

ment, Edmunson *et al.* (1979) compared the anti-tumour effects of cyclophosphamide alone and cyclophosphamide combined with adriamycin. The trial involved 26 women with minimal residual disease and who had experienced surgical excision of all tumour masses greater than 2 cm in diameter. Following surgery, the patients were further classified according to whether the residual disease was completely or partially excised. The age of the patient and their performance status were also recorded at the start of the trial. The response variable was the survival time in days following randomisation to one or other of the two chemotherapy treatments. The variables in the data set are therefore as follows:

- Time*: Survival time in days,  
*Status*: Event indicator (0 = censored, 1 = uncensored),  
*Treat*: Treatment (1 = single, 2 = combined),  
*Age*: Age of patient in years,  
*Rdisease*: Extent of residual disease (1 = incomplete, 2 = complete),  
*Perf*: Performance status (1 = good, 2 = poor).

The data, which were obtained from Therneau (1986), are given in Table 5.6.

In modelling these data, the factors *Treat*, *Rdisease* and *Perf* each have two levels, and will be fitted as variates that take the values given in Table 5.6. This does of course mean that the baseline hazard function is not directly interpretable, since there can be no individual for whom the values of all these variates are zero. From both a computational and interpretive viewpoint, it is more convenient to relocate the values of the variables *Age*, *Rdisease*, *Perf* and *Treat*. If the variable  $\text{Age} - 50$  is used in place of *Age*, and unity is subtracted from *Rdisease*, *Perf* and *Treat*, the baseline hazard then corresponds to the hazard for an individual of age 50 with incomplete residual disease, good performance status, and who has been allocated to the cyclophosphamide group. However, the original variables will be used in this example.

We begin by identifying which prognostic factors are associated with the survival times of the patients. The values of the statistic  $-2 \log \hat{L}$  on fitting a range of models to these data are given in Table 5.7.

When Weibull models that contain just one of *Age*, *Rdisease* and *Perf* are fitted, we find that both *Age* and *Rdisease* lead to reductions in the value of  $-2 \log \hat{L}$  that are significant at the 5% level. After fitting *Age*, the variables *Rdisease* and *Perf* further reduce  $-2 \log \hat{L}$  by 1.903 and 0.048, respectively, neither of which is significant at the 10% level. Also, when *Age* is added to the model that already includes *Rdisease*, the reduction in  $-2 \log \hat{L}$  is 13.719 on 1 d.f., which is highly significant ( $P < 0.001$ ). This leads us to the conclusion that *Age* is the only prognostic variable that needs to be incorporated in the model.

The term associated with the treatment effect is now added to the model. The value of  $-2 \log \hat{L}$  is then reduced by 2.440 on 1 d.f. This reduction of 2.440 is not quite large enough for it to be significant at the 10% level ( $P = 0.118$ ).

Table 5.6 Survival times of ovarian cancer patients.

Patient	Time	Status	Treat	Age	Rdisease	Perf
1	156	1	1	66	2	2
2	1040	0	1	38	2	2
3	59	1	1	72	2	1
4	421	0	2	53	2	1
5	329	1	1	43	2	1
6	769	0	2	59	2	2
7	365	1	2	64	2	1
8	770	0	2	57	2	1
9	1227	0	2	59	1	2
10	268	1	1	74	2	2
11	475	1	2	59	2	2
12	1129	0	2	53	1	1
13	464	1	2	56	2	2
14	1206	0	2	44	2	1
15	638	1	1	56	1	2
16	563	1	2	55	1	2
17	1106	0	1	44	1	1
18	431	1	1	50	2	1
19	855	0	1	43	1	2
20	803	0	1	39	1	1
21	115	1	1	74	2	1
22	744	0	2	50	1	1
23	477	0	1	64	2	1
24	448	0	1	56	1	2
25	353	1	2	63	1	2
26	377	0	2	58	1	1

Table 5.7 Values of  $-2 \log \hat{L}$  on fitting models to the data in Table 5.6.

Variables in model	$-2 \log \hat{L}$
none	59.534
<i>Age</i>	43.566
<i>Rdisease</i>	55.382
<i>Perf</i>	58.849
<i>Age</i> , <i>Rdisease</i>	41.663
<i>Age</i> , <i>Perf</i>	43.518
<i>Age</i> , <i>Treat</i>	41.126
<i>Age</i> , <i>Treat</i> , <i>Treat</i> $\times$ <i>Age</i>	39.708

Collett Modelling Survival Data  
2nd Edition

There is therefore only very slight evidence of a difference in the effect of the two chemotherapy treatments on the hazard of death.

For comparison, when *Treat* alone is added to the null model, the value of  $-2 \log \hat{L}$  is reduced from 59.534 to 58.355. This reduction of 1.179 is certainly not significant when compared to percentage points of the chi-squared distribution on 1 d.f. Ignoring *Age* therefore leads to an underestimate of the magnitude of the treatment effect.

To explore whether the treatment difference is consistent over age, the interaction term formed as the product of *Age* and *Treat* is added to the model. On doing so,  $-2 \log \hat{L}$  is only reduced by 1.419. This reduction is nowhere near being significant and so there is no need to include an interaction term in the model.

The variable *Treat* will be retained in the model, since interest centres on the magnitude of the treatment effect. The fitted model for the hazard of death at time  $t$  for the  $i$ th individual is then found to be

$$\hat{h}_i(t) = \exp\{0.144 \text{ Age}_i - 1.023 \text{ Treat}_i\} \hat{\lambda} \hat{\gamma} t^{\hat{\gamma}-1},$$

where  $\hat{\lambda} = 5.645 \times 10^{-9}$  and  $\hat{\gamma} = 1.822$ . In this model, *Treat* = 1 for cyclophosphamide alone and *Treat* = 2 for the combination of cyclophosphamide with adriamycin. The hazard for a patient on the single treatment, relative to one on the combined treatment, is therefore estimated by

$$\hat{\psi} = \exp\{(-1.023 \times 1) - (-1.023 \times 2)\} = 2.78.$$

This means that a patient receiving the single chemotherapy treatment is nearly three times more likely to die at any given time than a patient on the combined treatment. Expressed in this way, the benefits of the combined chemotherapy treatment sound to be great. However, when account is taken of the inherent variability of the data on which these results are based, this relative hazard is only significantly greater than unity at the 12% level ( $P = 0.118$ ).

The median survival time can be estimated for patients of a given age on a given treatment from the equation

$$\hat{t}(50) = \left\{ \frac{\log 2}{\hat{\lambda} \exp(0.144 \text{ Age} - 1.023 \text{ Treat})} \right\}^{1/\hat{\gamma}}.$$

For example, a woman aged 60 (*Age* = 60) who is given cyclophosphamide alone (*Treat* = 1) has an estimated median survival time of 423 days, whereas someone of the same age on the combination of the two chemotherapy treatments has an estimated median survival time of 741 days. Confidence intervals for these estimates can be found using the method illustrated in Example 5.6.

### 5.7 The Gompertz proportional hazards model

Although the Weibull model is the most widely used parametric proportional hazards model, the Gompertz model has found application in demography and

the biological sciences. Indeed the distribution was introduced by Gompertz in 1825, as a model for human mortality.

The hazard function of the Gompertz distribution is given by

$$h(t) = \lambda e^{\theta t},$$

for  $0 \leq t < \infty$ , and  $\lambda > 0$ . In the particular case where  $\theta = 0$ , the hazard function has a constant value,  $\lambda$ , and the survival times then have an exponential distribution. The parameter  $\theta$  determines the shape of the hazard function, positive values leading to a hazard function that increases with time. The hazard function can also be expressed as  $h(t) = \exp(\alpha + \theta t)$ , which shows that the log-hazard function is linear in  $t$ . On the other hand, from equation (5.7), the Weibull log-hazard function is linear in  $\log t$ . Like the Weibull hazard function, the Gompertz hazard increases or decreases monotonically.

The survivor function of the Gompertz distribution is given by

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

and the corresponding density function is

$$f(t) = \lambda e^{\theta t} \exp \left\{ \frac{\lambda}{\theta} (1 - e^{\theta t}) \right\}.$$

The  $p$ th percentile is such that

$$t(p) = \frac{1}{\theta} \log \left\{ 1 - \frac{\theta}{\lambda} \log \left( \frac{100 - p}{100} \right) \right\},$$

from which the median survival time is

$$t(50) = \frac{1}{\theta} \log \left\{ 1 + \frac{\theta}{\lambda} \log 2 \right\}.$$

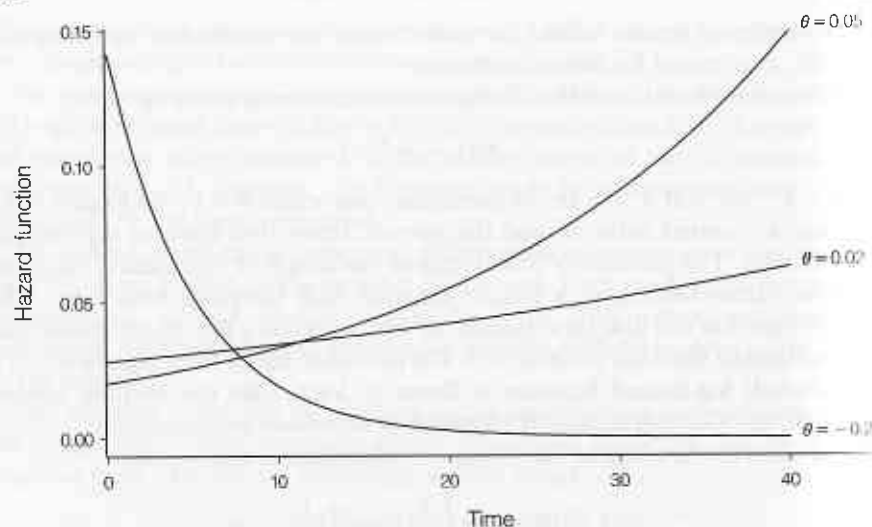
A plot of the Gompertz hazard function for distributions with a median of 20 and  $\theta = -0.2, 0.02$  and  $0.05$  is shown in Figure 5.15. The corresponding values of  $\lambda$  are 0.141, 0.028, and 0.020.

It is straightforward to see that the Gompertz distribution has the proportional hazards property, described in Section 5.4, since if we take  $h_0(t) = \lambda e^{\theta t}$ , then  $\psi h_0(t)$  is also a Gompertz hazard function with parameters  $\psi\lambda$  and  $\theta$ .

The general Gompertz proportional hazards model, for the hazard of death at time  $t$  for the  $i$ th of  $n$  individuals, is expressed as

$$h_i(t) = \exp(\beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi}) \lambda e^{\theta t},$$

where  $x_{1i}, x_{2i}, \dots, x_{pi}$  are the values of  $p$  explanatory variables  $X_1, X_2, \dots, X_p$  for the  $i$ th individual,  $i = 1, 2, \dots, n$ , and the  $\beta$ 's,  $\lambda$  and  $\theta$  are unknown parameters. The model can be fitted by maximising the likelihood function given in expression (5.12) or (5.13). The  $\beta$ -coefficients are interpreted as log-hazard ratios, and alternative models are compared using the approach described in Section 5.6. No new principles are involved.



**Figure 5.15** Hazard functions for a Gompertz distribution with a median of 20 and  $\theta = -0.2, 0.02$  and  $0.05$ .

#### Example 5.11 Chemotherapy in ovarian cancer patients

In Example 5.10 on the survival times of ovarian cancer patients, a Weibull proportional hazards model that contained the variables *Age* and *Treat* was fitted. For comparison, a Gompertz proportional hazards model that contains these two variables is now fitted. Under this model, the fitted hazard function for the  $i$ th patient is

$$\hat{h}_i(t) = \exp\{0.122 \text{ Age}_i - 0.848 \text{ Treat}_i\} \hat{\lambda} \exp(\hat{\theta}t),$$

where  $\hat{\lambda} = 1.706 \times 10^{-6}$  and  $\hat{\theta} = 0.00138$ . The change in the value of  $-2 \log L$  on adding *Treat* to the Gompertz proportional hazards model that contains *Age* alone is now 1.686 ( $P = 0.184$ ). The hazard ratio for the treatment effect, which is now  $\exp(0.848) = 2.34$ , is therefore smaller and less significant under this model than it was for the Weibull model.

### 5.8 Model choice

One attraction of the proportional hazards model for survival data is that it is not necessary to adopt a specific probability distribution for the survival times. However, when a Weibull distribution is appropriate for the observed survival data, the parametric version of the proportional hazards model provides a more suitable basis for modelling the data.

Diagnostic plots based on the log-cumulative hazard function, described in Section 5.4.1, may throw light on whether the assumption of Weibull survival times is plausible, but as has already been pointed out, this technique is often not informative in the presence of explanatory variables that affect survival

times. In such circumstances, to help choose between the Cox and Weibull proportional hazards models, it can be useful to fit the Cox regression model and examine the shape of the baseline hazard function. The fitted Weibull baseline cumulative hazard function, or the fitted baseline survivor function, can also be compared with the corresponding estimates for the Cox regression model, as described in Section 5.5.3.

A suitable analysis of residuals, to be discussed in Chapter 7, can be used to investigate whether one model fits better than the other. However, it will only be in exceptional circumstances that model-checking diagnostics provide convincing evidence that one or other of the two models is more acceptable.

In general, discrimination between a Cox and a Weibull proportional hazards model will be difficult unless the sample data contain a large number of death times. In cases where there is little to choose between the two models in terms of goodness of fit, the standard errors of the estimated  $\beta$ -parameters in the linear component of the two models can be compared. If those for the Weibull model are substantially smaller than those for the Cox model, the Weibull model would be preferred on grounds of efficiency. On the other hand, if these standard errors are similar, the Cox model is likely to be the model of choice in view of its less restrictive assumptions.

### 5.9 Further reading

The properties of the exponential, Weibull and Gompertz distributions are presented in Johnson and Kotz (1970). A thorough discussion of the theory of maximum likelihood estimation is included in Barnett (1999) and Cox and Hinkley (1974), and a useful summary of the main results is contained in Hinkley, Reid and Snell (1991). Numerical methods for obtaining maximum likelihood estimates, and the Newton-Raphson procedure in particular, are described by Everitt (1987) and Thisted (1988), for example; see also the description in Section 3.3.3 of Chapter 3. Byar (1982) presents a comparison of the Cox and Weibull proportional hazards models. One other distribution with the proportional hazards property is the *Pareto distribution*. This model is rarely used in practice, but see Davis and Feldstein (1979) for further details.