

3.8 Weibull regression in censored 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 $H(t) = \int_0^t h(s)ds$, implicitly assuming that the study started at time $t = 0$. The loglikelihood 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 i th patient is assumed to be on the form

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

Here, $h_0(t)$ is the “baseline” hazard function (common to all patients) and η_i is a linear predictor. 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 analyze 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

$$\eta_i = \beta_0 + \beta_{\text{sex}} \cdot \text{sex}_i + \beta_{\text{age}} \cdot \text{age}_i + \beta_{\text{D}} \mathbf{x}_i + u_i, \quad (3.3)$$

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. Posterior means as calculated by WinBUGS are also shown in the table, and are similar to the maximum likelihood estimates.

	β_0	β_{age}	β_1	β_2	β_3	β_{sex}	r	σ
ADMB-RE	-4.344	0.003	0.1208	0.6058	-1.1423	-1.8767	1.1624	0.5617
Std. dev.	0.872	0.0137	0.5008	0.5011	0.7729	0.4754	0.1626	0.297
BUGS	-4.6	0.003	0.1329	0.6444	-1.168	-1.938	1.215	0.6374
Std. dev.	0.8962	0.0148	0.5393	0.5301	0.8335	0.4854	0.1623	0.357