference is mainly based on the principle of *maximum likelihood* and on *Bayesian inference*.

II.1. Maximum likelihood

Suppose that f_X depends on an unknown (possibly vector) parameter θ_1 and f_T depends on θ_2 . Denote $\theta = (\theta_1, \theta_2)$, $\theta_1 = (\theta_{11}, \dots, \theta_{1k})$. Let us further denote

$$U = \{j : I_j = 1\}, \quad C = \{j : I_j = 0\},$$

the sets of uncensored and censored observations, respectively. The likelihood function based on sample (II.1) is then

$$(II.1.1) \quad L(\theta) = \prod_{j \in U} f_X(W_j; \theta_1) \prod_{j \in C} R_X(W_j; \theta_1) \prod_{j \in U} R_T(W_j; \theta_2) \prod_{j \in C} f_T(W_j; \theta_2)$$

which can be written in the form of the product of sub-likelihood functions:

$$L(\theta) = L_1(\theta_1) L_2(\theta_2).$$

We may look on θ_2 as on a nuisance parameter. In practice it is convenient to work with the logarithm of the sublikelihood

$$(II.1.2) \quad \ell_1(\theta_1) = \ln L_1(\theta_1).$$

Assuming the differentiability of (II.1.2), the maximum likelihood equations read

$$(II.1.3) \quad \frac{\partial \ell_1(\theta_1)}{\partial \theta_{1i}} = 0, \quad i = 1, \dots, k.$$

Usually these equations must be solve numerically. Note that the solution to (II.1.3) does not depend on the nuisance distribution F_T . We bring up three useful approaches to the inference.

(i) **Asymptotics of maximum likelihood.** The inference is based on the asymptotic argument for maximum likelihood estimation under general regularity conditions:

$$(II.1.4) \quad \sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} N(0, J^{-1}(\theta)) \quad \text{as } n \rightarrow \infty,$$

where $J(\theta)$ is the *Fisher information matrix* with the elements

$$(II.1.5) \quad J_{ij}(\theta) = -E \frac{\partial^2 \ln h(W, I; \theta)}{\partial \theta_i \partial \theta_j}, \quad i, j = 1, \dots, \dim \theta.$$

If there is no functional dependence between θ_1 and θ_2 then the Fisher information matrix takes the form

$$(II.1.6) \quad J(\theta) = \begin{pmatrix} J_1(\theta_1) & 0 \\ 0 & J_2(\theta_2) \end{pmatrix}$$

and hence the inverse has the same form. Note, however, that $J_1(\theta_1)$ depends on the nuisance distribution. In this case,

$$(II.1.7) \quad \sqrt{n}(\hat{\theta}_1 - \theta_1) \xrightarrow{D} N(0, J_1^{-1}(\theta_1)) \quad \text{as } n \rightarrow \infty.$$

Usually the information matrix is not known. To make the inference we need an estimate for it. There are two natural ways to do so. First, we can calculate the estimate $\hat{\theta}$ and substitute it into $J(\theta)$. Hence the resulting estimate is $J(\hat{\theta})$. Since the explicit formulas for the elements of the Fisher information are often rather complex expressions, this approach usually requires numerical integration. The second method is more optimistic from the computational point of view. It simply estimates the expected value in (II.1.5) by its sample counterpart, the sample mean:

$$(II.1.8) \quad \widehat{J}_{ij}(\theta) = -\frac{1}{n} \sum_{\ell=1}^n \frac{\partial^2 \ln h(W_\ell, I_\ell; \theta)}{\partial \theta_i \partial \theta_j} =_{def} \hat{J}_{ij}, \quad i, j = 1, \dots, \dim \theta.$$

We will denote this matrix as \hat{J} .

(ii) Likelihood ratio test. Denote $\ell(\theta) = \ln L(\theta)$. Let the null hypothesis be $H_0 : \theta_1 = \theta_1^{(0)}$ where $\theta_1^{(0)}$ is a hypothetical value. Suppose that $\hat{\theta}_2^{(0)}$ is the maximum likelihood estimate calculated from the likelihood equations where $\theta_1^{(0)}$ has been substituted instead of θ_1 . Under H_0 the statistic

$$(II.1.9) \quad W_{LR}(\theta_1^{(0)}) = 2[\ell(\hat{\theta}) - \ell(\theta_1^{(0)}, \hat{\theta}_2^{(0)})]$$

has approximately a χ^2 -distribution on $\dim \theta_1$ degrees of freedom. The same argument may also be applied if some of the parameters θ_1 are nuisance. Obviously, in case of no functional dependence between θ_1 and θ_2 , (II.1.9) reduces to

$$(II.1.10) \quad W_{LR}(\theta_1^{(0)}) = 2[\ell_1(\hat{\theta}_1) - \ell_1(\theta_1^{(0)})]$$

(iii) Wald test. Let \widehat{J}^{-1} be the inverse of the matrix with the elements defined in (II.1.8) and let K_1 be the leading ($k \times k$) submatrix of it. Then the Wald statistic

$$(II.1.11) \quad W(\theta_1^{(0)}) = n(\hat{\theta}_1 - \theta_1^{(0)})^T K_1^{-1}(\hat{\theta}_1 - \theta_1^{(0)})$$

has again approximately a χ^2 -distribution on $\dim \theta_1$ degrees of freedom.

Example II.1.1. Exponential distribution. Suppose $\mathcal{L}(X) = \text{Exp}(\theta)$ with the survival function

$$R_X(x) = \begin{cases} \exp(-x/\theta) & x > 0, \\ 1 & x \leq 0. \end{cases}$$

The maximum likelihood estimate of θ is

$$(II.1.12) \quad \hat{\theta} = \frac{\sum_{j=1}^n W_j}{\sum_{j=1}^n I_j}$$

provided that there is at least one uncensored observation. The maximum likelihood estimate of the survival function at a given mission time $x > 0$ is

$$(II.1.13) \quad \hat{R}_X(x) = \exp(-x/\hat{\theta}). \quad \diamond$$

Theorem II.1.1. For (II.1.12) we have

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} N\left(0, \frac{\theta^2}{P(X < T)}\right) \text{ as } n \rightarrow \infty.$$

For (II.1.13) and a fixed $x > 0$ we have

$$\sqrt{n}(\hat{R}_X(x) - R_X(x)) \xrightarrow{D} N\left(0, \left(\frac{x}{\theta}\right)^2 R_X^2(x)/P(X < T)\right) \text{ as } n \rightarrow \infty.$$

Proof. Observe that

$$EW = \int_0^\infty R_{1W}(w)dw = \int_0^\infty R_X(w)R_T(w)dw = \int_0^\infty e^{-w/\theta} R_T(w)dw,$$

$$P(X < T) = \iint_{x < t} dF_X(x)dF_T(t) = \frac{1}{\theta} \int_0^\infty e^{-w/\theta} R_T(w)dw$$

so that

$$\theta = \frac{EW}{P(X < T)}.$$

Hence $E(W - \theta I) = 0$ and $\text{var}(W - \theta I) = E(W - \theta I)^2$. Since

$$EI^2 = P(X < T), \quad EWI = \int_0^\infty \int_{(0,1)} w i h(w, i) d1 dw = \int_0^\infty w f_X(w) R_T(w) dw,$$

($d1$ standing for counting measure)

$$EW^2 = 2 \int_0^\infty w R_{1W}(w)dw = 2\theta \int_0^\infty w f_X(w) R_T(w)dw,$$

we get $\text{var}(W - \theta I) = \theta^2 P(X < T)$. Let \bar{W} and \bar{I} be the means of W_j 's and I_j 's, respectively. Then

$$\sqrt{n}(\hat{\theta} - \theta) = \sqrt{n} \left(\frac{\bar{W}}{\bar{I}} - \frac{EW}{EI} \right) = \sqrt{n} \frac{\bar{W}EI - \bar{I}EW}{\bar{I}EI}.$$

From the central limit theorem for i.i.d. random variables with positive and finite variance it follows that

$$\sqrt{n}(\bar{W}EI - \bar{I}EW) \xrightarrow{D} N(0, \text{var}(WEI - IEW)) \text{ as } n \rightarrow \infty.$$

Since $1/(\bar{I}EI) \xrightarrow{P} 1/(EI)^2 = 1/P^2(X < T)$ we have

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} N\left(0, \frac{1}{(EI)^2} \text{var}(W - \theta I)\right) = N\left(0, \frac{\theta^2}{P(X < T)}\right).$$

To prove the asymptotic normality of \hat{R}_X it is sufficient to employ Theorem 6a.2 (i) (Rao, 1973). \square

A more detailed asymptotic analysis of the estimates may be performed by methods in (Hurt, 1986).

Remark II.1.1. Type I censoring (or time censoring) is a particular case of random censorship. Instead of a random time censor we suppose that the time censor is a fixed constant $T > 0$, say. The results for random censorship can almost literally be transferred to Type I censorship. This is the analogue of the last Theorem:

$$(II.1.14) \quad \sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} N\left(0, \frac{\theta^2}{(1 - e^{-T/\theta})}\right) \quad \text{as } n \rightarrow \infty.$$

Remark II.1.2. In practice we estimate $P(X < T)$ by $\hat{p} = \frac{1}{n} \sum_{j=1}^n I_j$. Hence

$$\sqrt{n\hat{p}}(\hat{\theta} - \theta)/\hat{\theta} \xrightarrow{D} N(0, 1).$$

Theorem II.1.2. Assume that $\mathcal{L}(X_1) = \text{Exp}(\theta)$ and $\mathcal{L}(T_1) = \text{Exp}(\theta/\gamma)$ (the Koziol-Green model). Then

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} N(0, \theta^2(1 + \gamma)) \quad \text{as } n \rightarrow \infty$$

and

$$(II.1.15) \quad \hat{\lambda} = \frac{n-1}{n} \frac{\sum_{j=1}^n I_j}{\sum_{j=1}^n W_j}$$

is the best unbiased estimate of the hazard rate $\lambda = \theta^{-1}$.

Proof. The asymptotic normality is a simple consequence of the previous Theorem. To see that (II.1.15) is unbiased, it suffices to calculate $E I = n/(1 + \gamma)$, $E W^{-1} = (1 + \gamma)/(n - 1)$ and take into account that I, W are independent under Koziol-Green model. Further, $(\sum_{j=1}^n I_j, \sum_{j=1}^n W_j)$ is the complete sufficient statistic for (θ, γ) , hence the result. \square

Example II.1.2. Proportional hazards. Similarly as in (I.5) let us consider a proportional hazards family

$$(II.1.16) \quad r(x; \lambda) = \lambda r_0(x), \quad \lambda > 0$$

where λ is an unknown parameter and $r_0(x)$ is known so called *baseline hazard rate*. The *baseline cumulative hazard rate* is $\Lambda_0(x) = \int_0^x r_0(t)dt$ with the corresponding survival function

$$R(x; \lambda) = \exp\{-\lambda\Lambda_0(x)\}.$$

Hence

$$\Lambda_0(x) = -\frac{1}{\lambda} \ln R(x; \lambda).$$

It is well known that if X has the survival function $R(x)$ then $\mathcal{L}((R(X))) = R(0, 1)$, where $R(0, 1)$ denotes the uniform distribution on $(0, 1)$. Hence

$$\mathcal{L}\left(-\frac{1}{\lambda} \ln R(X; \lambda)\right) = \text{Exp}\left(\frac{1}{\lambda}\right).$$

This is an important fact since instead of processing the original data X_j 's we can treat the transformed data $\Lambda_0(X_j)$'s which follow the exponential distribution with unknown λ . Hence we can apply the above methods. It is also possible to use this fact to separate aspects of the shapes of the marginal distributions of X, Y when constructing a *dependence measure*: instead of $\text{cov}(X, Y)$ use $\text{cov}(\Lambda_X(x), \Lambda_Y(y))$. See (Prentice and Cai, 1992) for details. \diamond

Example II.1.3. Three-parameter Weibull distribution. Suppose $\mathcal{L}(X) = W(\alpha, \theta, \beta)$ with the survival function

$$(II.1.17) \quad R_X(x) = \begin{cases} \exp\left\{-\left(\frac{x-\alpha}{\theta}\right)^\beta\right\} & x > \alpha, \\ 1 & x \leq \alpha. \end{cases} \quad \diamond$$

The following result may be found in (Anděl M., 1992):

Theorem II.1.3. Suppose that at least one observation is uncensored. Then the maximum likelihood estimates $\hat{\alpha}$, $\hat{\theta}$, $\hat{\beta}$ of the parameters in (II.1.17) may be found as a solution to the equations

$$(II.1.18) \quad \hat{\alpha} = \min\{W_j : j \in U\},$$

$$(II.1.19) \quad \frac{1}{\hat{\beta}} + \frac{1}{|U|} \sum_{j \in U} \ln(W_j - \hat{\alpha}) - \frac{\sum_{j \in U} (W_j - \hat{\alpha})^{\hat{\beta}} \ln(W_j - \hat{\alpha}) + \sum_{j \in D} (W_j - \hat{\alpha})^{\hat{\beta}} \ln(W_j - \hat{\alpha})}{\sum_{j \in U} (W_j - \hat{\alpha})^{\hat{\beta}} + \sum_{j \in D} (W_j - \hat{\alpha})^{\hat{\beta}}} = 0,$$

$$(II.1.20) \quad \hat{\theta} = \left[\frac{1}{|U|} \left(\sum_{j \in U} (W_j - \hat{\alpha})^{\hat{\beta}} + \sum_{j \in D} (W_j - \hat{\alpha})^{\hat{\beta}} \right) \right]^{1/\hat{\beta}},$$

where $D = \{j : I_j = 0, W_j \geq \hat{\alpha}\}$, and $|U| = \sum_{j=1}^n I_j$ is the number of uncensored observations.

Proof. See loc. cit. \square

Remark II.1.3. It is important to emphasize that actually only one from the above equations, i. e., (II.1.19), is to be solved numerically.

Remark II.1.4. Note that for the two-parameter Weibull distribution with $\alpha = 0$, (II.1.19) and (II.1.20) are the usual maximum likelihood equations if we put $\hat{\alpha} = 0$. The sums over both U and D are simply sums over all observations.

Example II.1.4. Weibull distribution under Koziol-Green model. Suppose $\mathcal{L}(X) = W(\theta, \beta)$ with the survival function

$$(II.1.21) \quad R_X(x) = \begin{cases} \exp\left\{-\left(\frac{x}{\theta}\right)^\beta\right\} & x > 0, \\ 1 & x \leq 0. \end{cases}$$

If we assume the Koziol-Green model, the distribution of the time censor is also Weibull, $W(\theta\gamma^{-1/\beta}, \beta)$. Altogether we have three unknown parameters. \diamond

Theorem II.1.4. Suppose that at least one observation is uncensored. Then the maximum likelihood estimates $\hat{\theta}$, $\hat{\beta}$ of the parameters in (II.1.21) under the Koziol-Green model may be found as a solution to the equations

$$(II.1.22) \quad \frac{1}{\hat{\beta}} + \frac{1}{n} \sum_{j=1}^n \ln W_j - \frac{\sum_{j=1}^n W_j^{\hat{\beta}} \ln W_j}{\sum_{j=1}^n W_j^{\hat{\beta}}} = 0,$$

$$(II.1.23) \quad \hat{\gamma} = \frac{n - |U|}{|U|},$$

$$(II.1.24) \quad \hat{\theta} = \left[\frac{1 + \hat{\gamma}}{n} \sum_{j=1}^n W_j^{\hat{\theta}} \right]^{1/\hat{\theta}}.$$

In this case, $\hat{\gamma}$ is the maximum likelihood estimate of γ .

Proof. Recall the form of the likelihood from Remark I.1. See (Anděl M., 1992) for further details. \square

The following Theorem gives an explicit form of the Fisher information matrix needed for the asymptotics of the maximum likelihood estimates.

Theorem II.1.5. Denote $a := \Gamma'(2) - \ln(1 + \gamma)$. Under the assumptions of the previous Theorem the Fisher information matrix takes the form

$$(II.1.25) \quad J(\theta, \beta, \gamma) = DAD$$

where

$$(II.1.26) \quad D = \text{diag}\left(\frac{\beta}{\theta}, \frac{1}{\beta}, p\right)$$

and

$$(II.1.27) \quad A = \begin{pmatrix} 1 & -a & -1 \\ -a & a^2 + \pi^2/6 & a \\ -1 & a & 1 + \frac{1}{\gamma} \end{pmatrix}.$$

Proof. The Fisher information matrix may be simply calculated using the facts that

$$(II.1.28) \quad \mathcal{L}\left(\left(\frac{W}{\theta}\right)^\beta\right) = \text{Exp}(p),$$

$\Gamma'(1) = -C$ (the Euler constant), $\Gamma''(1) = \frac{\pi^2}{6} + C^2$. Also the relationships $\Gamma'(2) = 1 - C$, $\Gamma''(2) = \Gamma''(1) + 2\Gamma'(1)$ simplify the calculations. \square

Example II.1.5. Extreme value distribution. We say that a random variable X follows the extreme value distribution type I with parameters $\lambda > 0$ and $\beta > 0$, symbolically $\text{Extrem}(\lambda, \beta)$, if its survival function is

$$(II.1.29) \quad R_X(x) = \exp\{-\lambda e^{\beta x}\}, \quad x \in R.$$

Suppose $\mathcal{L}(Y) = W(\theta, \beta)$ with the survival function (II.1.21). Put $X = \ln Y$, $\lambda = \theta^{-\beta}$. Then $\mathcal{L}(X) = \text{Extrem}(\lambda, \beta)$. Obviously,

$$EX = -\frac{1}{\beta}(\ln \lambda + C), \quad \text{var} X = \frac{\pi^2}{6\beta^2}.$$

Under the Koziol-Green model, $\mathcal{L}(T) = \text{Extrem}(\gamma\lambda, \beta)$. \diamond

Theorem II.1.6. Suppose that at least one observation is uncensored. Then the maximum likelihood estimates $\hat{\beta}$, $\hat{\lambda}$ of the parameters of $\text{Extrem}(\lambda, \beta)$ under the Koziol-Green model may be found as a solution to the equations

$$(II.1.30) \quad \frac{1}{\hat{\beta}} + \frac{1}{n} \sum_{j=1}^n W_j - \frac{\sum_{j=1}^n e^{\hat{\beta} W_j} W_j}{\sum_{j=1}^n e^{\hat{\beta} W_j}} = 0,$$

$$(II.1.31) \quad \hat{\gamma} = \frac{n - |U|}{|U|},$$

$$(II.1.32) \quad \hat{\lambda} = \frac{n}{(1 + \hat{\gamma}) \sum_{j=1}^n e^{\hat{\beta} W_j}}$$

In this case, $\hat{\gamma}$ is the maximum likelihood estimate of γ . Denote $a := \Gamma'(2) - \ln[\lambda(\gamma + 1)]$. The Fisher information matrix takes the form

$$(II.1.33) \quad J(\lambda, \beta, \gamma) = DAD$$

where

$$(II.1.34) \quad D = \text{diag}\left(\frac{1}{\lambda}, \frac{1}{\beta}, p\right)$$

and

$$(II.1.35) \quad A = \begin{pmatrix} 1 & a & 1 \\ a & a^2 + \pi^2/6 & a \\ 1 & a & 1 + \frac{1}{\gamma} \end{pmatrix}.$$

Proof. The proof is analogous to the proof of the previous Theorem. See (Hurt et al., 1991) for details. \square

II.2. Bayesian estimation

Since processes we study in survival analysis are rather evolutionary than revolutionary, a prior knowledge of the previous survival characteristics may be utilized for the estimating of the current ones. There is a lot of papers concerning Bayesian inference for the exponential distribution but little is known for perhaps the most interesting case of the Weibull distribution. At least the results gained are not too satisfactory for practice. The reason sounds naturally. Due to the simple form of the sufficient statistic for the unknown parameter, the inference under exponential distribution often leads to explicit formulæ. For the Weibull distribution, the only sufficient statistic for all two or three parameters is the *whole sample*. Moreover, no natural conjugate family of prior distributions is available³ (Barlow, Proschan 1988). Some suggestions may be found in (Martz, Waller 1982). Most recent papers do not carry too much novelty.

For the sake of simplicity we deal with a scalar parameter only. Suppose that $\{F(\cdot; \theta_1)\}_{\theta_1 \in \Theta}$ is a family of absolutely continuous survival distributions in question. The unknown parameter

³We do not consider the trivial case β known.

θ_1 is supposed to be a random variable with the Lebesgue density $q(t)$. From the Bayes theorem it follows that the a posteriori density of θ_1 is

$$(II.2.1) \quad \psi(\theta_1 | (II.1)) = \frac{L_1(\theta_1)q(\theta_1)}{\int_{\Theta} L_1(t)q(t)dt}.$$

The loss function $l(\theta_1, \tilde{\theta}_1)$ is a function defined on $\Theta \times \tilde{\Theta}$ where $\tilde{\Theta}$ is a decision space, the space of reasonable estimators in our case. The most common is weighted quadratic loss function

$$(II.2.2) \quad l(\theta_1, \tilde{\theta}_1) = w(\theta_1)(\theta_1 - \tilde{\theta}_1)^2$$

where w is properly chosen weight function. It is well-known that under general conditions the Bayes estimator minimizing the Bayes risk

$$(II.2.3) \quad B(\tilde{\theta}_1) = \int_{\Theta} w(t)(t - \tilde{\theta}_1)^2 \psi(t | (II.1)) dt$$

is

$$(II.2.4) \quad \hat{\theta}_1 = \frac{\int_{\Theta} tw(t)\psi(t | (II.1)) dt}{\int_{\Theta} w(t)\psi(t | (II.1)) dt}.$$

Example II.2.1. Exponential distribution. Assume that the survival time is distributed as $\text{Exp}(1/\lambda)$. The natural conjugate prior for the hazard rate λ is the gamma-distribution with the a priori parameters a and p , say, i. e., with the a priori density

$$(II.2.5) \quad q(t) = \frac{a^p}{\Gamma(p)} t^{p-1} e^{-at} \quad t > 0.$$

The a posteriori distribution of λ under random censorship possesses the density

$$(II.2.6) \quad \psi(\lambda | (II.1)) = \text{const} \lambda^{\sum_{j=1}^n I_j + p - 1} e^{-\lambda(\sum_{j=1}^n W_j + a)}$$

which is the gamma density with parameters $\sum_{j=1}^n W_j + a$ and $\sum_{j=1}^n I_j + p$. The Bayes estimate of λ under the quadratic loss function with the constant weight function is

$$(II.2.7) \quad \hat{\lambda} = \frac{\sum_{j=1}^n I_j + p}{\sum_{j=1}^n W_j + a}.$$

The a priori density of $\theta = \frac{1}{\lambda}$ corresponding to (II.2.5) is

$$(II.2.8) \quad q_{\theta}(t) = \frac{1}{\Gamma(p)} a^p e^{-a/t} t^{-p-1}, \quad t > 0$$

which is the density of the inverted gamma distribution with parameters a and p and the expectation $\frac{a}{p-1}$ for $p > 1$. Hence the a posteriori distribution of θ is also the inverted gamma distribution with parameters $\sum_{j=1}^n W_j + a$ and $\sum_{j=1}^n I_j + p$ so that the Bayes estimate of θ is

$$(II.2.9) \quad \hat{\theta} = \frac{\sum_{j=1}^n W_j + a}{\sum_{j=1}^n I_j + p - 1}.$$

Let us turn to the estimation of the survival function $R_X(x; \lambda) = e^{-\lambda x}$ at the fixed mission time x . Without any loss of generality we can set $x := 1$ otherwise we can change the time scale. The a priori density of $R = e^{-\lambda}$ is

$$(II.2.10) \quad q_R(r) = \frac{a^p}{\Gamma(p)} r^{a-1} (-\ln r)^{p-1}, \quad 0 < r < 1$$

so that the a posteriori density of R is

$$(II.2.11) \quad \psi_R(r) = \frac{1}{\Gamma(\sum_{j=1}^n I_j + p)} \left(\sum_{j=1}^n W_j + a \right)^{\sum_{j=1}^n I_j + p} r^{\sum_{j=1}^n W_j + a - 1} (-\ln r)^{\sum_{j=1}^n I_j + p - 1},$$

$$0 < r < 1.$$

Therefore the Bayes estimate of R is

$$(II.2.12) \quad \hat{R} = \left(\frac{\sum_{j=1}^n W_j + a}{\sum_{j=1}^n W_j + a + 1} \right)^{\sum_{j=1}^n I_j + p}.$$

The similar approach may also be used in estimating from the counting processes and for repairable systems characteristics (Franz, 1990). In terms of the reliability parameter r (corresponding to the estimated reliability $R = e^{-\lambda}$), the weighted quadratic loss function becomes

$$(II.1.13) \quad \ell(r, \tilde{r}) = w(r)(r - \tilde{r})^2, \quad 0 \leq r, \tilde{r} \leq 1.$$

By a suitable choice of the weight function $w(r)$ we can put emphasis on the tails of R . One possible choice leading to the explicit form of the estimators is $w(r) = r^{-\alpha} (-\ln r)^{-\beta}$, $\alpha > 0$, $\beta > 0$. It is obvious that w is U-shaped. Higher values of α or β give greater importance to values R close to 0 or 1, respectively. For $\alpha = 0.5$, $\beta = 0.4$ or $\alpha = 0.62$, $\beta = 0.5$ (e.g.), the weight function is almost symmetric. The Bayes estimate of R under the loss function (II.2.13) is

$$(II.2.14) \quad \hat{R}_w = \begin{cases} \left(\frac{\sum_{j=1}^n W_j + a - \alpha}{\sum_{j=1}^n W_j + a - \alpha + 1} \right)^{\sum_{j=1}^n I_j + p - \beta} & \sum_{j=1}^n W_j + a - \alpha > 0 \\ & \sum_{j=1}^n I_j + p - \beta > 0, \\ 1 & \sum_{j=1}^n I_j + p - \beta \leq 0, \\ 0 & \text{otherwise. } \diamond \end{cases}$$

The asymptotic properties of the Bayes estimates are similar to those for the maximum likelihood ones. The development of the asymptotic results is more difficult from the technical point of view, however. Sometime, a general argument concerning the equivalence of the asymptotic behaviour of maximum likelihood and Bayes estimates may be utilized.

III. NONPARAMETRIC INFERENCE

(i) Estimation of the survival function. The most popular estimator of the survival function is Kaplan-Meier or product limit estimator. This estimator can be obtained as a "nonparametric" maximum likelihood estimator (Cox, Oakes 1984). Assume that $W_{(1)}, \dots, W_{(n)}$ are order statistics

of the sample W_1, \dots, W_n , $W_{(0)} = 0$. Let $I_{(j)}$ be the indicator of the event $\{W_{(j)} = X_{(j)}\}$. The Kaplan-Meier estimate of the survival function is defined as

$$(III.1) \quad \hat{R}_{KM}(x) = \begin{cases} \prod_{j: W_{(j)} \leq x} \left(\frac{n-j}{n-j+1} \right)^{I_{(j)}} & x < W_{(n)}, \\ 0 & x \geq W_{(n)}, \end{cases}$$

with empty product defined as 1. In case of complete sample the Kaplan-Meier estimator coincides with the usual empirical distribution function. Little is known about small sample properties. Asymptotics was first given by (Breslow and Crowley, 1974) with an error in the proof. See (Csörgő and Horváth L., 1981) for a correct proof and further results.

Theorem III.1. Let c be such that $R_W(c) > 0$. Then

$$\sqrt{n}(\hat{R}_{KM}(\cdot) - R_X(\cdot)) \xrightarrow{D} V \quad \text{as } n \rightarrow \infty.$$

in Skorohod $D[0, c]$ space where V is a zero mean Gaussian process with the covariance function

$$(III.2) \quad EV(s)V(t) = -R_X(s)R_X(t) \int_0^s \frac{dR_X(x)}{R_X^2(x)R_T(x)}, \quad 0 \leq s \leq t \leq c.$$

Proof. See the references given above. \square

Formula (III.2) has its empirical analogue known as the Greenwood's formula which serves as an approximate estimate of the variance of the Kaplan-Meier estimate:

$$(III.3) \quad \widehat{\text{var}} \hat{R}_{KM}(x) = \hat{R}^2(x) \sum_{j: W_{(j)} \leq x} \frac{I_{(j)}}{(n-j)(n-j+1)}.$$

Example III.1. Making use of this Theorem it can be shown that under the Koziol-Green model the asymptotic variance (as. var) of $\hat{R}_{KM}(x)$ is

$$(III.4) \quad \text{as. var} (\hat{R}_{KM}(x)) = \frac{p}{n} e^{-2x/\theta} (e^{x/\theta} - 1). \quad \diamond$$

(ii) Estimation of the cumulative hazard rate. The cumulative hazard rate may be estimated by the Nelson-Aalen estimator

$$(III.5) \quad \hat{\Lambda}_{KM}(x) = \sum_{j: W_{(j)} \leq x} \frac{I_{(j)}}{n-j+1}.$$

Also this estimator can be obtained as a "nonparametric" maximum likelihood estimator.

Theorem III.2. Let c be such that $R_W(c) > 0$. Then

$$\sqrt{n}(\hat{\Lambda}_{KM}(\cdot) - \Lambda_X(\cdot)) \xrightarrow{D} U \quad \text{as } n \rightarrow \infty.$$

in Skorohod $D[0, c]$ space where U is a zero mean Gaussian process with the covariance function

$$(III.6) \quad EV(s)V(t) = - \int_0^s \frac{dR_X(x)}{R_X^2(x)R_T(x)}, \quad 0 \leq s \leq t \leq c.$$

Proof. See (Andersen et al. 1992). \square

Remark III.1. Recall that $\Lambda_X(x) = -\ln R_X(x)$. There is a close connection between \widehat{R}_{KM} and $\widehat{\Lambda}_{KM}$. Using a Taylor formula we can write

$$(III.7) \quad -\ln \widehat{R}_{KM}(x) = - \sum_{j:W_{(j)} \leq x} \ln \left(\frac{n-j}{n-j+1} \right)^{I_{(j)}} =$$

$$- \sum_{j:W_{(j)} \leq x} \ln \frac{n-j+1-I_{(j)}}{n-j+1} = - \sum_{j:W_{(j)} \leq x} \ln \left(1 - \frac{I_{(j)}}{n-j+1} \right) \doteq$$

$$\sum_{j:W_{(j)} \leq x} \frac{I_{(j)}}{n-j+1}$$

which is the Nelson-Aalen estimator. The left-hand side of (III.7) is called *Peterson estimator* of the cumulative hazard rate. Sometimes

$$(III.8) \quad \widehat{R}_N(x) = \exp^{-\widehat{\Lambda}_{KM}(x)}$$

is called the *Nelson estimator* of the survival function. If we plot both the Kaplan-Meier and Nelson estimate of the survival function we can not observe a substantial difference. The only visible distinction is in the right-hand tail area where (III.1) lies under (III.8).

(iii) **Estimation under the Koziol-Green model.** Remember the relationship $R_X = R_W^p$. Since we dispose of the complete sample of W_j 's, we can estimate R_W by the empirical survival function and p by the relative frequency of uncensored observations and therefore to create the estimator for R :

$$(III.9) \quad \widehat{R}_{KG}(x) = [\widehat{R}_W(x)]^{\widehat{p}}$$

where $\widehat{R}_W(x) = \frac{1}{n} \sum_{j=1}^n I\{W_j > x\}$, $\widehat{p} = \frac{1}{n} \sum_{j=1}^n I_j$. Like the above estimators, (III.9) is the maximum likelihood estimator in the class of nonparametric estimators restricted to the Koziol-Green model.

Theorem III.3. Let c be such that $R_X(c) > 0$ and $p \in (0, 1)$. Then

$$\sqrt{n}(\widehat{R}_{KG}(\cdot) - R_X(\cdot)) \xrightarrow{\mathcal{D}} V \quad \text{as } n \rightarrow \infty.$$

in Skorohod $D[0, c]$ space where V is a zero mean Gaussian process with the covariance function

$$(III.10) \quad EV(s)V(t) = R_X(s)R_X(t) \left[p^2 \frac{F_W(s)}{R_W(s)} + p(1-p) \ln R_W(s) \ln R_W(t) \right]$$

$$0 \leq s \leq t \leq c.$$

Proof. See (Herbst, 1991). \square

The cumulative hazard rate may be estimated in the same way by

$$(III.11) \quad \widehat{\Lambda}_{KG}(x) = \widehat{p} \sum_{j:W_j \leq x} \frac{1}{n-j+1}.$$

Theorem III.4. Let c be such that $R_X(c) > 0$ and $p \in (0, 1)$. Then

$$\sqrt{n}(\hat{\Lambda}_{KG}(\cdot) - \Lambda_X(\cdot)) \xrightarrow{D} Z \quad \text{as } n \rightarrow \infty.$$

in Skorohod $D[0, c]$ space where Z is a zero mean Gaussian process with the covariance function

$$(III.12) \quad E Z(s)Z(t) = p^2 \frac{F_W(s)}{R_W(s)} + p(1-p) \ln R_W(s) \ln R_W(t),$$

$$0 \leq s \leq t \leq c.$$

Proof. See (Anděl M., 1992). \square

Remark III.2. It is not difficult to show that both \hat{R}_{KG} , $\hat{\Lambda}_{KG}$ have greater asymptotic efficiency than \hat{R}_{KM} , $\hat{\Lambda}_{KM}$, respectively. This assertion is based on an interesting inequality

$$\frac{r}{1-r} \ln^2 r < 1, \quad \forall r \in (0, 1).$$

(iv) **Estimation of moments.** The r^{th} -moment about the origin of a nonnegative random variable with the survival function R_X may be calculated utilizing the well-known formula

$$(III.13) \quad \mu'_r = r \int_0^\infty x^{r-1} R_X(x) dx.$$

A natural empirical analogue of (II.1) is obtained by substituting the Kaplan-Meier estimator instead of R_X :

$$(III.14) \quad \hat{\mu}'_{r,KM} = r \int_0^\infty x^{r-1} \hat{R}_{KM}(x) dx = \sum_{j=1}^n \hat{R}_{KM}(W_{(i-1)}) [W_{(i)} - W_{(i-1)}].$$

Despite the asymptotic normality of the last estimate can also be established, it is beyond the scope of the present paper to formulate precise results.

(v) **Estimation of moments under the Koziol-Green model.** The same idea leads to the estimate

$$(III.15) \quad \hat{\mu}'_{r,KG} = \sum_{j=1}^n \hat{R}_{KG}(W_{(i-1)}) [W_{(i)} - W_{(i-1)}] = \sum_{j=1}^n \left[1 - \frac{i-1}{n}\right] \hat{p} [W_{(i)} - W_{(i-1)}].$$

Also in this case a detailed analysis of the asymptotics is rather complicated, see (Herbst, 1992b, 1992c).

(vi) **Test of fit with the Koziol-Green model.** The test of a Kolmogorov-Smirnov type has been developed by (Herbst 1992a). It is based on characterizations in Theorems I.2 and I.3. Define the empirical conditional distribution functions

$$(III.16) \quad \hat{F}_1(x) = \frac{1}{|U|} \sum_{j=1}^n I\{W_j \leq x, I_j = 1\}$$

and

$$(III.17) \quad \hat{F}_0(x) = \frac{1}{n - |U|} \sum_{j=1}^n I\{W_j \leq x, I_j = 0\}.$$

Denote

$$(III.18) \quad \Delta_n = \sup_x |\hat{F}_1(x) - \hat{F}_0(x)|$$

the Kolmogorov-Smirnov statistic for the conditional empirical distribution functions \hat{F}_1 and \hat{F}_0 .

Theorem III.5. *If there exists a positive γ such that (I.4) is true then for $x > 0$*

$$(III.19) \quad P \left(\sqrt{\frac{|U|(n-|U|)}{n}} \Delta_n \leq x \right) \longrightarrow K(x) \quad \text{as } n \longrightarrow \infty$$

where $K(x) = \sum_{j=-\infty}^{\infty} (-1)^j \exp(-2j^2 x^2)$ is the Kolmogorov distribution function.

Proof. For the proof see (Herbst 1992a). \square

In loc. cit. the following assertion concerning the exact distribution of the test statistic can be found.

Theorem VI.2. *If there exists a positive γ such that (I.4) is true then the conditional distribution $\mathcal{L}(\Delta_n | |U| = k)$ is the same as the distribution of the usual Kolmogorov-Smirnov statistic in case of two independent samples of size k and $n - k$, respectively.*

(vii) **Nonparametric estimation of the density and hazard rate.** Successfully the kernel-type estimators are used. Suppose K be a positive symmetric kernel, $\int K(t)dt = 1$. For the density f_X , a kernel-type estimator is

$$(III.20) \quad \hat{f}_X(x) = -h_n^{-1} \int_{-\infty}^{\infty} K[(x-t)/h_n] d\hat{R}_{KM}(t),$$

where \hat{R}_{KM} is the Kaplan-Meier estimator. The integral in the last formula may be reduced to the finite sum in this case. For the hazard rate r_X , a kernel-type estimator is

$$(III.21) \quad \hat{r}_X(x) = -h_n^{-1} \int_{-\infty}^{\infty} K[(x-t)/h_n] \hat{R}_{KM}^{-1}(t) d\hat{R}_{KM}(t), \quad dx > 0, R_X(x) > 0.$$

See (Padgett, 1988) and (Tanner and Wong, 1983). For a review of the ideas for complete samples see (Antoch, 1986).

IV. PACKAGES

This is only to make some remarks on four packages which were not selected at random at all. Rather they reflect the situation with available packages in this country.

IV.1. SOLO

There is a part of the package called *Survival analysis*. As a rule, survival data should enter as two vectors – the first one containing W_j 's and the second one containing I_j 's. The estimation module supplies the user with estimates of parameters for exponential, Weibull, Rayleigh, Gamma, normal, and lognormal distribution as well as the Kaplan-Meier estimate with the cumulative hazard function. Beware of the parameters! They are often defined in a different way than we are accustomed to.

The modules *Two-Group Nonparametric Tests* and *Multi-Group Nonparametric Tests* contain analogues to the Wilcoxon and Kruskal-Wallis tests, respectively, for comparing survival distributions. Since neither the tests nor a motivation for them are clearly described in the User's Manuals, we will give a heuristic derivation for the two-sample problem here. The ideas involved are common for all the tests that can be found in the packages mentioned in this paper.

Suppose that we have two samples 1 and 2 of (I.1) type with the sizes n_1 and n_2 , respectively:

$$(IV.1.1) \quad (W_1^i, I_1^i), \dots, (W_{n_1}^i, I_{n_1}^i) \quad i = 1, 2.$$

If F_1 and F_2 are the distribution functions of the samples 1 and 2, respectively, the null hypothesis is

$$H_0 : F_1 = F_2.$$

Let $Z_1 < \dots < Z_L$ be the ordered (actually) observed distinct survival times from the sample formed by pooling the samples 1 and 2 together, and let $D_{i\ell}$ and $Y_{i\ell}$, denote the number of observed survival times and the number of individuals still at risk, respectively, in sample i at time Z_ℓ , $\ell = 1, \dots, L$, $i = 1, 2$, formally

$$(IV.1.2) \quad D_{i\ell} = \sum_{j=1}^{n_i} I\{W_j^i = X_j^i, W_j^i \leq Z_\ell\}, \quad Y_{i\ell} = \sum_{j=1}^{n_i} I\{W_j^i \geq Z_\ell\}.$$

For every ℓ we may create a 2×2 contingency table:

Survival	Sample 1	Sample 2	Total
Yes	$D_{1\ell}$	$D_{2\ell}$	D_ℓ
No	$Y_{1\ell} - D_{1\ell}$	$Y_{2\ell} - D_{2\ell}$	$Y_\ell - D_\ell$
Total	$Y_{1\ell}$	$Y_{2\ell}$	Y_ℓ

As usually, "Survival Yes" means that we have observed a survival time, or better to say, death. Given $Y_{i\ell}$, the $D_{i\ell}$ have a binomial distribution with number of trials $Y_{i\ell}$. The Fisher exact test for testing the hypothesis of the equality of binomial parameters in this setting is based on conditioning on D_ℓ and on using the resulting conditional hypergeometric distribution for $D_{1\ell}$. Under the hypothesis, $D_{1\ell}$ has the conditional expectation

$$E_{1\ell} = D_\ell \frac{Y_{1\ell}}{Y_\ell}$$

and the conditional variance

$$V_{1\ell} = D_\ell \frac{Y_{1\ell} Y_{2\ell}}{Y_\ell^2} \frac{Y_\ell - D_\ell}{Y_\ell - 1}.$$

The test statistic

$$(IV.1.3) \quad Q = \frac{\sum_{\ell=1}^L (D_{1\ell} - E_{1\ell})}{\sqrt{\sum_{\ell=1}^L V_{1\ell}}}$$

has approximately $N(0, 1)$ under the null hypothesis. The statistic (IV.1.3) is the standardized two-sample logrank statistic. For the two-sample problem, the tests in the packages are based on statistics of the similar form

$$(IV.1.4) \quad W = c \sum_{\ell=1}^L w(Z_\ell)(D_{1\ell} - E_{1\ell})$$

where c is a normalizing constant and $w(\cdot)$ is a properly chosen weight function which also gives the name to the test. The reader will find a variety of combinations of the following names: *Peto J.*, *Peto R.*, *Wilcoxon*, *Gehan*, *Cox*, *Mantel*, *Haenszel*, *Tarone*, *Ware*, *Breslow*. Tests differ in their power with respect to the alternatives. For a lucid explanation see (Miller, 1981). The multi-group nonparametric tests are based on similar, but inevitably nonlinear statistics which lead to an approximate chi-squared distribution.

In the *Cox's Proportional Hazards Regression* module it is supposed that the hazard rate is of the form

$$(IV.1.5) \quad \lambda(x; \mathbf{x}) = \lambda_0(x) \exp(\mathbf{x}'\boldsymbol{\beta})$$

where \mathbf{x} is a vector of covariates, $\boldsymbol{\beta}$ is a vector of unknown parameters, and λ_0 is a baseline arbitrary hazard rate. Only ranks of the survival times are used to estimate $\boldsymbol{\beta}$.

A simple, but powerful method is given in *Censored Regression* module, the idea of which comes from (Schmee, Hahn 1979). In the usual regression model

$$Y = X\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

with Y possibly censored, the censored observations are treated as survival times in the first step so that we begin with n regression equations

$$(IV.1.6) \quad W_j = \mathbf{x}'_j\boldsymbol{\beta} + \varepsilon_j, \quad j = 1, \dots, n.$$

Denote $X = (\mathbf{x}_1, \dots, \mathbf{x}_n)'$. The initial estimate of $\boldsymbol{\beta}$ is the usual least squares estimate

$$\hat{\boldsymbol{\beta}} = (X'X)^{-1}X'W$$

where $W = (W_1, \dots, W_n)'$. Based upon this initial fit, the expected survival time for every censored observation is estimated by

$$(IV.1.7) \quad \widehat{W}_j = \mathbf{x}'_j\hat{\boldsymbol{\beta}}, \quad j \in C.$$

Let $\hat{\sigma}$ is an estimate of $\text{var } \widehat{W}_j = \mathbf{x}'_j \text{cov}(\hat{\boldsymbol{\beta}}) \mathbf{x}_j$. Considering the j^{th} censored observation as truncated at W_j ,

$$(IV.1.8) \quad \widetilde{W}_j = \widehat{W}_j + \hat{\sigma} \varphi \left(\frac{W_j - \widehat{W}_j}{\hat{\sigma}} \right) / \left[1 - \Phi \left(\frac{W_j - \widehat{W}_j}{\hat{\sigma}} \right) \right], \quad j \in C$$

serves as an estimate of the expected survival time based on the truncated normal distribution. These estimates are introduced in the regression model instead of the original time censors, and the usual regression procedure is repeated. This procedure is iterated until convergence is achieved.

The last module deals with *Probit Analysis*.

Résumé: There is a variety of fundamental methods of survival analysis in this package, the convergence of methods which need iterations is reliable enough, and with a basic knowledge of the *SOLO* system handling the survival analysis modules is almost effortless. Despite the '*SOLO* Survival Analysis' Manual is not a product of professional statisticians, the author recommends *SOLO* as a very comfortable tool in this field.

IV.2. BMDP

Since the methods used in this package are briefly but rigorously described in the Appendix to the Manual, we only make some comments. A big part of the corresponding programme 1L is devoted to nonparametric analysis based on *life table computations* and on *product-limit computations*. Tests for comparing the distributions allow data to be stratified and organized in ordered treatment groups. Further methods included in the package are Cox proportional hazards model and accelerated failure time model which will be discussed in the next Section.

IV.3. SAS

Algorithms for statistical analysis of survival data are contained in SAS/STAT LIFEREG Procedure. Data may be right-, left-, or interval-censored. Everything is based on an accelerated failure time model in which it is assumed that the survival time X can be expressed in the form

$$(IV.3.1) \quad X = \exp(\mathbf{x}'\boldsymbol{\beta})X_0$$

where \mathbf{x} is the vector of covariates, $\boldsymbol{\beta}$ a vector of unknown parameters, and X_0 is the survival time sampled from the baseline distribution with zero covariates. The baseline distribution is supposed to be of the known form with unknown parameters. This is important since by setting all covariates equal to zero, we may obtain estimates of the parameters of the baseline distribution. Optional logarithmic transformation $Y = \ln X$ gives the linear model

$$(IV.3.2) \quad Y = \mathbf{x}'\boldsymbol{\beta} + Y_0$$

where $Y_0 = \ln X_0$ plays the role of the error term. The user can choose extreme value, normal, and/or logistic baseline distribution for Y_0 in (IV.3.2). After the logarithmic transform, he can choose among exponential, Weibull, log-normal, and log-logistic distributions for X_0 in (IV.3.1). The parameters are estimated by the maximum likelihood method applied to (IV.3.2) described above. As common with SAS, it offers a lot of options for input, output, and computations.

IV.4. SPSS

Only life table methods are in SPSS. The results are summarized in tables containing the empirical densities, hazard rates, and survival functions. Also a comparison of survival functions is possible.

V. CONCLUDING REMARKS

This is a non-representative selection of some trends in the field, which are, by the author's opinion, prospective for further research and applications.

(i) **Product-integration.** A unifying theory of product-integration with survival analysis applications has been given by (Gill, Johansen 1990). The theory is based on the product-integral which generalized the discrete product operator Π in a similar way as the usual integral \int generalizes the operator \sum . Since the Kaplan-Meier estimator may be expressed in terms of product-integral, the properties of it can be derived by making use of the general results valid for the product-integral.

(ii) **Counting processes and the martingale approach.** "The martingale approach has proved remarkably successful in yielding results about statistical methods for many problems arising in censored data. Martingale methods can be used to obtain simple expressions for the moments of complicated statistics, to calculate and verify asymptotic distributions for test statistics and estimators, to examine the operating characteristics of nonparametric testing methods and semiparametric censored data regression methods, and even provide a basis for graphical diagnostics in model building with counting process data." (Fleming and Harrington, 1991)

Here we briefly sketch the counting process approach to the cumulative hazard rate. The first step is to construct an integral representation for statistics calculated from censored data. For $j = 1, \dots, n$ define

$$(V.1) \quad N_j(x) = I\{W_j \leq x, I_j = 1\}, \quad Y_j(x) = I\{W_j \geq x\},$$

$$N = \sum_{j=1}^n N_j, \quad Y = \sum_{j=1}^n Y_j.$$

Using this notation, the Nelson-Aalen estimator of Λ (here as well as in the sequel we omit index x) may be expressed as

$$(V.2) \quad \hat{\Lambda}(x) = \int_0^x \frac{I\{Y(t) > 0\}}{Y(t)} dN(t)$$

($0/0 = 0$, by definition). (V.3) "estimates" the random quantity

$$(V.3) \quad \Lambda^*(x) = \int_0^x I\{Y(t) > 0\} r(t) dt.$$

We have

$$(V.4) \quad \hat{\Lambda}(x) - \Lambda(x) = \int_0^x \frac{I\{Y(t) > 0\}}{Y(t)} \{dN(t) - Y(t)r(t)dt\} = \sum_{j=1}^n \int_0^x \frac{I\{Y(t) > 0\}}{Y(t)} dM_j(t),$$

where

$$(V.5) \quad M_j(x) = N_j(x) - \int_0^x Y_j(t) d\Lambda(t) = N_j(x) - A_j(x),$$

say. The process M_j is a martingale, and A_j is called compensator. The martingale approach to statistical models for counting processes is useful only if the compensator is known or can be computed (at least theoretically). Denote $\Delta\Lambda(x) = \Lambda(x) - \Lambda(x-)$. The following assertions may be found in (Fleming and Harrington, 1991):

$$(V.6) \quad E\hat{\Lambda}(x) = E\Lambda^*(x),$$

$$(V.7) \quad E\hat{\Lambda}(x) - E\Lambda(x) = - \int_0^x P^n(W < t) d\Lambda(t) \geq -P^n(W < x)\Lambda(x),$$

$$(V.8) \quad \sigma_*^2(x) = E\{\sqrt{n}\{\hat{\Lambda}(x) - \Lambda(x)\}\}^2 = E\left[n \int_0^x \frac{I\{Y(t) > 0\}}{Y(t)} \{1 - \Delta\Lambda(t)\} d\Lambda(t)\right].$$

The last quantity should approach

$$(V.9) \quad \sigma^2(x) = \int_0^x P^{-1}(W \geq t) [1 - \Delta\Lambda(t)] d\Lambda(t)$$

for large n . Further, $E n[\Lambda^*(x) - \Lambda(x)]^2 \rightarrow 0$ as $n \rightarrow \infty$ if $P(W \geq x) > 0$, and

$$(V.10) \quad \sqrt{n}[\hat{\Lambda}(x) - \Lambda(x)] = \frac{1}{\sqrt{n}} \sum_{j=1}^n \int_0^x \frac{n}{Y(t)} dM_j(t)$$

where M_j 's are independent identically distributed. Since $n/Y(x) \rightarrow 1/P(W \geq x)$, we might expect that (V.10) is approximately distributed as $N(0, \sigma^2(x))$ for large n which is indeed the case.

(iii) **Left truncated right censored data.** In this context, left truncation is considered in a slightly different way than usually. Left truncation by t_j means that X_j is only observable if $X_j \geq t_j$. The data consist of n observations

$$(X_1^0, t_1^0), \dots, (X_n^0, t_n^0) \quad \text{with} \quad X_j^0 \geq t_j^0.$$

We can regard the observed sample as being generating by a larger sample of independent random variables

$$(X_j, t_j), \quad j = 1, \dots, m(n)$$

where $M(n) = \inf\{m : \sum_{j=1}^m I\{X_j^0 \geq t_j^0\} = n\}$. In (Lynden-Bell, 1971), an analogue to the Kaplan-Meier estimator has been derived. Such a kind of experiment has applications in astronomy when small objects are not observed. The properties similar to those of the Kaplan-Meier estimator have been studied by many authors. Afterwards, the right censoring of the data was also considered and the research proceeds in the same way as for the usual Kaplan-Meier estimator.

REFERENCES

1. Antoch J., *Estimation of Density*, Proceedings of ROBUST'86 Winter School (J. Antoch and J. Jurečková, ed.), JČSMF, Praha, 1986, pp. 1-9. (Czech)
2. Anděl M., *Properties of the estimates for Weibull Distribution*, Diploma Thesis (1992), Charles University, Faculty of Mathematics and Physics, Prague. (Czech)
3. Andersen P. K., Borgan Ø., Gill R. D., Keiding N. L., *Statistical Models Based on Counting Processes*, Springer, New York 1992, (to appear).
4. Antoch J., *Programme Equipment for Statistical Data Analysis*, Computer Echo III no. 2 (1992), 40-43. (Czech)
5. Barlow R. E., Proschan F., *Life Distribution Models and Incomplete Data*, Handbook of Statistics (P. R. Krishnaiah and C. R. Rao, ed.), Elsevier, Amsterdam, 1988, pp. 225-249.
6. Chen Y. Y., Hollander M., Langberg N. A., *Small Sample Results for the Kaplan-Meier Estimator*, J. Am. Stat. Assoc. 77 (1982), 141-144.
7. Cox D. R., Oakes D., *Analysis of Survival Data*, Chapman and Hall, London, 1984.
8. Csörgő S., Horváth L., *On the Koziol-Green for Random Censorship*, Biometrika 68 (1981), 391-401.
9. Fleming T. R., Harrington D. P., *Counting Processes and Survival analysis*, Wiley, New York, 1991.
10. Franz J., *Admissible Estimators in Sequential Plans for Exponential-type Processes*, Scand. J. Statist. 17 (1990), 275-285.
11. Gill R. D., Johansen S., *A Survey of Product-Integration with a View toward Application in Survival Analysis*, Annals of Statistics 18 no. 4 (199), 1501-1555.
12. Herbst T., *Estimation of Lifetime Characteristics from Randomly Censored Samples*, PhD Thesis (1991), Charles University, Faculty of Mathematics and Physics, Prague.
13. Herbst T., *Test of Fit with the Koziol-Green for Random Censorship*, Statistics & Decisions 10 (1992a), 163-171.
14. Herbst T., *Estimation of Moments under Koziol-Green Model for Random Censorship*, Commun. Statist.-Theory Meth. 21 no. 3 (1992b), 613-624.
15. Herbst T., *On Estimation of Residual Moments under Koziol-Green Model of Random Censorship*, Commun. Statist.-Theory Meth. (1992c) (to appear).
16. Hjort N. L., *Semiparametric Estimation of Parametric Hazard Rates*, Statistical Research Report no. 6 (1991), Department of Mathematics, University of Oslo.
17. Hurt J., *Theory of Reliability*, Státní pedagogické nakladatelství, Praha, 1984. (Czech)
18. Hurt J., *Comparison of Some Reliability Estimators in the Exponential Case under Random Censorship*, Proc. of the 5th Pannonian Symp. on Math. Stat. (W. Grossman, J. Mogyoródi, I. Vincze, W. Wertz, ed.), Visegrád, Hungary, 1985, pp. 255-266.
19. Hurt J., *Bayes Methods in Reliability Analysis*, Statistik und Informatik 20 no. 1-2 (1990), 65-66.
20. Hurt J., Dahms R., Samek P., Veselý P., *On Inference for Extreme Value Distribution under Random Censorship*, (Unpublished manuscript, to appear).
21. Kaplan E. L., Meier P., *Nonparametric Estimator from Incomplete Observations*, J. Amer. Statist. Ass. 53 (1958), 457-481.

22. Koziol J. A., Green S. B., *A Cramér-von Mises Statistic for Randomly Censored Data*, *Biometrika* 63 (1976), 465-474.
23. Lai T. L., Ying Z., *Estimating a Distribution Function with Truncated and Censored Data*, *The Annals of Statistics* 19 no. 1 (1991), 417-442.
24. Lawless J. F., *Statistical Models and Methods for Lifetime Data*, Wiley, New York, 1982.
25. Lee E., *Statistical Methods for Survival Data Analysis*, Wadsworth, Belmont (CA), 1984.
26. Lynden-Bell D., *A Method of Allowing for Known Observational Selection in Small Samples Applied to 3CR Quasars*, *Monthly Notices Roy. Astronomy Soc.* 155 (1971), 95-118.
27. Martz H. F., Waller R. A., *Bayesian Reliability Analysis*, Wiley, New York, 1982.
28. Miller R.G. jr., *Survival Analysis*, Wiley, New York, 1981.
29. Padgett W. J., *Nonparametric Estimation of Density and Hazard Rate Functions when Samples are Censored*, *Handbook of Statistics* (P. R. Krishnaiah[†] and C. R. Rao, ed.), Elsevier, Amsterdam, 1988, pp. 313-331.
30. Prentice R. L., Cai J., *Covariance and Survivor Function Estimation Using Multivariate Failure Time Data*, *Biometrika* 79 no. 3 (1992), 495-512.
31. Rao C. Radhakrishna, *Linear Statistical Inference and Its Applications*, Second edition, Wiley, New York, 1973.
32. Schmee J., Hahn G. J., *A Simple Method for Regression Analysis with Censored Data*, *Technometrics* 21 no. 4 (1979), 417-432.
33. Shorack G. R., Wellner, J. A., *Empirical Processes with Applications to Statistics*, Wiley, New York, 1986.
34. Tanner M. A., Wong W. H., *The Estimation of the Hazard Function from Randomly Censored Data by the Kernel Method*, *The Annals of Statistics* 11 (1983), 989-993.

DEPARTMENT OF STATISTICS, CHARLES UNIVERSITY, SOKOLOVSKÁ 83, 186 00 PRAHA 8, CZECHOSLOVAKIA
 E-mail: HURT@CSPGUK11.BITNET