Fahrmeir:

Discrete failure time models

Sonderforschungsbereich 386, Paper 91 (1997)

Online unter: http://epub.ub.uni-muenchen.de/

Discrete failure time models

Ludwig Fahrmeir, Universität München

Introduction

Most methods for analyzing failure time or event history data are based on time as a continuously measured variate. A basic assumption for large parts of theory is that failure times are untied, see Andersen et al. [2]. In practice, there is always some smallest time unit, so that ties can occur. A moderate number of ties, while banned in theory, can be treated by appropriate modifications. If many ties occur, e.g. due to grouping in larger time units or intervals, or if time is truly discrete, then discrete survival or failure time models are more consistent with the data. Such situations arise in medical work when patients are followed up at fixed intervals like months, in certain biostatistical problems, for example human fertility studies and time to pregnancy (Scheike and Jensen [19]), or in labor market studies where duration of unemployment is measured in weeks, at best, or in months. We review parametric models and outline recent nonparametric approaches. More details, in particular for parametric models, are given e.g. in Fahrmeir and Tutz [11], ch. 9, and further references cited there.

Basic concepts

Let time be divided into intervals $[a_0, a_1)$,$[a_1, a_2)$,. . .,$[a_{q-1}, a_q)$,$[a_q, \infty)$. Usually $a_0 = 0$ is assumed, and a_q denotes the final follow up. Identifying the discrete time index t with interval $[a_{t-1}, a_t)$, a discrete failure time T is considered, where $T = t$ denotes failure within interval $t = [a_{t-1}, a_t)$. The basic quantity characterizing T is the *discrete hazard function*

$$\alpha(t) = Pr(T = t|T \geq t), \quad t = 1, \ldots, q\,, \tag{0.1}$$

which is the conditional probability for the risk of failure in interval t given the interval is reached. The *discrete survivor function* for reaching interval t

is

$$S(t) = Pr(T \geq t) = \prod_{s=1}^{t-1} [1 - \alpha(s)] \,, \qquad (0.2)$$

and the unconditional probability for failure at t is $Pr(T = t) = \alpha(t)S(t)$.

For a homogeneous population, discrete failure time data are given by $(t_i, \delta_i), i = 1, \dots, n$, where $t_i = \min(T_i, C_i)$ is the minimum of survival time T_i and censoring time C_i, and δ_i is the indicator variable for failure ($\delta_i = 1$) or censoring ($\delta_i = 0$). For the following we assume that censoring occurs at the end of the intervals, otherwise appropriate modifications have to be made. Simple estimates for $\alpha(t)$ are *crude death rates* $\hat{\alpha}(t) = d_t/n_t$, where n_t is the size of the population at risk and d_t the number of observed failures in $[a_{t-1}, a_t)$. The so-called *standard life table estimate* replaces n_t by $n_t - w_t/2$, where w_t is the number of censored observations in $[a_{t-1}, a_t)$, thereby assuming that censored observations are under risk for half the interval. In particular for large t, where the size n_t of the risk set often becomes small, these estimates may be quite unsteady, and smoothing by one of the nonparametric methods outlined further below will be appropriate. This is illustrated in Figure 1.

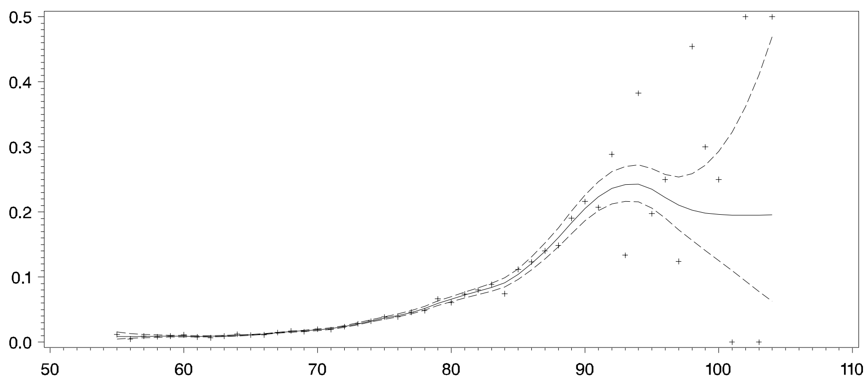

Figure 0.1: Posterior mean estimates (solid line) and pointwise two standard deviation confidence bands together with crude death rates (plus)

It shows crude death rates at age t in years for a population of retired American white females together with a smoothed estimate. A look at the data

(Green and Silverman [13], p. 101) shows that n_t becomes rather small for higher age t, resulting in unstable estimates towards the end of the observation period.

Discrete failure time data can also be described by *discrete-time counting processes** $N_i(t), i = 1, \ldots, n$, defined by $N_i(0) = 0$ and

$$\Delta N_i(t) = N_i(t) - N_i(t-1) = \begin{cases} 1 & \text{if individual } i \text{ is at risk and fails at } t \\ 0 & \text{else} \end{cases}$$

for $t \geq 1$, see e.g. Arjas and Haara [3]. Thus, for every individual i under risk at t, the value $\Delta N_i(t)$ can be considered as the outcome of a binary experiment, with $Pr(\Delta N_i(t) = 1) = \alpha(t)$. The sum $N(t) = \sum_i N_i(t)$ counts the number of observed failures up to t, and crude death rates can be derived as nonparametric maximum likelihood estimators, in analogy to the Nelson-Aalen estimator* for continuous time.

In most studies a vector of possibly time-dependent basic or derived covariates $\mathbf{x}_{it}$ is observed in addition to failure times. Time-dependent components of $\mathbf{x}_{it}$ are assumed to be fixed within interval t. Then the hazard function for survival time T_i of individual i will generally depend on covariates and is defined by

$$\alpha_i(t|\mathbf{x}_{it}^*) = Pr(T_i = t | T_i \geq t, \mathbf{x}_{it}^*), \quad t = 1, \ldots, q\,, \tag{0.3}$$

where $\mathbf{x}_{it}^* = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{it})$ denotes the history of covariates up to time t. Expressions for the survivor function (0.2) and for $Pr(T = t)$ have to be modified accordingly. Also, the sequence of binary experiments for $\Delta N_i(t), t \geq 1$, will depend on $\mathbf{x}_{it}$. Unless separate analysis for homogeneous subgroups can be carried out, it is natural to describe the dependence of conditional probabilities of failure by binary regression models. Let F_{t-} denote the history of events registered up to time t, but excluding the failure at t, and let $r_i(t)$ denote a risk indicator with $r_i(t) = 1$ if individual i is at risk in interval t, $r_i(t) = 0$ otherwise. Then it will be assumed that the conditional probability of failures can be expressed as

$$Pr(\Delta N_i(t) = 1 | F_{t-}) = r_i(t)\alpha_i(t|\mathbf{x}_{it}^*)$$

with hazard functions linked to a time-varying predictor η_{it} by

$$\alpha_i(t|\mathbf{x}_{it}^*) = h(\eta_{it}) \tag{0.4}$$

through a suitable link function h, for example the logistic function. The predictor η_{it} is modelled parametrically or nonparametrically as a function of time t and basic or derived covariates $\mathbf{x}_{it}$.

Parametric models

This section deals mainly with parametric models (0.4), where the predictor has the common linear parametric form

$$\eta_{it} = \mathbf{z}_{it}^{'}\boldsymbol{\beta} \tag{0.5}$$

as in generalized linear models*, with the design vector $\mathbf{z}_{it}$ formed from basic covariates. In many applications the linear predictor is chosen as

$$\eta_{it} = \beta_{0t} + \mathbf{x}_{it}^{'}\boldsymbol{\beta}_x\,, \tag{0.6}$$

where $\boldsymbol{\beta}_x$ is a vector of covariate effects and $\beta_{0t}, t = 1, \ldots, q$ is a time-varying baseline effect. Model (0.6) can be written in the form (0.5) by defining $\mathbf{z}_{it}^{'} = (0, \ldots, 1, \ldots, 0, \mathbf{x}_{it}^{'})$, $\boldsymbol{\beta}^{'} = (\beta_{01}, \ldots, \beta_{0q}, \boldsymbol{\beta}_x^{'})$. Other predictors are discussed further below.

Different discrete-time failure models are determined by choice of the link function h. Most common are discrete proportional hazards and logistic models.

The discrete proportional hazards model

Suppose that an underlying continuous failure time obeys a proportional hazard or relative risk model $\alpha_0(t)\exp(\mathbf{x}_{it}^{'}\boldsymbol{\beta}_x)$. If time T can only be observed as a discrete random variable, $T = t$ denoting failure in $[a_{t-1}, a_t)$, this yields the *discrete* proportional hazards model

$$\alpha(t|\mathbf{x}_{it}) = 1 - \exp[-\exp(\beta_{0t} + \mathbf{x}_{it}^{'}\boldsymbol{\beta}_x)]\,, \tag{0.7}$$

with baseline effects

$$\beta_{0t} = \log \int_{a_{t-1}}^{a_t} \alpha_0(t)dt$$

derived from the baseline function $\alpha_0(u)$ (see e.g. Kalbfleisch and Prentice [17]). An alternative formulation of (0.7) is the complementary log-log model* $\log\{-\log[1-\alpha(t|\mathbf{x}_{it}^{'})]\} = \beta_{0t} + \mathbf{x}_{it}^{'}\boldsymbol{\beta}_x$. The parameter vector $\boldsymbol{\beta}_x$ is unchanged by the transition to the discrete version, so that the same analysis as with the proportional hazard model is possible as far as the influence of covariates is concerned. However, $\boldsymbol{\beta}_x$ and time-varying effects β_{0t} have now to be estimated jointly. If the number of intervals is large, then the dimension of $\beta_{01}, \ldots, \beta_{0q}$ may become dangerously high, often even leading to nonexistence of ML estimates. Then more parsimonious parametric forms like polynomials $\beta_{0t} = \beta_0 + \ldots + \beta_k t^k$, piecewise constant effects or regression splines with only a few cut points are preferable. Often, cubic-linear splines of the form $\beta_{0t} = \beta_0 + \beta_1 t + \beta_2(t-t_c)_-^2 + \beta_3(t-t_c)_-^3$, are useful, where $(t-t_c)_- = \min(t-t_c, 0)$ and t_c is a cut-point. The baseline effect is cubic before t_c and linear after t_c. Such a simple spline model is more robust against few data at the end of the observation period than polynomials, and it is a smooth function, compared to piecewise constant modelling. Of course, other forms of regression splines may be considered. Also one may use the numerically more stable B-spline basis instead of the truncated power form, compare Sleeper and Harrington [20] in a continuous-time setting. By appropriate definition of the design vector, regression spline models can also be written in linear parametric from (0.5).

The logistic model

An alternative model is the logistic model for the discrete hazard

$$\alpha(t|\mathbf{x}_{it}) = \frac{\exp(\beta_{0t} + \mathbf{x}_{it}^{'}\boldsymbol{\beta}_x)}{1 + \exp(\beta_{0t} + \mathbf{x}_{it}^{'}\boldsymbol{\beta}_x)}\,, \qquad (0.8)$$

considered by Thompson [21] and, in slightly different form, by Cox [5]. For short intervals, this model becomes rather similar to the discrete proportional hazards model. An advantage of the logistic model is that the covariate effects $\boldsymbol{\beta}_x$ can be estimated semiparametrically, considering baseline effects β_{0t} as nuisance parameters and leaving them unspecified as in the continuous-time proportional hazards model, see Cox and Oakes [6].

Other discrete-time failure models result for other choices of h. Very flexible models are obtained if the link is an element of a parametric family of link functions. Examples are the model of Aranda-Ordaz (see Fahrmeir and Tutz [11], p. 318) or the families considered by Czado [7].

Although choice of the link function is an important issue, we feel that careful modelling of the predictor is often even more essential. To simplify discussion, we consider only two covariates x and w, where x is a continuous variable like tumor size or hormone concentration and w is binary, indicating for example sex or treatment group.

Models with *time-varying effects* are obtained by assuming

$$\eta_{it} = \beta_{0t} + \beta_1 x_i + \beta_{2t} w_i \,, \tag{0.9}$$

where β_{2t} could be the time-varying effect of a therapy, possibly decreasing with time. Alternatively, the term $\beta_{2t} w_i$ may be considered as a particular form of interaction between time t and the covariate w. The function β_{2t} may be modelled parametrically similarly as the baseline effect β_{0t}. A more detailed discussion of parametric time-varying effects is in Yamaguchi [24]. If the simple linear form $\beta_1 x_i$ for the influence of x is too restrictive, one may also try to replace it by a nonlinear smooth function $\beta_1(x)$ like in generalized additive models. As in Hastie and Tibshirani [15], one may go a step further and consider *varying coefficient models* of the form

$$\eta_{it} = \beta_{0t} + \beta_1(x_i) + \beta_2(x_i) w_i + \beta_{3t} w_i \,. \tag{0.10}$$

Here the smooth function β_2 may be viewed as an effect of w varying over x, or it is interpreted as an interaction term between the continuous covariate x and the binary covariate w. Without further prior knowledge it will often be difficult to specify certain parametric forms for the smooth functions in (0.9), (0.10). Instead, it will be reasonable to explore patterns with nonparametric approaches outlined in the next section and to proceed then with a simpler parametric likelihood-based inference.

Likelihood inference

Under appropriate conditions on censoring and covariate processes, the log-likelihood reduces to the common form known for binary regression models. Introducing the indicators

$$\mathbf{y}_i = (y_{i1}, \ldots, y_{it}) = \begin{cases} (0, \ldots, 0) & , \quad \delta_i = 0 \\ (0, \ldots, 0, 1) & , \quad \delta_i = 1 \end{cases}$$

it is proportional to

$$l(\boldsymbol{\beta}) = \sum_{i=1}^{n} \sum_{s=1}^{t_i} \{ y_{is} \log \alpha_i(s|\mathbf{x}_{is}^*) + (1 - y_{is}) \log[1 - \alpha_i(s|\mathbf{x}_{is}^*)] \} \,.$$

Arjas and Haara [3] give a careful discussion of assumptions leading to $l(\boldsymbol{\beta})$ as a (partial) log-likelihood. They will generally hold for noninformative random censoring and time-independent or external covariates, but can become critical for internal covariates. In particular, the likelihood is valid in the presence of ties, by making the weak assumption that failures at t are conditionally independent given covariates and past failures. By appropriate construction of design vectors $\mathbf{z}_{it}$, the parameters β can then be estimated with software for binary regression models, and other tools of likelihood inference for these models may be adopted, see Fahrmeir and Tutz [11], ch. 9.

Nonparametric approaches

Often, the common assumptions of linearity, additivity and time-constancy of effects are definitely violated and parametric specifications of more flexible models like (0.9) or (0.10) may be difficult. In this situation nonparametric approaches provide useful tools for detecting and exploring nonlinear or time-dependent effects. We outline the roughness penalty approach*, leading to spline-type smoothing and related Bayesian nonparametric techniques. Other methods are based on discrete kernels (e.g. Fahrmeir and Tutz [11], ch. 5, 9), or local likelihoods (Wu and Tuma [23], in a continuous-time setting, Tutz [22]). Consider models like (0.9) or (0.10) with unknown parameter vector $\boldsymbol{\beta}$ and unknown "smooth functions" $\beta_1, \beta_2, \ldots, \beta_q$ of time or continuous covariates. The roughness penalty approach maximizes a penalized log-likelihood criterion

$$pl(\boldsymbol{\beta}, \beta_1, \ldots, \beta_p) = l(\boldsymbol{\beta}, \beta_1, \ldots, \beta_p) - \sum_{j=1}^{p} \lambda_j J(\beta_j)\,,$$

where $J(\boldsymbol{\beta}_j)$ are roughness penalties and λ_j are smoothing parameters. A simple roughness penalty for a time-varying effect $\beta_{jt}, t = 1, \ldots, q$, is

$$J(\beta_j) = \sum_{s=2}^{q} \frac{(\beta_{jt} - \beta_{j,t-1})^2}{a_t - a_{t-1}}\,. \tag{0.11}$$

The same form may be used for a function $\beta_j(x)$ of some continuous covariate x. Another common penalty is

$$J(\beta_j) = \int [\beta_j(x)'']^2 dx$$

leading to cubic smoothing splines (see e.g. Green and Silverman [13]). Kiefer [18] proposes a discrete proportional hazards model with time-varying effects β_{jt} and penalty function (0.11). Dannegger, Klinger and Ulm [8] use the roughness penalty approach to explore nonlinear and time-varying effects of risk factors in a breast cancer study with monthly data. Related Bayesian nonparametric approaches put smoothness priors on β_{jt} or $\beta_j(x)$ and estimation is based on posteriors given the data. If, for example, a random walk of first order

$$\beta_{jt} = \beta_{j,t-1} + (a_t - a_{t-1})^{1/2} v_t \quad , \quad v_t \sim N(0, 1/\lambda_j)$$

is taken as smoothness prior for the sequence $\{\beta_{jt}\}$, then the posterior mode or MAP estimate is identical to the penalized likelihood estimate with penalty (0.11), see Fahrmeir [9]. Full posterior analyses can be carried out with MCMC* (Markov Chain Monte Carlo) techniques, see Fahrmeir and Knorr-Held [10] and, in the related context of generalized additive models, Biller and Fahrmeir [4]. Nonparametric methods are also useful for smoothing hazard functions in the absence of covariates. The smooth curve in Figure 1 is the posterior mean estimate obtained from a Bayesian MCMC approach. The corresponding cubic spline smoother is very close, see Green and Silverman [13].

Some extensions

More complex discrete-time event history data

Discrete failure time models can be extended similarly as continuous-time models. Often one may distinguish between several types $R \in \{1, \ldots, m\}$ of failure or terminating events. For example, in a medical study there may be several causes of death, or in studies on unemployment duration one may consider full time and part time jobs that end the unemployment duration. The basic quantities for *models with multiple modes of failure** are now cause-specific hazard functions

$$\alpha_{ir}(t|\mathbf{x}_{it}^*) = Pr(T = t, R = r|T \geq t, \mathbf{x}_{it}^*)\,, \qquad (0.12)$$

i.e. conditional probabilities for failure of type r in interval t. Multicategorical response models can be used for regression analysis of cause-specific hazard functions. A common candidate for unordered events is the multinomial logit model (e.g. Allison [1]). Other discrete choice models like a probit or a nested multinomial logit model (Hill et al. [16]) may also be considered. If events are ordered, ordinal response models (e.g. Fahrmeir and Tutz [11]) are appropriate. Again parametric and nonparametric approaches are possible. Penalized likelihood and Bayesian smoothing techniques with models for time-varying effects are applied to unemployment durations in Fahrmeir and Wagenpfeil [12], Fahrmeir and Knorr-Held [10].

Discrete failure time models can also be extended to general multiepisode-multistate models or, in counting process terminology, marked point processes. Hamerle [14] studies parametric regression analysis for such discrete event history data, but generally much less theoretical or applied work has been done here.

Unobserved heterogeneity and frailty models

The above model specifications assume that individual heterogeneity can be described by observed variables. However, it is likely that not all relevant variables are included in a regression model. The conventional approach to account for neglected heterogeneity or frailty is to include individual-specific parameters into the predictor, i.e. modifying η_{it} to $\eta_{it}^h = \eta_{it} + \theta_i$, and to assume that the individual-specific parameters are i.i.d. random variables from a prior density function f, i.e. a normal density. Estimation can then be based on approaches for generalized mixed models with random effects, and recent MCMC methods seem particularly well suited. However, for single-spell failure time models the estimates can be very dependent on the choice of the prior specification. More experience is needed here. The problem becomes less severe with repeated events.

References

[1] Allison, P. D. (1982). Discrete–time methods for the analysis of event histories. In *Sociological Methodology*, Vol.13, S. Leinhardt, ed., Jossey–Bass, San Francisco, pp. 61–98.

[2] Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (1993). *Statistical models based on counting processes.* Springer Verlag, New York.

[3] Arjas, E., and Haara, P. (1987). A logistic regression model for hazard: Asymptotic results. *Scandinavian Journal of Statistics* **14**, 1–18.

[4] Biller, C., and Fahrmeir, L. (1997). Bayesian spline-type smoothing in generalized regression models. To appear in *Computational Statistics.*

[5] Cox, D. R. (1972). Regression models and life tables. *Journal of the Royal Statistical Society* **B 34**, 187–220. (with discussion).

[6] Cox, D. R., and Oakes, D. (1984). *Analysis of survival data.* Chapman and Hall, London.

[7] Czado, C. (1992). On link selection in generalized linear models. *Springer Lecture Notes in Statistics* **78**, 60–65.

[8] Dannegger, F., Klinger, A., and Ulm, K. (1995). Identification of prognostic factors with censored data. Discussion Paper 11, Sonderforschungsbereich 386, Ludwig-Maximilians Universität München.

[9] Fahrmeir, L. (1994). Dynamic modelling and penalized likelihood estimation for discrete time survival data. *Biometrika* **81**, 317–330.

[10] Fahrmeir, L., and Knorr-Held, L. (1997). Dynamic discrete-time duration models: estimation via Markov Chain Monte Carlo. To appear in *Sociological Methodology.*

[11] Fahrmeir, L., and Tutz, G. (1994). *Multivariate Statistical Modelling Based on Generalized Linear Models.* Springer–Verlag, New York.

[12] Fahrmeir, L., and Wagenpfeil, S. (1996). Smoothing hazard functions and time-varying effects in discrete duration and competing risk models. To appear in *Journal of the American Statistical Association.*

[13] Green, P. J., and Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models.* Chapman and Hall, London.

[14] Hamerle, A. (1986). Regression analysis for discrete event history or failure time data. *Statistical Papers* **27**, 207–225.

[15] Hastie, T., and Tibshirani, R. (1993). Varying–coefficient models. *Journal of the Royal Statistical Society B* **55**, 757–796.

[16] Hill, D. H., Axinn, W. G., and Thornton, A. (1993). Competing hazards with shared unmeasured risk factors. In *Sociological Methodology*, Vol.23, P. V. Marsden, ed., Basil Blackwell, Oxford, pp. 245–277.

[17] Kalbfleisch, J. D., and Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data.* John Wiley & Sons, New York.

[18] Kiefer, N. M. (1990). Econometric methods for grouped duration data. In *Panel data and labor market studies*, J. Hartog, G. Ridder, and J. Theeuwes, eds., Elsevier Science Publisher, pp. 97–117.

[19] Scheike, T. H., and Jensen, T. K. (1997). A discrete survival model with random effects: An application to time to pregnancy. *Biometrics* **53**, 349–360.

[20] Sleeper, L. A., and Harrington, D. P. (1990). Regression splines in the Cox model with application to covariate effects in liver disease. *Journal of the American Statistical Association* **85**, 941–949.

[21] Thompson Jr., W. A. (1977). On the treatment of grouped observations in life studies. *Biometrics* **33**, 463–470.

[22] Tutz, G. (1995). Dynamic modelling of discrete duration data: a local likelihood approach. Report 95-15, Institut für Quantitative Methoden, Tech. Univ. Berlin.

[23] Wu, L. L., and Tuma, N. B. (1991). Assessing bias and fit of global and local hazard models. *Sociological Methods and Research* **19**, 354–387.

[24] Yamaguchi, K. (1993). Modeling time–varying effects of covariates in event–history analysis using statistics from the saturated hazard rate model. In *Sociological Methodology*, Vol.23, P. V. Marsden, ed., Basil Blackwell, Oxford, pp. 279–317.