On Bayes Assessing of Goodness-of-fit in Lifetime Models

P. Volf, J. Timková
Institute of Information Theory and Automation, Prague, Czech Republic
Email: volf@utia.cas.cz

ABSTRACT: The contribution deals with statistical tests of model fit. We recall the notion of martingale residuals defined for lifetime models based on failure intensities. Analysis of these residual processes has already been studied by many authors. Nevertheless, Bayes approach to this problem is just developing. We shall present a Bayes procedure of estimation in Cox's and AFT models semi-parametric models. Our main concern is Bayes construction of residual processes and goodness-of-fit tests based on them. The method will be illustrated on an example with randomly generated data.

1 MARTINGALE RESIDUALS

In order to introduce the notion of martingale residuals, we shall first consider a standard survival data case, without any dependence on covariates. Let us imagine that a set of i.i.d. random variables T_i, survival times of n objects of the same type, is observed. Alternatively, we may consider their **counting processes** $N_i(t)$, each having maximally 1 count (at the time of failure, T_i), or it can be censored without failure. Further, let us consider also indicator processes (of being in risk) $Y_i(t)$, $Y_i(t) = 0$ after failure or censoring, $Y_i(t) = 1$ otherwise. As lifetimes are i.i.d., corresponding counting processes have the same common **hazard rate** $h(t) \geq 0$. Cumulated hazard rate is then $H(t) = \int_0^t h(s)ds$. It follows that the intensity of $N_i(t)$ is $a_i(t) = h(t) \cdot Y_i(t)$. Notice a difference between those two notions: The hazard rate is a characteristics of distribution, namely here $h(t) = -dln\bar{F}(t)/dt$ where $\bar{F}(t) = 1 - F(t)$ is a survival function, complement to distribution function, while intensity depends on realization of process $Y_i(t)$. It is assumed that data are observed on a finite time interval $t \in [0, T]$, $N_i(0) = 0$.

Let us define also sums of individual characteristics, namely counting process $N(t) = \sum_{i=1}^n N_i(t)$ counting number of failures, further $Y(t) = \sum_{i=1}^n Y_i(t)$, cumulated intensities $A_i(t) = \int_0^t a_i(s)ds$ and $A(t) = \sum_{i=1}^n A_i(t)$, so that here $A(t) = \int_0^t h(s)Y(s)ds$.

In theoretical studies on lifetime models, many results are based on martingale – compensator decomposition of counting process, namely that $N_i(t) = A_i(t)+M_i(t)$, so that also $N(t) = A(t) + M(t)$, where $M_i(t)$, $M(t)$ are martingales with zero means, conditional variance processes (conditioned by corresponding filtration, a nondecreasing set of σ-algebras $\mathcal{F}(t^-)$) are $\langle M_i\rangle(t) = A_i(t)$, $\langle M\rangle(t) = A(t)$. Naturally, martingales have non-correlated increments, and $M_i(t)$ are also non-correlated mutually (for different i).

Then it is quite reasonable to consider a residual process (martingale residuals)

$$R(t) = N(t) - \hat{A}(t) = M(t) + A(t) - \hat{A}(t)$$

as a tool for testing model fit. Here $\hat{A}(t)$ is estimated cumulated intensity. Hence, residual process is constructed from observed data, and its properties depend mainly on properties of estimator of cumulated hazard rate, because $\hat{A}(t) = \int_0^t Y(s)d\hat{H}(s)$. Tests are then performed either graphically or numerically, critical borders for assessing the goodness-of-fit are based on asymptotic properties of estimates.

1.1 *Properties of residuals*

The most common estimator of cumulated hazard rate $H(t)$ is the Nelson-Aalen estimator, which has the form:

$$\hat{H}(t) = \int_0^t \sum_{i=1}^n \frac{dN_i(s)}{\sum_{j=1}^n Y_j(s)} = \int_0^t \frac{dN(s)}{Y(s)},$$

so that it is a piecewise constant function with jumps $d\hat{H}(s) = dN(s)/Y(s)$ at times where failures have occurred. Its asymptotic properties, namely uniform on $[0,T]$ consistency in probability and asymptotic normality when $n \to \infty$, are well known (for review of survival analysis, see for instance Kalbfleisch and Prentice, 2002). More precisely, the following convergence in distribution on $[0,T]$ to Brown process $\mathcal{B}$ holds:

$$\sqrt{n}(\hat{H}(t) - H(t)) \to \mathcal{B}(V(t)),$$

$$V(t) = \int_0^t \frac{h(s)ds}{c_0(s)},$$

where we assume the existence of $c_0(s) = P - lim \frac{Y(s)}{n}$, uniform in $[0,T]$, $c_0(s) \geq \varepsilon > 0$. Hence, it is possible to construct Kolmogorov-Smirnov type confidence bands for $H(t)$ as well as point-wise confidence intervals. Consistent (again uniform in $0,T$) estimator of $V(t)$ is available, too: $\hat{V}(t) = \int_0^t \frac{n\,dN(s)}{Y(s)^2}$.

In the present contribution we are interested mainly in properties of residual process $R(t) = N(t) - \hat{A}(t)$. Notice that here $\hat{A}(t) = N(t)$ directly, so that it is preferred to construct residuals in data subgroups (strata), $S \subset \{1,..,n\}$. Thus, let us define

$$R_S(t) = N_S(t) - \hat{A}_S(t) = M_S(t) + A_S(t) - \hat{A}_S(t),$$

where we denoted again $N(t) = \sum_{i=1}^n N_i(t)$, $N_S(t) = \sum_{i\in S} N_i(t)$, similarly for $Y(t)$, $M(t)$, $A(t)$, $\hat{A}(t)$. As $\hat{A}_S(t) =$

$$= \int_0^t \sum_{i\in S} d\hat{H}(r) Y_i(r) = \int_0^t \frac{dN(r)}{Y(r)} \cdot Y_S(r) =$$

$$= \int_0^t \frac{dH(r)Y(r) + dM(r)}{Y(r)} \cdot Y_S(r) =$$

$$= A_S(t) + \int_0^t \frac{dM(r)}{Y(r)} \cdot Y_S(r),$$

we obtain that (with notation $\bar{S}$ - complement of S)

$$R_S(t) = M_S(t) - \int_0^t \frac{dM(r)}{Y(r)} \cdot Y_S(r) =$$

$$= \int_0^t \frac{dM_S(r)Y_{\bar{S}}(r) - dM_{\bar{S}}(r)Y_S(r)}{Y(r)}.$$

From its structure it follows that process $R_S(t)$ has non-correlated increments, conditioned variance (by σ-algebras $\mathcal{F}(t^-)$) of $\frac{1}{\sqrt{n}} dR_S(t)$ is

$$\frac{dH(t)}{nY(t)^2}(Y_{\bar{S}}(t)Y_S(t)^2 + Y_{\bar{S}}(t)^2 Y_S(t))$$

$$\sim dH(t)\frac{c_S(t)c_{\bar{S}}(t)}{c_0(t)},$$

where we again assume that there exist P-limits $Y_S(t)/n \to c_S(t), Y_{\bar{S}}(t)/n \to c_{\bar{S}}(t)$, $Y(t)/n \to c_0(t)$, uniform in $t \in [0,T]$, bounded away from zero. Then $\frac{1}{\sqrt{n}} R_S(t) \to \mathcal{B}(V_R(t))$ converges to Brown process, too, and asymptotic variance function $V_R(t)$ is consistently estimable by $\hat{V}_R(t) =$

$$\int_0^t \frac{d\hat{H}(r)Y_S(r)Y_{\bar{S}}(r)}{nY(r)} = \int_0^t \frac{dN(r)Y_S(r)Y_{\bar{S}}(r)}{nY(r)^2}.$$

Hence, if assumptions of our model hold, the process

$$\frac{1}{\sqrt{n}} \frac{R_S(t)}{(1 + \hat{V}_R(t))}$$

should behave asymptotically as Brown bridge process. It can be tested by the Kolmogorov-Smirnov criterion (or other similar, as Cramer-von Mises test). As we assume a simple model of survival times without any non-heterogeneity, the method can be used for assessing the homogeneity for different subsamples S.

The case considered in the present part was rather simple, in such a case the tests of model fit can be performed directly with the aid of estimated cumulative hazard rates or distribution functions (recall well known Product limit estimate of Kaplan and Maier). However, in cases of regression models the test are not so straightforward. That is why we shall continue by description of Bayes variant of residual analysis.

2 BAYES RESIDUALS

Bayes approach to statistical analysis assumes that all models components (i.e. the parameters as well as non-parametrized parts) as random quantities, initially with a prior probability distribution. The result of statistical analysis is then a posterior distribution of those model components, i.e. their estimate is a distribution. Actually it is the likelihood function 'modulated' by prior distribution.

From another point of view, it is possible to say that while the "standard statistics" studies the variation of data and its consequence when inserted to given functions (estimators), in Bayes statistics the main concern is variation of 'parameters', data are taken as fixed.

Today, Bayes analysis is often supported by the MCMC (Markov Chain Monte Carlo) methods. They are based on algorithms of random sampling (Gibbs sampler, Metropolis-Hastings procedure) and are used for obtaining approximate representation of posterior distribution. More about MCMC can be found elsewhere, for instance in Gamerman (1997).

In the case considered here we deal with nonparametric hazard rate. For Bayes solution, its representation can be made from piecewise constant functions (or from splines or from other functional basis), as in Arjas and Gasbarra (1994). Parameters are then points of changes of hazard rate, also their number in $[0, T]$, and levels of hazard rate in intervals between these points. Arjas and Gasbarra (1994) show how MCMC generation can follow Gibbs sampler combined with an "accept-reject" sampling method.

Once we have posterior sample of 'hazard rates', $h^{(m)}(t)$, (i.e. last M representatives of aposteriori distribution obtained by MCMC procedure), we can construct from them a sample of cumulated intensities in subgroup S and corresponding residuals:

$$A_S^{(m)}(t) = \int_0^t h^{(m)}(r) Y_S(r) dr,$$

$$R_S^{(m)}(t) = N_S(t) - A_S^{(m)}(t).$$

2.1 *Bayes confidence regions*

Point-wise (at each t) sample quantiles of $R_S^{(m)}(t)$ are obtained immediately, showing so called credibility intervals (Bayesian version of confidence intervals) for $R_S(t)$. Methods for construction of confidence bands (of Bayes type) on the whole interval $[0, T]$ are studied intensively nowadays. Theoretically, this problem is connected with the concept of 'depth of data' (see for instance Zuo and Serfling, 2000). Practically, the approach corresponds to construction of multivariate quantiles, for instance in the following way: Let us consider a sample of functions $f^{(m)}(x)$, $m = 1, .., M$, given empirically by values at the same set of points $x_j, j = 1, .., J$. For each $k < M/2$ point-wise sample k/M and $(M-k)/M$ quantiles (i.e. at each x_j) can be constructed. If we join them to bands, we can try to find such k that, approximately, a given proportion (95%, say) of functions lies inside. As an additional finer criterion we can compare numbers of points at which the quantiles are crossed.

3 CASE OF REGRESSION MODELS

In the follow-up, we shall assume that distribution of time-to failure depends on covariates. It means that we have to select a regression model for hazard rate and after its evaluation it is necessary to test the model fit. We shall discuss here several types of regression models, with the focus on Cox regression model and the accelerated failure time model. More details about regression models in survival analysis can be found in many monographs, let us again mention Kalbfleish and Prentice (2002).

3.1 *Additive regression model*

In this (also Aalen's) model, hazard function is specified as $h(t, z) = z \cdot \beta(t)$, where z represents values of covariates, $\beta(t)$ are functions of time, both z and β are p-dimensional. Their domains should ensure that $h(t, z) \geq 0$. As a rule, the first covariate component is taken fixed to 1, so that $\beta_1(t)$ has the meaning of a 'baseline' hazard function. In the sequel, by index i, $i = 1, .., n$ we shall denote individual objects, while by k, $k = 1, .., p$ components of vectors β, z.

The covariates themselves, $Z_i(t)$, can be different for each object and can change in time. Individual intensity of $N_i(t)$ is then

$$a_i(t) = Z_i(t) \cdot \beta(t) \cdot Y_i(t), \;\; i = 1, \ldots, n.$$

Cumulated functions $B_k(t) = \int_0^t \beta_k(s)\, ds$ are estimated by weighted least squares method. As $dN_i(t) = X_i(t)dB(t) + dM_i(t)$, where $X_i(t) = Z_i(t) \cdot Y_i(t)$, then $\hat{B}(t) =$

$$= \int_0^t (X(r)'W(r)X(r))^{-1}X(r)'W(r)dN(r),$$

where $W(r)$ is a matrix of weights; the simplest choice $W(r) = I_n$, identity matrix, optimal weights are $W(r) = diag\{1/a_i(r)\}$, in practice $\hat{a}_i(r)$ are used, computation is iterated.

Consistency and asymptotic normality of $\hat{B}(t)$ are straightforward, it holds that the term

$$\sqrt{n}(\hat{B}(t) - B(t)) = \sqrt{n} \int_0^t \overline{X}(r)dM(r),$$

where $\overline{X}(r) = (X(r)'W(r)X(r))^{-1}X(r)'W(r)$, is asymptotically distributed as a Gaussian process with independent increments (Brown process), its covariance function is estimable consistently by empirical version of

$$n \int_0^t \overline{X}(s)\, D(s, B(s))\, \overline{X}'\, ds$$

provided its limit exists, uniformly on $[0, T]$. Here $D(s, B(s))$ is a diagonal matrix with components $a_i(s)$.

It follows that the case is similar to the case of non-parametrized hazard rate treated in the 1-st part. Therefore it is possible to derive asymptotic distribution of residuals, again coinciding with Brown process distribution. It is described in detail in Volf (1996). Even Bayes residual analysis can follow the same scheme as in the preceding part, each function $\beta_k(t)$ has to be modelled separately, for instance by Arjas and Gasbarra (1994) approach.

4 COX REGRESSION MODEL

The case differs in certain aspects from preceding one, it is caused by more complicated asymptotic properties. The hazard rate is specified as $h(t, z) = h_0(t)exp(z \cdot \beta)$, with processes of covariates $Z_i(t)$ and parameter β (both p-dimensional), $h_0(t)$ is a baseline hazard rate, a nonnegative function.

The intensity of i-th process $N_i(t)$ is

$$a_i(t) = h(t, Z_i(t)) \cdot Y_i(t).$$

Parameter β is estimated from partial log-likelihood $L_p =$

$$\sum_{i=1}^{n} \int_0^T log\{\frac{exp(Z_i(t)\beta)}{\sum_{k=1}^{n} exp(Z_k(t)\beta) \cdot Y_k(t)}\}dN_i(t),$$

by an iterative procedure (of Newton-Raphson, as a rule), cumulated baseline hazard $H_0(t) = \int_0^t h_0(r)dr$ is then estimated as

$$\hat{H}_0(t) = \int_0^t \frac{dN(r)}{\sum_{k=1}^{n} exp(Z_k(t)\hat{\beta}) \cdot Y_k(t)}.$$

Theory on properties of estimates is collected elsewhere, first time the results has been established by Andersen and Gill (1982). Estimates are consistent, asymptotically normal, however, neither $\sqrt{n}(\hat{H}_0(t) - H_0(t))$ nor residual process are martingales.

4.1 *Residuals in Cox model*

Residuals are sometimes formulated more generally, as

$$dR(t) = \sum_{i=1}^{n} K_i(t) \cdot (dN_i(t) - d\hat{A}_i(t)),$$

with some (convenient) 'weight' processes $K_i(t)$, for instance if $K_i(t) = Z_i(t)$ (p-dimensional), $R(t)$ is then estimated score process (the first derivative) of L_p, while $K_i(t) = 1[i \in S]$ yields stratified residuals. Stratified residuals (the simplest case) are then expressed as $dR_S(t) =$

$$dM_S(t) + dH_0(t)C_S(\beta_0, t) - d\hat{H}_0(t) \cdot C_S(\hat{\beta}, t),$$

where $d\hat{H}_0(t) = \frac{dN(t)}{C(\hat{\beta},t)}$ and
$C_S(\beta, t) = \sum_{i \in S} exp(Z_i(t)\beta) \cdot Y_i(t)$,
$C(\beta, t) = \sum_{i=1}^{n} exp(Z_i(t)\beta) \cdot Y_i(t)$.

If we take approximately $\hat{\beta} \sim \beta_0$, we obtain expression similar to previous case without regression. Exact approach uses Taylor expansion of the last term at β_0. $R_S(t)/\sqrt{n}$

is then expressed with the aid of a martingale and a nonrandom function, with asymptotic distribution as a Gaussian process, however with rather complicated covariance structure. Hence, random generation of would-be residual processes with 'ideal' distribution under hypothesis of model fit is possible (but not easy). It is actually a bootstrapping, by which we obtain a sample of 'ideal' residual processes. Then, for instance their absolute maxima are compared with maximal residual computed from our data. Or certain their characteristics can be compared. That is why practical tests of Cox model fit is often performed just graphically, comparing visually how far are residuals in group S from zero line, or, equivalently, $\hat{A}_S(t)$ from $N_S(t)$. Thus, it seems that in the case of Cox model the Bayes analysis could offer an easiest tool for model fit assessing.

4.2 *Bayes procedure in Cox model*

Let us summarize briefly starting points and procedure of Bayes analysis in the Cox model setting, namely the variant applied in the following numerical example. MCMC algorithm of Metropolis-Hastings was used to obtain samples representing posterior distributions of β. Values of β-s were proposed from a prior (a sufficiently wide uniform, in our case), accepted or rejected with the use of partial likelihoods proportion. Then, to each β, a representation of $H_0(t)$ was generated, similarly as in the 1-st part. i.e. from a piecewise constant prior. Finally, we obtained a sample of both, $\beta^{(m)}, h_0^{(m)}(t)$, $m = 1, .., M$, from them the intensities and residuals (in a group S, say) were derived,

$$A_S^{(m)}(t) =$$
$$= \int_0^t h_0^{(m)}(r) \sum_{i \in S} exp\left(Z_i(r)\beta^{(m)}) \cdot Y_i(r)\right) dr,$$
$$R_S^{(m)}(t) = A_S^{(m)}(t) - N_S(t),$$

and used for model fit assessing.

5 AFT REGRESSION MODEL

In Accelerated Failure Time model it is assumed that individual hazard is changed by a factor depending on covariates. Quite commonly this factor has the form $exp(\alpha z)$, where z is a covariate, here constant in time. It follows that the distribution of time to failure T_i of an object with covariate value z_i has distribution function $F_i(t) = F_0(t \cdot e^{\alpha z})$, where F_0 characterizes a baseline distribution (of a random variable T_0 with covariate $z = 0$). It also means that $T_{0i} = T_i \cdot exp(\alpha z_i)$ is an i.i.d. representation of T_0. Logarithmic transformation yields that

$$\log T_i = -\alpha \cdot z_i + \log T_{0i}. \qquad (1)$$

Statistical inference based on (1) has to deal with unknown distribution of $\log(T_0)$, analysis is not straightforward and could be complicated further by the presence of censored data. Another approach uses semi-parametric likelihood based on hazard rates, similarly as in the case of Cox model. Namely, let $\{T_i, z_i, d_i, i = 1, .., n\}$ be observed times to failures or censoring of i-th object, their covariates, indicators of censoring, respectively, then the likelihood reads

$$L = \prod_{i=1}^n h_i(T_i)^{d_i} \cdot \exp(-\int_0^{T_i} h_i(t)dt), \qquad (2)$$

where $h_i(t) = h_0(t \cdot \exp(\alpha z_i)) \cdot \exp(\alpha z_i)$ is the hazard rate of i-th object at time t, h_0 is the baseline hazard rate of T_0. Theory of estimation and asymptotic properties are derived in Lin et al (1998) and further developed also in Bagdonavicius and Nikulin (2002, Ch. 6). Nevertheless, as the practical computation of asymptotic characteristics is rather complicated, Bayes approach can offer a reasonable alternative.

5.1 *Bayes analysis in AFT model*

Let us consider a MCMC procedure alternating two steps. In the first step, parameter α is fixed (from preceding iteration) and baseline hazard rate $h_0(t)$ is updated. Observed times T_i can be transformed to the scale of variable T_0: $T_i^0 = T_i \cdot \exp(\alpha z_i)$. Then, the likelihood of h_0 is

$$L = \prod_{i=1}^n [h_0(T_i^0) \cdot e^{\alpha z_i}]^{d_i} \cdot \exp(-\int_0^{T_i^0} h_0(t)dt). \qquad (3)$$

Likelihood (3) can be used either for direct computation of the Nelson-Aalen estimate of cumulated $H_0(t)$, or for innovation of Bayes estimator of $h_0(t)$, again for instance in the same manner as proposed by Arjas and Gasbarra (1994).

The second step updates estimate of parameter α, while current estimate of h_0 is taken as fixed. Values of α are proposed from a prior distribution, Metropolis-Hastings acceptance rule uses the proportion of likelihoods (3) with newly proposed and current values.

MCMC procedure yields a sequence of estimates $\alpha^{(m)}, h_0(t)^{(m)}$, $h_0(t)^{(m)}$ are piecewise constant nonnegative functions defined on selected finite interval. Hence, corresponding m-th estimate of the intensity of failure for i-th object equals $a_i^{(m)}(t) = h_0^{(m)}(t \cdot \exp(\alpha^{(m)} z_i)) \cdot \exp(\alpha^{(m)} z_i)$ on $[0, T_i]$ and equals zero for $t > T_i$. Cumulated intensities are then $A_i^{(m)}(t) = \int_0^t a_i^{(m)}(s) ds$, and , finally, the residual processes in a subset $S \subset \{1, 2, ..., n\}$ are again differences

$$R_S^{(m)}(t) = \sum_{i \in S} [A_i^{(m)}(t) - N_i(t)], \qquad (4)$$

where $N_i(t)$ is again a counting process of failure of i-th object, i.e. with maximally one step $+1$ at T_i provided $d_i = 1$. As it has been said, in practice the residuals are plotted against counts $N_S(t)$, not against time t.

Let us here also recall another approach to Bayes analysis in AFT model. It utilizes logarithmic model formulation (1). Instead with baseline hazard rate we then deal with baseline density for $\log T_0$. Often, its prior is constructed as a mixture of Gauss densities with weights given by Dirichlet distributions (as for instance in Gelfand and Kottas, 2003). However, complications caused by censoring remain and have to be overcome by an additional generation of would-be non-censored values, i.e. by a data augmentation. It is actually a randomized version of the EM algorithm used in non-Bayes inference.

6 EXAMPLE

A sample of $n = 100$ data was generated randomly from the following AFT model: Baseline distribution of $\log T_0$ followed normal distribution with $\mu = 0, \sigma = 0.5$, values of covariate Z were distributed uniformly in $(0, 2)$, corresponding accelerating parameter was set to

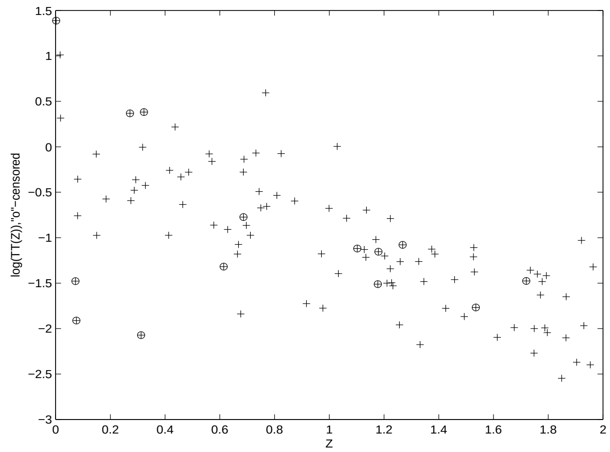

Figure 1: Data: Covariate is on x axis, log of survival on y axis, censored items are denoted by 'o'.

$\alpha = 1$. Further, values $T_i^* = T_{0i} exp(\alpha z_i)$ were randomly right-censored by i.i.d. variables distributed uniformly in $(min T_i^*, max T_i^*)$. Final censored data (in log scale) are displayed in Figure 1.

Then, data were analyzed in the framework of both AFT and Cox model, by Bayes approach described in preceding parts. In the AFT setting, together 5000 MCMC iterations were performed, last 2000 were used for the analysis. Posterior representation of α had the mean 1.0105 and standard deviation 0.0309. Figure 2. shows characteristics of posterior sample of residual processes, namely their pointwise medians and then also approximate 95% credibility bands, i.e. such a region that approximately 95% of residual processes lied fully inside. The bands are dashed, they are plotted against counts $N_S(t)$, separately for 2 groups $Z < 1$, $Z > 1$). It is seen that graphs are concentrated around zero, thus assessing good AFT model fit. Several trajectories of residual processes are displayed by dots.

In the Cox model framework, 5000 β-s were generated, again last 2000 were taken as a representation of posterior distribution of β. It had the mean 2.3665 and standard deviation 0.2688. Further, to each $\beta^{(m)}$ 200 instances of $H_0(t)$ were generated, we always took just the last of them. In such a way, a representation with M=2000 members was obtained. Figure 3. shows again the characteristics of posterior sample of residual processes, i.e. their point-

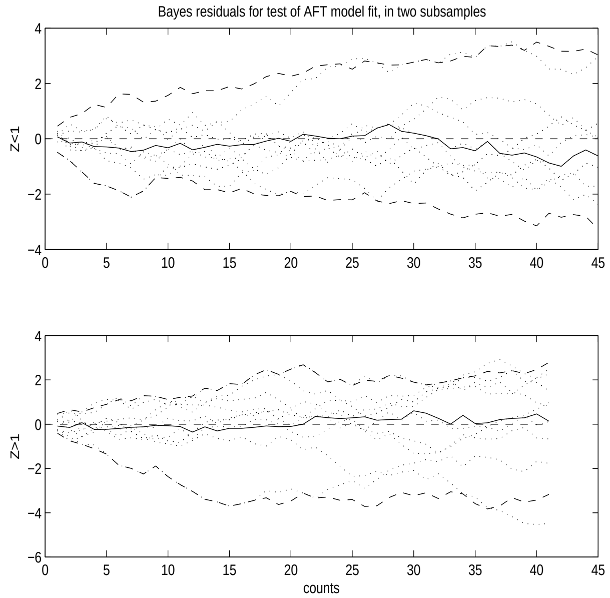

Figure 2: Characteristics of residuals in AFT model, in 2 subgroups with $Z < 1$ and $Z > 1$.

wise medians and approximate 95% credibility region, for two groups with $Z < 1$, $Z > 1$). Similarly as above, several trajectories of residual processes are displayed by dots. Departures of sample of residuals from the zero level is now rather significant, especially for low times in the first group. Graphs indicate that Cox model overestimates the failure intensity for small covariate values and also underestimate it for larger covariate values.

In the Cox model setting, standard analysis was performed, too. It yielded the estimate $\beta = 2.3687$ with asymptotic standard deviation 0.2786.

7 CONCLUSION

Models of lifetime often have to reflect a dependence on covariates. The present contribution was devoted to models goodness-of-fit tests based on martingale residuals. The main objective was to present Bayes variants of statistical analysis in the framework of semiparametric regression models and to show that especially in the cases of Cox and AFT regression model, Bayes approach could be a reasonable alternative to standard methods. On the other hand, the use of semi-parametric models together with an approach based on random generation is not convenient for analysis of cases of rare events with small amount of noncensored data.

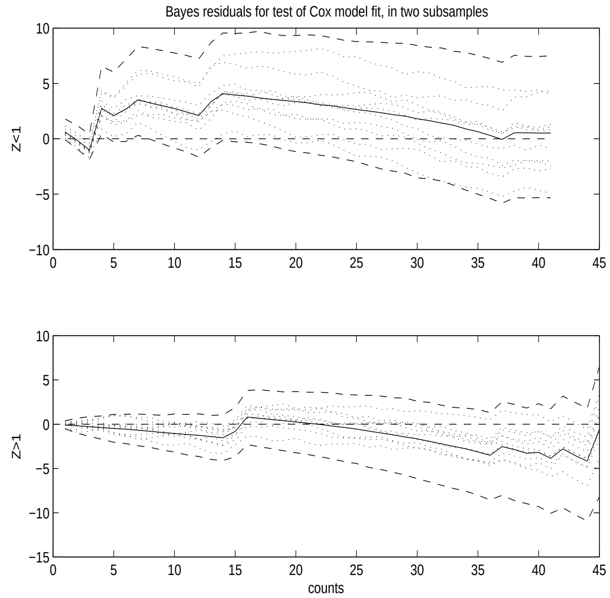

Figure 3: Characteristics of residuals in Cox model, in 2 subgroups with $Z < 1$ and $Z > 1$.

Acknowledgement: The research has been supported by the project of MŠMT ČR No. 1M06047, The Center of Quality and Reliability.

REFERENCES

Arjas, E. & Gasbarra, D. 1994. Nonparametric bayesian inference from right censored survival data using Gibbs sampler. *Statist. Sinica* 4: 505-524.

Andersen, P.K. & Gill, R. 1982. Cox's regression model for counting processes: a large sample study. *Ann. Statist.* 10: 1100-1120.

Bagdonavicius, V. & Nikulin, M. 2002. *Accelerated Life Models, Modeling and Statistical Analysis.* Boca Raton: Chapman&Hall/CRC.

Gamerman, D. 1997. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference.* Boca Raton: Chapman & Hall/CRC.

Gelfand, A.E. & Kottas, A. 2003. Bayesian semiparametric regression for median residual life. *Scand. J. Statist.* 30: 651-665.

Kalbfleisch, J.D. & Prentice, R.L. 2002. *The Statistical Analysis of Failure Time Data.* New York: Wiley.

Lin, D.Y., Wei, L.J. & Ying Z. 1998. Accelerated failure time models for counting processes. *Biometrika* 85: 605-618.

Volf P. 1996. Analysis of generalized residuals in hazard regression models. *Kybernetika* 32: 501-510.

Zuo Y. & Serfling R. 2000. General notion of statistical depth function. *Ann. Statist.* 28: 461-482.