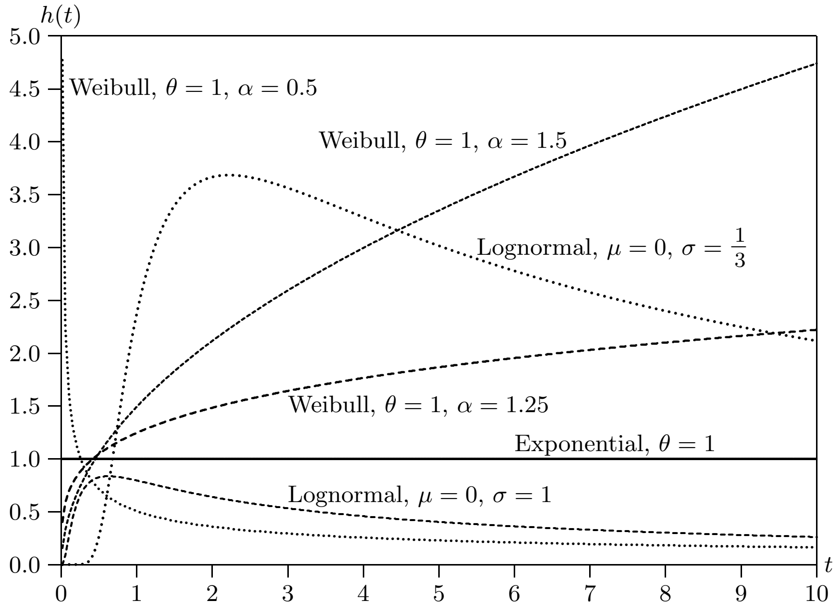

Figure 11.4 Various hazard functions

Maximum Likelihood Estimation

It is reasonably straightforward to estimate many duration models by maximum likelihood. In the simplest case, the data consist of n observations t_i on observed durations, each with an associated regressor vector $\boldsymbol{X}_i$. Then the loglikelihood function for $\boldsymbol{t}$, the entire vector of observations, is just

$$\ell(\boldsymbol{t}, \boldsymbol{\theta}) = \sum_{i=1}^{n} \log f(t_i \,|\, \boldsymbol{X}_i, \boldsymbol{\theta}), \tag{11.84}$$

where $f(t_i \,|\, \boldsymbol{X}_i, \boldsymbol{\theta})$ denotes the density of t_i conditional on the data vector $\boldsymbol{X}_i$ for the parameter vector $\boldsymbol{\theta}$. In many cases, it may be easier to write the loglikelihood function as

$$\ell(\boldsymbol{t}, \boldsymbol{\theta}) = \sum_{i=1}^{n} \log h(t_i \,|\, \boldsymbol{X}_i, \boldsymbol{\theta}) + \sum_{i=1}^{n} \log S(t_i \,|\, \boldsymbol{X}_i, \boldsymbol{\theta}), \tag{11.85}$$

where $h(t_i \,|\, \boldsymbol{X}_i, \boldsymbol{\theta})$ is the hazard function and $S(t_i \,|\, \boldsymbol{X}_i, \boldsymbol{\theta})$ is the survivor function. The equivalence of (11.84) and (11.85) is ensured by (11.81), in which the hazard function was defined.

As with other models we have looked at in this chapter, it is convenient to let the loglikelihood depend on explanatory variables through an index function. As an example, suppose that duration follows a Weibull distribution, with a parameter θ_i for observation i that has the form of the exponential mean function (11.48), so that $\theta_i = \exp(\boldsymbol{X}_i\boldsymbol{\beta}) > 0$. From (11.83) we see that the hazard and survivor functions for observation i are

$$\alpha \exp(\alpha \boldsymbol{X}_i\boldsymbol{\beta})t^{\alpha-1} \quad \text{and} \quad \exp\bigl(-t^{\alpha}\exp(\alpha\boldsymbol{X}_i\boldsymbol{\beta})\bigr),$$

respectively. In practice, it is simpler to absorb the factor of α into the parameter vector $\boldsymbol{\beta}$, so as to yield an exponent of just $\boldsymbol{X}_i\boldsymbol{\beta}$ in these expressions. Then the loglikelihood function (11.85) becomes

$$\ell(\boldsymbol{t},\boldsymbol{\beta},\alpha) = n\log\alpha + \sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{\beta} + (\alpha-1)\sum_{i=1}^{n}\log t_i - \sum_{i=1}^{n} t_i^{\alpha}\exp(\boldsymbol{X}_i\boldsymbol{\beta}),$$

and ML estimates of the parameters α and $\boldsymbol{\beta}$ are obtained by maximizing this function in the usual way.

In practice, many data sets contain observations for which t_i is not actually observed. For example, if we have a sample of people who entered unemployment at various points in time, it is extremely likely that some people in the sample were still unemployed when data collection ended. If we omit such observations, we are effectively using a truncated data set, and we therefore obtain inconsistent estimates. However, if we include them but treat the observed t_i as if they were the lengths of completed spells of unemployment, we also obtain inconsistent estimates. In both cases, the inconsistency occurs for essentially the same reasons as it does when we apply OLS to a sample that has been truncated or censored; see Section 11.6.

If we are using ML estimation, it is easy enough to deal with duration data that have been censored in this way, provided we know that censorship has occurred. For ordinary, uncensored observations, the contribution to the loglikelihood function is a contribution like those in (11.84) or (11.85). For censored observations, where the observed t_i is the duration of an **incomplete spell**, it is the logarithm of the probability of censoring, which is the probability that the duration exceeds t_i, that is, the log of the survivor function. Therefore, if U denotes the set of uncensored observations, the loglikelihood function for the entire sample can be written as

$$\ell(\boldsymbol{t},\boldsymbol{\theta}) = \sum_{i\in U}\log h(t_i \,|\, \boldsymbol{X}_i,\boldsymbol{\theta}) + \sum_{i=1}^{n}\log S(t_i \,|\, \boldsymbol{X}_i,\boldsymbol{\theta}). \tag{11.86}$$

Notice that uncensored observations contribute to both terms in equation (11.86), while censored observations contribute only to the second term. When there is no censoring, the same observations contribute to both terms, and the loglikelihood function (11.86) reduces to (11.85).