A.1.4 Weibull regression in survival analysis

Model description A typical setting in survival analysis is that we observe the time point t at which the death of a patient occurs. Patients may leave the study (for some reason) before they die. In this case the survival time is said to be censored, and t refers to the time point when the patient left the study. The indicator variable δ is used to indicate whether t refers to the death of the patient ($\delta = 1$) or to a censoring event ($\delta = 0$). The key quantity in modelling the probability distribution of t is the hazard function $h(t)$, which measures the instantaneous death rate at time t. We also define the cumulative hazard function $\Lambda(t) = \int_0^t h(s)ds$, implicitly assuming that the study started at time $t = 0$. The log likelihood contribution from our patient is $\delta \log(h(t)) - H(t)$. A commonly used model for $h(t)$ is Cox's proportional hazard model, in which the hazard rate for the ith patient is assumed to be on the form

$$h_i t = h_0(t) \exp(\eta_i), \qquad i = 1, \dots n.$$

Here, $h_0(t)$ is the "baseline" hazard function (common to all patients) and $\eta_i = \mathbf{X}_i \beta$, where $\mathbf{X}_i$ is a covariate vector specific to the ith patient and β is a vector of regression parameters. In this example we shall assume that the baseline hazard belongs to the Weibull family: $h_0(t) = rt^{r-1}$ for $r > 0$.

In the collection of examples following the distribution of WinBUGS this model is used to analyse a dataset on times to kidney infection for a set of $n = 38$ patients (Kidney: Weibull regression with random effects, Examples Volume 1, WinBUGS 1.4). The dataset contains two observations per patient (the time to first and second recurrence of infection). In addition there are three covariates: *age* (continuous), *sex* (dichotomous) and *type of disease* (categorical, four levels), and an individual-specific random effect $u_i \sim N(0, \sigma^2)$. Thus, the linear predictor becomes

$$\eta_i = \beta_0 + \beta_{\text{sex}} \cdot \text{sex}_i + \beta_{\text{age}} \cdot \text{age}_i + \beta_{\text{D}}\, \mathbf{x}_i + u_i,$$

where $\beta_{\text{D}} = (\beta_1, \beta_2, \beta_3)$ and $\mathbf{x}_i$ is a dummy vector coding for the disease type. Parameter estimates are shown in the table below.

	β_0	β_{age}	β_1	β_2	β_3	β_{sex}	r	σ
ADMB-RE	-4.3440	0.0030	0.1208	0.6058	-1.1423	-1.8767	1.1624	0.5617
Std. dev.	0.8720	0.0137	0.5008	0.5011	0.7729	0.4754	0.1626	0.2970
BUGS	-4.6000	0.0030	0.1329	0.6444	-1.1680	-1.9380	1.2150	0.6374
Std. dev.	0.8962	0.0148	0.5393	0.5301	0.8335	0.4854	0.1623	0.3570

Files http://otter-rsch.com/admbre/examples/kidney/kidney.html

Bibliography

ADMB Development Core Team (2009), *An Introduction to AD Model Builder*, ADMB project.

ADMB Foundation (2009), 'ADMB-IDE: Easy and efficient user interface', *ADMB Foundation Newsletter* **1**, 1–2.

Eilers, P. & Marx, B. (1996), 'Flexible smoothing with B-splines and penalties', *Statistical Science* **89**, 89–121.

Harvey, A., Ruiz, E. & Shephard, N. (1994), 'Multivariate stochastic variance models', *Review of Economic Studies* **61**, 247–264.

Hastie, T. & Tibshirani, R. (1990), *Generalized Additive Models*, Vol. 43 of *Monographs on Statistics and Applied Probability*, Chapman & Hall, London.

Kuk, A. Y. C. & Cheng, Y. W. (1999), 'Pointwise and functional approximations in Monte Carlo maximum likelihood estimation', *Statistics and Computing* **9**, 91–99.

Lin, X. & Zhang, D. (1999), 'Inference in generalized additive mixed models by using smoothing splines', *J. Roy. Statist. Soc. Ser. B* **61**(2), 381–400.

Pinheiro, J. C. & Bates, D. M. (2000), *Mixed-Effects Models in S and S-PLUS*, Statistics and Computing, Springer.

Rue, H. & Held, L. (2005), *Gaussian Markov random fields: theory and applications*, Chapman & Hall/CRC.

Ruppert, D., Wand, M. & Carroll, R. (2003), *Semiparametric Regression*, Cambridge University Press.

Skaug, H. & Fournier, D. (2006), 'Automatic approximation of the marginal likelihood in non-gaussian hierarchical models', *Computational Statistics & Data Analysis* **56**, 699–709.

Zeger, S. L. (1988), 'A regression-model for time-series of counts', *Biometrika* **75**, 621–629.