MODELLING MARKOVIAN MIGRATION IN FINANCE AND MEDICINE

Rafael Weißbach

Dept. Statistics, Dortmund University of Technology,
44221 Dortmund, GERMANY, Rafael.Weissbach@tu-dortmund.de

ABSTRACT

For a time-continuous finite-state Markov process, constant transition intensities simplify the calculation of migration matrices. For two applications I investigate the adequacy of the model, first, rating migrations in finance, and second, cancer survival in medicine. In finance, ratings have usually more than one transient state. In contrast, in medicine "alive" as single transient state is common. Still, nonparametric methodology, like the Nelson-Aalen estimate, or likelihood ratio testing is very similar for both situations. The re-analysis of several data shows that the constant intensity is usually to be rejected, however alternative models need arguments of the specific field. Nonparametric smoothing may an intermediate step towards such models.

1. INTRODUCTION

I have two areas of application in mind when analyzing transition data, rating transitions (or migrations) and cancer survival. The evolvement of ratings, see e.g. [1] for a description, is depicted in Figure 1. Statistical analysis can be based on counting the transitions from one state to any other (for all combinations). Of interest for the financial analyst is typically P(rating after year 1 is j | rating at credit origination is h). Statistically, counting deaths,

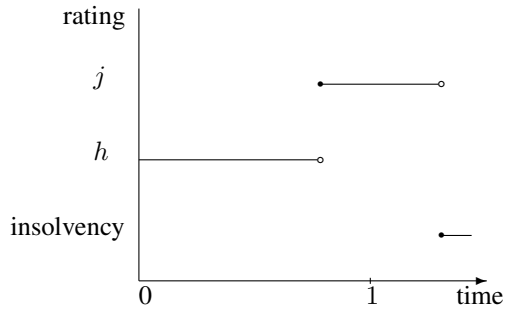

Figure 1. Exemplary evolvement of rating for a counterpart

i.e. transitions from "alive" to "dead", is very similar to counting transitions between rating classes. The analysis of deaths is very common in cancer studies [2, 3, 4] and usually performed without a parametric assumption as long as no covariates are incorporated. The Kaplan-Meier estimate [5] and the log-rank test [6] for equality between

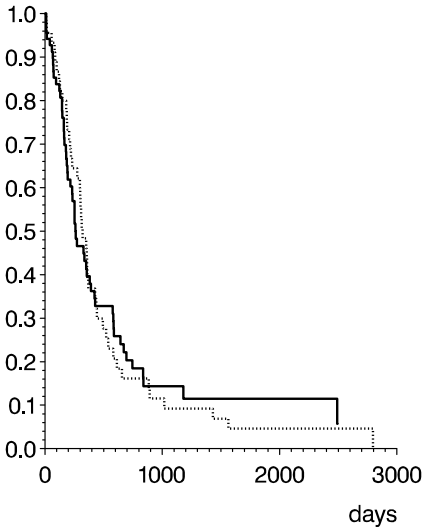

Figure 2. Kaplan-Meier estimation of survival function for patients with low concentration in Metallothionein ($\leq 10\%$) (solid line) and with high concentration ($> 10\%$) (dashed line)

groups have become very popular. However, more grades of "alive" are also conceivable in medical studies. As typical in cancer studies, survival is compared in between two groups. For the data in [4] the Kaplan-Meier estimates are shown in Figure 2.

However, data in finance and medicine are often quit different as the following comparison of data sets shows. Rating systems typically have several states (or 'rating classes' as is the terminology in finance). E.g. for the data that are re-analyzed here, one finds 8 ordered transient rating states (and 1 insolvency state) [7] or even 25 transient rating states [8]. For cancer studies, there is mostly one "alive" state and one "dead'" state [2, 4] The sample sizes are also quit different. Sample sizes of $3,700$ [7] and 100 [8] are found in finance, whereas in medicine smaller sizes like 99 [2], 118 [4], and occasionally up to 542 [9] are common. Unfortunately, sample sizes are not the only determinant for the adequacy of asymptotic results in statistics. Essentially, it is the cross-product between number of states, sample size, overall follow-up time and the overall intensity level that determine the adequate analysis. For instance, 3,700 rating histories might be followed over

seven years (1997-2003) [7] and still the number of transitions can be below the number of observed processes, namely 2,750. On the other hand, 100 process followed-up for 12 years (1994-2006) can result in 240 transitions [8]. In the medical examples, the amount of information may be easier displayed by the rate of censoring and the follow-up time. For the rectal cancer 75% censoring and at median follow-up of five years indicates little chance for complex models [2]. In the bladder cancer study [4], 10% censoring at a maximal follow-up duration of nine years already shows how lethal the cancer (in this study) is. The skin cancer [9] has much better perspectives, so that 66% censoring (at a median follow-up of six years) reduces the sample size of 540 patients considerably. We see that the data specifics do reduce the amount of generally useful methodology. However, we will see in the following that some methods are applicable to all data sets.

2. INTENSITY ESTIMATION

The intensity of transition is now the mathematical nomenclature, steaming from the analysis of markov chains and the Poisson process theory. In financial and medical statistics it is usually called the hazard rate [6]. We want to model the hazard rate and in order not to imposing a parametric model a priori, I like to resort to nonparametric statistics first. One is interested in the hazard rate

$$\begin{aligned}\lambda(t) &:= \lim_{dt\to 0}\frac{P(T\in[t,t+dt)|T\geq t)}{dt}\\ &= \frac{f(t)}{1-F(t)}\end{aligned}$$

of some transition time T having density $f(\cdot)$ and cumulative distribution function $F(\cdot)$. The hazard rate describes the instantaneous probability to migrate, e.g. to default or to die, conditional on the fact that the transition has not yet taken place. This probability may change over time t.

2.1. Nonparametric smoothing

If the process is homogeneous, the intensity is constant and may be estimated by maximum likelihood [10]. This implies the 1-year migration probability to be constant. As a consequence, t-years migration probabilities are initially identical to one year after the origin of time. The Kaplan-Meier estimate is easily generalized to t-year migration probabilities. The homogenous markov model can be checked graphically, see Figure 3 for the data in [7]. Here we see that the homogeneity is not very convincing.

The typical analysis is now to estimate the hazard rate by smoothing methods. In kernel smoothing, the Nelson-Aalen estimate $\hat{\Lambda}(\cdot)$ of the cumulative hazard rate $\Lambda(t) := \int_0^t \lambda(s)ds$ is convoluted with a kernel $K(\cdot)$ [11]:

$$\lambda_n(t) := \int_{\mathbb{R}_0^+} K_{h^{NN}}(t-s)d\hat{\Lambda}(s) \qquad (1)$$

As the fixed bandwidth is difficult to apply to failure time data, I propose to use the nearest-neighbor bandwidth

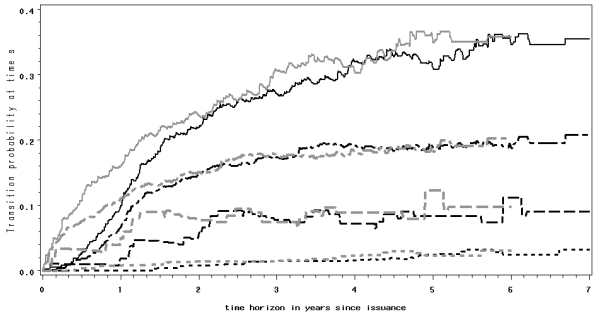

Figure 3. Transition probabilities $p_{hj}(s, s+\psi)$, $h, j \in K$ for various rating combinations h, j and starting times $s = 0$ (black lines) and $s = 1$ (grey lines) after time of issuance plotted against time horizon Ψ: $2 \to 3$ (black solid line), $4 \to 3$ (black long-short dashed line), $7 \to 8$ (grey dashed line), $7 \to 6$ (black dashed line), $2 \to 7$ (black short-dashed line)

$h^{NN}(t)$ defined by

$$\inf\left\{r>0 : |S_n\left(t-\frac{r}{2}\right) - S_n\left(t+\frac{r}{2}\right)| \geq \frac{k}{n}\right\}$$

with Kaplan-Meier $S_n(\cdot)$.

The consistency for the estimate has the following uniform rate [11].

$$\sup_{t\in[a,b]} |\lambda_n(t)-\lambda(t)| \leq D_1\sqrt{\frac{\log(n)}{k}} + D_2\frac{k}{n} \quad a.s.$$

Constants D_1 and D_2 depend on kernel and intensity.

Here the second term in the boundary represents the bias. However, the bias is in practice a mayor concern also in finite samples. First I will now compare the bias of the estimate (1) with a fixed-bandwidth estimate. As distribution family I select the exponentiated Weibull distribution [12]. It is defined in terms of the survival function

$$S(t) = 1 - \left(1-\exp\left(-\left(\frac{t}{\gamma}\right)^{\kappa}\right)\right)^{\theta},$$

with $0 < x < \infty$, $\kappa > 0$, $\theta > 0$ and $\gamma > 0$. So that the family contains the original Weibull distribution for $\theta = 1$. Four shapes of the hazard rate modeled in this family are identifiable by parameter space segments: Increasing, decreasing, unimodal and convex (bath-tube) shaped. The limiting lines are $\kappa = 1$ and $\kappa\theta = 1$. The convex hazard rate is the most challenging of the four. It has a steep increase towards the left boundary, the time-origin. Additionally, it increases steeply towards the right, where fading data will constitutes the main problem. For the sake of brevity, I restrict the discussion to the estimation of that type, with the parameters $\kappa = 5$, $\theta = 0.1$ and $\gamma = 100$.

Preliminary simulations have shown that it is advisable to restrict estimation to the inner 80% area. For the left and the right 10% quantiles, the boundary bias prohibits a sound interpretation. For the selected convex hazard rate the inner 80% area is

$$[F^{-1}(0.1), F^{-1}(0.9)] = [1; 84.5].$$

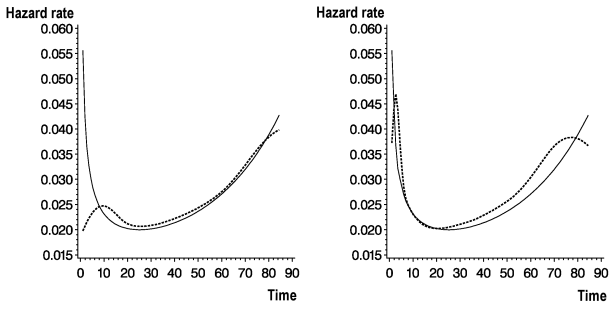

Figure 4. Fixed bandwidth (left) vs. k-nearest neighbor bandwidth (right) (solid line is true hazard)

A typical defect on failure time data is censoring. In detail, we assume to observe right-censored default times $X_i = \min(T_i, C_i)$ with the censoring indicator $\delta_i = 1_{\{X_i=T_i\}}$. The default times $T_1, \ldots, T_n$ are assumed to be independent and identically distributed with distribution function F, the censoring time $C_1, \ldots, C_n$ are assumed to be independent and identically distributed with distribution function G. The T_i's are assumed to be independent of the censoring times C_i's. The studies cited in the Introduction illustrate that, at least, censoring rates between 10% and 75% are conceivable. As an example, censoring of 40% is used here. A simple means to achieve this rate is choosing survival times T_i and censoring times C_i both due to the exponentiated Weibull family with similar parameters. Starting from an expected 50% censoring for similar distributions, one may use the monotony of the expected rate of censoring, i.e. of $P(C_i > T_i)$, with respect to the parameter θ, in order to obtain the desired 40% by simulation. The resulting parameter set is $(5, 0.15, 100)$. Sample sizes n of 50 and 100 observations mimic the situation in [8]. A medium size of 300 observations resembles the situation in [7]. To my experience, 500 simulation runs are sufficient for point estimation purposes, for the 300-observation situation 250 replications reduce the time of computation.

As random number generator, I use the generator of uniformly distributed random number on $[0, 1]$ as implemented in SAS/IML and map to the distribution family via the inverse cumulative distribution function

$$F^{-1}(u) = \gamma[-\log(1 - u^{\frac{1}{\theta}})]^{\frac{1}{\kappa}}, \quad 0 < u < 1.$$

Using a sample size of 300, and averaging over 250 random samples, makes the bias comparable. For the fixed bandwidth I use Silverman's rule of thumb [13] for the bandwidth selection and for the nearest neighbor bandwidth I use a cross-validatory selector [14]. The nearest neighbor bandwidth apparently avoids boundary bias problems, which the fixed bandwidth has at the origin (Figure 4).

Silently, I have already made use of bandwidth selectors. However, bandwidth selection, or in the case of the nearest neighbor, the selection of the number of neighbors is critical.

Two types of bandwidth selectors are found frequently in the literature. First of all, one may assume a - counterfactual - distribution. Minimization of, say, the mean integrated squared error with respect to the bandwidth results in an analytical expression for the optimal bandwidth [15, 16]. Without a particular distributional assumption, the second idea is to numerically minimize an error by cross-validation, usually, leave-one-out estimation is used here [17, 18, 19, 20].

For the fixed bandwidth b in kernel density estimation, Silverman's rule-of-thumb assumes normality of the data and minimizes asymptotically the mean integrated squared error [16]. The solution h^{RoT} is explicitly given and results in close-from expression for the for the hazard rate under random censoring and for the nearest neighbor bandwidth [9]. The adoption to the nearest neighbor bandwidth is achieved by identifying the fixed bandwidth to imply a linear approximation of the cumulative distribution function, upon which can be improved by stochastic approximation, with the empirical process. The number of nearest neighbors is given by

$$k^{RoT} = \left[n \cdot |\hat{\beta}| \cdot h^{RoT} \right],$$

with $\hat{\beta}$ as regression slope through the points $(X_i, S_n(X_i))$ of the Kaplan-Meier survival estimate.

Bandwidth selection for the nearest neighbor bandwidth in hazard rate estimation can be implemented by cross-validation in [14]. Maximizing a leave-one-out likelihood results in an optimal bandwidth and minimizes asymptotically the expected Kullback-Leibler loss for the corresponding density [21]. In detail, the likelihood is decomposed into the hazard rate and the survival function, both estimated by cross-validation.

$$k^{CV} = argmax_{k \in \{1,\ldots,n\}} \prod_{i=1}^{n} \lambda_n^{-i}(X_i)^{\delta_i} S_n^{-i}(X_i).$$

Here $\lambda_n^{-i}(X_i)$ estimates the hazard rate at time X_i by (1) on basis of the entire sample except for X_i, the same applies for the Kaplan-Meier estimate $S_n^{-i}(\cdot)$ of the survival function. Enumeration over the possible numbers of nearest neighbors yields the optimum. For an example of obviously not constant hazard rates see Figure 5. E.g. two Weibull distributions with proportional hazards may be more adequate in the case of skin cancer modelling.

3. GOODNESS-OF-FIT TESTING

A first step towards parametric models is goodness-of-fit testing. For a given parametric model a test can be developed that against the hypothesis of that particular model. For the case of one transient Markov state, the Kolmogorov-Smirnov principle is applicable to test globally consistent e.g. for a constant hazard rate [8]. For the general multi-state process a test against a piece-wise constant hazard rate [7].

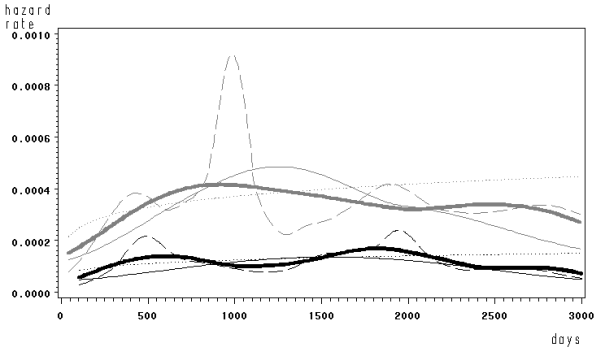

Figure 5. Hazard rate estimates for survival with melanoma in the presence of ulceration (gray lines) and in the absence of ulceration (black lines): (i) double-smoothing approach (solid bold lines), (ii) fixed-bandwidth estimate using the normal-scale rule (solid thin lines), (iii) and nearest-neighbor estimate using cross-validation (dashed lines)). The parametric estimates (Weibull) with a constant hazard ratio of three are indicated (dotted lines).

3.1. Kolmogorov-Smirnov

The one-sample log-rank test assesses equality of the hazard to a known hazard rate $\lambda_0(\cdot)$ [22]. (Especially its two-sample formulation is frequently used in the biometry of cancer research (see e.g. [4])). The idea is to compare the increments of the unrestricted estimator of the cumulative hazard rate with those of the estimator parametrically arising from the hazard rate model under the hypothesis $H_0: \lambda(t) = \lambda_0(t)$.

I want to test the hazard rate for constancy. As a prerequisite we need to recall the maximum likelihood estimate for an exponential distribution with density $f(t) = \lambda e^{-\lambda t}$:

$$\lambda_n := \frac{n_u}{\sum_{i=1}^n X_i}.$$

The number of uncensored observations is defined as $n_u := \sum_{i=1}^n \delta_i$. How to construct the test statistic for the hypothesis

$$H_0: \quad \text{“Is } \lambda(t) \text{ constant?”} \quad \Leftrightarrow \quad \lambda(t) \equiv \lambda_0 \quad \text{unknown?}$$

In detail, the differences of the increments of the unrestricted estimator of the cumulative hazard and those of the estimator arising for an exponential distribution are added up: In the non-parametric analysis the process of counting the number of uncensored events generalizes the “number of events” n_u in:

$$N(t) := \sharp\{i : X_i \le t, \delta_i = 1\}.$$

The population under risk at time t generalizes the “exposure to risk” $\sum X_i$:

$$Y(t) := \sharp\{i : X_i \ge t\}.$$

The increment of the compensator of the counting process, $E(dN(t) \mid \mathcal{F}_t) = Y(t)\lambda(t)$, depends on the hazard rate. The Nelson-Aalen estimate can now be written as

$$\hat{\Lambda}(t) = \int_{0<s\le t} \frac{J(t)}{Y(t)} dN(s),$$

where $J(t) := I\{Y(t) > 0\}$ and $\frac{0}{0} := 0$ (see [23, 24]). The definition of $J(t)$ “stops” the estimation of the cumulative hazard rate at the largest observation $X_{(n)}$ (irrespective of censoring).

Under the null hypothesis this yields

$$\Lambda_0^\star(t) \quad = \quad \int_{0<s\le t} J(s)\lambda_n ds.$$

This is the parametric estimate for the cumulative hazard rate, again capped at $X_{(n)}$.

The log-rank statistic weights the incremental differences with the number of observations under risk and adds them [6]:

$$\begin{aligned} Z(t) \quad &:= \quad \int_0^t Y(s) d\{\hat{\Lambda}(s) - \Lambda_0^\star(s)\} \\ &= \quad \int_0^t dN(s) - Y(s)\lambda_n ds, \end{aligned}$$

where

$$E^\star(t) := \int_0^t Y(s)\lambda_n ds$$

is the expected number of defaults in $[0, t]$ under hypothesis. Let us look at the Kolmogorov-Smirnov-type test statistic of the process $Z(t)$. For this purpose define

$$\hat{\sigma}^2 := \int_0^\tau \frac{dN(s)}{n\lambda_n^2}.$$

For hazard rates bounded away from 0 on $[0, \tau]$ it follows due to [25] that, under the null hypothesis of a constant hazard rate, the process

$$(Z(t)/(\sqrt{n}\lambda_n\hat{\sigma}))_{t\in[0,\tau]}$$

converges weakly in $D[0, \tau]$ to a Brownian bridge, say W^0, in time $p(t) = t/\tau$, i.e.

$$\frac{Z(t)}{\sqrt{n}\lambda_n\hat{\sigma}} \stackrel{d}{\Longrightarrow} W^0 \circ p.$$

According to the continuous mapping theorem (see [26]) the Kolmogorov-Smirnov-type statistic is

$$K_n \quad = \quad \max_{0\le t\le\tau} \left| \frac{Z(t)}{\sqrt{n}\lambda_n\hat{\sigma}} \right| \quad \stackrel{d}{=} \quad \max_{0\le t\le 1} |W^0(t)|$$

In particular, if k_α denotes the $1-\alpha$ quantile of the distribution of the random variable $\max_{0\le t\le 1} |W^0(t)|$ the rule $\phi((X_1, \delta_1), \ldots (X_n, \delta_n))$ defined by

$$I_{\{\max_{0\le t\le\tau} \left| \int_0^t dN(s) - Y(s)\lambda_n ds \right| > k_\alpha \hat{\sigma}\sqrt{n}\lambda_n\}}$$

corresponds to the Kolmogorov-Smirnov statistic and constitutes a test with asymptotic level α. A typical choice for τ is the maximum over all observations X_i. By this, no information is ignored in the construction of the statistic. The quantiles k_α can be found in [26], p. 143. For example, the 95% quantile is given by $k_{0.05} \approx 1.36$.

For a 10% sub-sample of the data in [7] one has e.g. for the transition 3→4 a test statistic that 72.38 that exceeds the critical value of 63.77.

3.2. Likelihood Ratio

Our goal is now to check homogeneity simultaneously in a multi-state process with state space K of k classes. Do transition probabilities, thus, exhibit time-invariant behavior? Therefore, the null hypothesis of time-homogeneity can be stated in terms of

$$H_0 : \forall h, j \in K : \exists q_{hj} \in \mathbb{R}_+ : q_{hj}(t) \equiv q_{hj} \forall t \in [0, T].$$

We are interested in the alternative that transition probabilities are time-dependent which can be approximated by piece-wise constant transition intensities $H_1 : \exists h, j \in K$, so that $q^b_{hj}(t)$ equals

$$\sum_{i=1}^{b} q_{hji} I_{[t_{i-1},t_i)}(t) : \exists i_1, i_2 = 1, \ldots, b : q_{hji_1} \neq q_{hji_2}$$

where $0 = t_0 \leq t_1 \leq t_1 \leq \ldots \leq t_b = \tau$ partitions the time-line. In order to constructing a likelihood ratio test (as in the discrete case [27]), following [28][pg. 296], I consider the partial-likelihood for n independent processes. The transition data for the sample are given by the multivariate counting process $\mathbf{N}$ with components $N_{hj}(t)$ that count the migrations for each state combination up to t. The likelihood gets the product of all migration times and over all rating combinations

$$\left(Y_h(t) \underbrace{d\hat{\Lambda}_{hj}(t)}_{q_{hj}(t)dt} \right)^{dN_{hj}(t)} \left(1 + \underbrace{d\hat{\Lambda}_{hh}(t)}_{q_{hh}(t)dt} \right)^{Y_h(t) - dN_h(t)}$$

where $N_h(t) := \sum_{j=1, j\neq h} N_{hj}(t)$.

The actual value of the likelihood ratio can be calculated by inserting in the likelihood the ML-estimators for the transition intensities $q_{hj}(t)$ [10] under the null and the alternative, if they exist.

With respect to the fact that transition intensities $q^b_{hj}(t)$, $h \neq j$ are piecewise constant under the alternative we can also apply this estimator to derive the ML-estimator for $q^b_{hj}(t)$. Therefore, we only have to find ML-estimators for the b parameters $q_{hji}, i = 1, \ldots, b$ of the step functions $q^b_{hj}(t)$. This can be accomplished by confining the considered time interval, where rating transitions are observed, from $[0, T]$ to the i-th time interval $[t_{i-1}, t_i)$ of the partition of $[0, T]$. Thereby, the ML-estimator of $q_{hji}, i = 1, \ldots, b$ is given by

$$\hat{q}^n_{hji} = \frac{N_{hj}(t_i-) - N_{hj}(t_{i-1}-)}{\int_{t_{i-1}-}^{t_i-} Y_h(t)dt},$$

for $\int_{t_{i-1}-}^{t_i-} Y_h(t)dt > 0$. It could be regarded as the ML-estimator $\hat{q}^n_{hj}$ for a time homogeneous Markov process on this restricted time interval, and thus only takes into account rating transitions that occur within $[t_{i-1}, t_i)$.

The test statistic of the likelihood ratio test for the preceding test problem is

$$\Phi := -2\ln(LR) = -2(\ln L_0 - \ln L_1).$$

Using standard theorems of likelihood testing [29], Φ is, under the null, asymptotically χ^2-distributed with $(b-1)(k-1)^2$ degrees of freedom.

E.g. in the rating data set of [7], the $p - value$ for yearly pieces in the intensity is below 0.001.

4. DISCUSSION

It must be admitted that many other parametric models, apart from the piece-wise constant model, appear to be useful. E.g. two Weibull distributions with proportional hazards may be adequate in the case of skin cancer modelling in two groups [9]. So that maybe a synthesis of parametric and non-parametric estimation, i.e. semi-parametric modeling, is a promising field. On the other hand, step-wise approximation may still be worth investigating and the optimal number of steps could act as bandwidth surrogate. More fundamentally, one must distinguish between significant effects and relevant effects (see e.g. [30]). Furthermore, the differences between parametric testing (e.g. with the likelihood ratio) and non-parametric testing must be further evaluated.

5. ACKNOWLEDGMENTS

I am grateful to O. Gefeller, H. Dette, C. Lawrenz, P. Tschiersch, A. Pfahlberg, M. Gordy, N. Hjort, Ø. Borgan and an anonymous referee for advice. Furthermore, the financial support of the Deutsche Forschungsgemeinschaft (SFB 475, "Reduction of complexity in multivariate data structures," project B1, and Grant Ge 637/3) is gratefully acknowledged.

6. REFERENCES

[1] FitchRatings, "Fitch ratings global corporate finance 2004 transition and default study," Research reports (credit market research), Fitch Ratings, 5 2005.

[2] K. Günther, J. Leier, G. Henning, A. Dimmler, R. Weißbach, W. Hohenberger, and R. Förster, "Prediction of lymph node metastasis in colorectal carcinoma by chemokine receptors CCR7 and CXCR4," *International Journal of Cancer*, vol. 116, pp. 726–733, 2005.

[3] W.M. Brueckl, J.M.M. Preuss, A. Wein, A. Jung, T. Braletz, R. Pflüger, G.H. Wiest, T. Kirchner, E.G. Hahn, W. Hohenberger, and K. Guenther, "Ki-67 and resection margin status are associated with disease free survival in desmoid tumor patients," *Anticancer Research*, vol. 21, pp. 3615–3620, 2001.

[4] L.L. Siu, D. Banerjee, R.J. Khurana, X. Pan, R. Pflüger, I.F. Tannock, and M.J. Moore, "The prognostic role of p53, metallothionein, p-glycoprotein, and mib-1 in muscle-invasive urothelial transitional cell carcinoma," *Clinical Cancer Research*, vol. 4, pp. 559–565, 1998.

[5] Ø. Borgan, "Three contributions to the encyclopedia of biostatistics: The Nelson-Aalen, Kaplan-Meier, and Aalen-Johansen estimators," Tech. Rep., Department of Mathematics, University of Oslo, 1997.

[6] P.K. Andersen, Ø. Borgan, R.D. Gill, and N. Keiding, *Statistical Models Based on Counting Processes*, Springer-Verlag New York, 1993.

[7] R. Weißbach, P. Tschiersch, and C. Lawrenz, "Testing time-homogeneity of rating transitions after origination of debt," *Empirical Economics*, vol. to appear, 2008.

[8] R. Weißbach and H. Dette, "Kolmogorov-smirnov-type testing for the partial homogeneity of markov processes - with application to credit risk," *Applied Stochastic Models in Business and Industry*, vol. 22, pp. 223–234, 2007.

[9] R. Weißbach, A. Pfahlberg, and O. Gefeller, "Double-smoothing in kernel hazard rate estimation," *Methods of Information in Medicine*, vol. 47, pp. 167–173, 2008.

[10] A. Albert, "Estimating the infinitesimal generator of a continuous time, finite state markov process," *Annals of Mathematical Statistics*, vol. 38, pp. 727 –753, 1962.

[11] R. Weißbach, "A general kernel functional estimator with general bandwidth – strong consistency and applications," *Journal of Nonparametric Statistics*, vol. 18, pp. 1–12, 2006.

[12] G.S. Mudholkar, D.K. Srivastava, and M. Freimer, "The exponential weibull family: A reanalysis of the bus-motor-failure data," *Technometrics*, vol. 37, pp. 436–445, 1995.

[13] M.P. Wand and M.C. Jones, *Kernel Smoothing*, Chapman & Hall, London, 1995.

[14] O. Gefeller, R. Pflüger, and T. Bregenzer, "The implementation of a data-driven selection procedure for the smoothing parameter in nonparametric hazard rate estimation using sas/iml software," in *Proceedings of the 13th SAS European Users Group International Conference, Stockholm*, H. Friedl, A. Berghold, and G. Kauermann, Eds. SAS Institute Inc. Carry, July 1996, pp. 1288–1300.

[15] E. Parzen, "On the estimation of a probability density function and the mode," *Annals of Mathematical Statistics*, vol. 33, pp. 1065–1076, 1962.

[16] B.W. Silverman, *Density Estimation*, Chapman & Hall, London, 1986.

[17] D.W. Scott and G.R. Terrell, "Biased and unbiased cross-validation in density estmation," *Journal of the American Statistical Association*, vol. 82, pp. 1131–1146, 1987.

[18] Y.S. Chow, S. Geman, and L.D. Wu, "Consistent cross-validated density estimation," *Annals of Statistics*, vol. 11, pp. 25–38, 1983.

[19] J.S. Marron, "A comparison of cross-validation techniques in density estimation," *Annals of Statistics*, vol. 15, pp. 152–163, 1987.

[20] P.N. Patil, "On the least square cross-validation bandwidth in hazard rate estimation," *Annals of Statistics*, vol. 21, pp. 1792–1810, 1993.

[21] P. Hall, "Cross-validation in density estimation," *Biometrika*, vol. 69, pp. 383–390, 1978.

[22] N.E. Breslow, "Analysis of survival data under the proportional hazards model," *International Statistical Review*, vol. 43, pp. 45–58, 1975.

[23] W. Nelson, "Theory and applications of hazard plotting for censored failure data," *Technometrics*, vol. 14, pp. 945–966, 1972.

[24] O.O. Aalen, "Nonparametric estimation of partial transition probabilities in multiple decrement models," *Annals of Statistics*, vol. 6, pp. 534–545, 1978.

[25] N.L. Hjort, "Goodness of fit tests in models for life history data based on cumulative hazard rates," *Annals of Statistics*, vol. 18, pp. 1221–1258, 1990.

[26] G.R. Shorack and J.A. Wellner, *Empirical processes with applications to statistics*, John Wiley & Sons, New York, 1986.

[27] N.M. Kiefer and C.E. Larson, "A simulation estimator for testing the time homogeneity of credit rating transitions," *Journal of Empirical Finance*, vol. 14, pp. 818–835, 2007.

[28] P. Andersen, Ø. Borgan, R. Gill, and N. Keiding, *Statistical Models Based on Counting Processes*, Springer, New York, 1993.

[29] R. J. Serfling, *Approximation theorems of mathematical statistics*, J. Wiley, New York, 1980.

[30] A. Munk and R. Pflüger, "$1-\alpha$ equivariant confidence rules for convex alternatives are $\alpha/2$-level tests – with applications to the multivariate assessment of bioequivalence," *Journal of the American Statistical Association*, vol. 94, pp. 1311–1320, 1999.