Observer Arrival Time on Days Where the Descent Time Was Not Observed

Day	Date	Arrival Time	Day	Date	Arrival Time	Day	Date	Arrival Time
1	1/12/63	0705	32	13/10/63	0840	63	2/5/64	1012
2	6/11/63	0710	33	4/7/64	0845	64	1/3/64	1018
3	24/10/63	0715	34	3/5/64	0850	65	17/10/63	1020
4	26/11/63	0720	35	25/5/64	0851	66	23/10/63	1020
5	18/10/63	0720	36	24/11/63	0853	67	25/7/64	1020
6	7/5/64	0730	37	15/7/64	0855	68	13/7/64	1031
7	7/11/63	0740	38	16/2/64	0856	69	8/6/64	1050
8	23/11/63	0750	39	10/3/64	0857	70	9/3/64	1050
9	28/11/63	0750	40	28/7/64	0858	71	26/4/64	1100
10	27/11/63	0753	41	18/6/64	0858	72	14/10/63	1205
11	28/5/64	0755	42	20/2/64	0858	73	18/11/63	1245
12	5/7/64	0757	43	2/8/64	0859	74	2/3/64	1250
13	28/3/64	0800	44	27/5/64	0900	75	8/5/64	1405
14	23/3/64	0805	45	28/10/64	0905	76	1/7/64	1407
15	26/10/63	0805	46	15/5/64	0907	77	12/10/63	1500
16	11/7/64	0805	47	10/5/64	0908	78	31/7/64	1531
17	27/7/64	0807	48	27/6/64	0915	79	6/10/63	1535
18	9/6/64	0810	49	11/10/63	0915	80	19/6/64	1556
19	24/6/64	0812	50	17/2/64	0920	81	29/6/64	1603
20	16/ 10/63	0812	51	22/10/63	0920	82	9/5/64	1605
21	25/2/64	0813	52	10/7/64	0925	83	9/10/63	1625
22	6/6/64	0814	53	14/7/64	0926	84	8/3/64	1625
23	22/11/63	0815	54	11/4/64	0931	85	11/2/64	1653
24	10/10/63	0815	55	23/5/64	0933	86	30/5/64	1705
25	2/11/63	0815	56	30/7/64	0943	87	5/3/64	1708
26	23/6/64	0817	57	18/7/64	0945	88	26/2/64	1722
27	24/4/64	0823	58	29/7/64	0946	89	4/5/64	1728
28	3/7/64	0830	59	16/7/64	0950	90	12/3/64	1730
29	29/4/64	0831	60	22/7/64	0955	91	25/10/63	1730
30	4/8/63	0838	61	15/10/63	0955	92	29/11/63	1750
31	7/10/63	0840	62	19/10/63	1005	93	22/2/64	1801
						94	22/3/64	1829

6 Topics in Univariate Estimation

6.1 Introduction

In Chapter 4, we presented two techniques for providing summary curves which tell us about the survival experience of a cohort of individuals. These two estimators were the Kaplan–Meier estimator, which provides an estimate of the survival function, and the Nelson–Aalan estimator, which provides an estimate of the cumulative hazard rate. These statistics are readily available in many statistical packages.

Although these two statistics provide an investigator with important information about the eventual death time of an individual, they provide only limited information about the mechanism of the process under study, as summarized by the hazard rate. The slope of the Nelson–Aalan estimator provides a crude estimate of the hazard rate, but this estimate is often hard to interpret. In section 6.2, we discuss how these crude estimates of the hazard rate can be smoothed to provide a better estimator of the hazard rate by using a kernel-smoothing technique.

In some applications of survival analysis, an investigator has available very precise information about the mortality rates in a historical control or standard population. It is of interest to compare the hazard rates in the sample group to the known hazard rates in the reference population to determine how the mortality experience of the experimental subjects differs. The "excess" mortality in the experimental group can

have either a multiplicative or additive effect on the reference hazard rate. In section 6.3, estimation techniques for both the additive and multiplicative models for excess mortality are developed.

In section 6.4, the problem of estimation of the survival function for right censored data is considered from a Bayesian perspective. In this framework, an investigator has some prior information on the survival function from results of similar studies, from a group of experts, or from some reference population. The prior information is combined with sample data to provide a posterior distribution of the survival function on which the estimation is based. The combination of prior and sample information can be done analytically by Bayes theorem or by a Monte Carlo method via the Gibbs sampler. Both methods are illustrated.

6.2 Estimating the Hazard Function

The Nelson–Aalen estimator $\tilde{H}(t)$, discussed in sections 4.2 or 4.6, provides an efficient means of estimating the cumulative hazard function $H(t)$. In most applications, the parameter of interest is not $H(t)$, but rather its derivative $h(t)$, the hazard rate. As noted earlier, the slope of the Nelson–Aalen estimator provides a crude estimate of the hazard rate $h(t)$. Several techniques have been proposed in the literature to estimate $h(t)$. In this section, we shall concentrate on the use of kernel smoothing to estimate $h(t)$.

Kernel-smoothed estimators of $h(t)$ are based on the Nelson–Aalen estimator $\tilde{H}(t)$ and its variance $\hat{V}[\tilde{H}(t)]$. The estimator $\tilde{H}(t)$ can be based on right-censored data (see section 4.2) or on left-truncated data (see section 4.6). Recall that, in either case, $\tilde{H}(t)$ is a step function with jumps at the event times, $0 = t_0 < t_1 < t_2 < \cdots < t_D$. Let $\Delta\tilde{H}(t_i) = \tilde{H}(t_i) - \tilde{H}(t_{i-1})$ and $\Delta\hat{V}[\tilde{H}(t_i)] = \hat{V}[\tilde{H}(t_i)] - \hat{V}[\tilde{H}(t_{i-1})]$ denote the magnitude of the jumps in $\tilde{H}(t_i)$ and $\hat{V}[\tilde{H}(t_i)]$ at time t_i. Note that $\Delta\tilde{H}(t_i)$ provides a crude estimator of $h(t)$ at the death times. The kernel-smoothed estimator of $h(t)$ is a weighted average of these crude estimates over event times close to t. Closeness is determined by a bandwidth b, so that event times in the range $t - b$ to $t + b$ are included in the weighted average which estimates $h(t)$. The bandwidth is chosen either to minimize some measure of the mean-squared error or to give a desired degree of smoothness, as illustrated in Example 6.2. The weights are controlled by the choice of a kernel function, $K()$, defined on the interval $[-1, +1]$, which determines how much weight is given to points at a distance from t. Common choices for the kernel

are the uniform kernel with

$$K(x) = 1/2 \quad \text{for } -1 \le x \le 1, \tag{6.2.1}$$

the Epanechnikov kernel with

$$K(x) = 0.75(1 - x^2) \quad \text{for } -1 \le x \le 1, \tag{6.2.2}$$

and the biweight kernel with

$$K(x) = \frac{15}{16}(1 - x^2)^2 \quad \text{for } -1 \le x \le 1. \tag{6.2.3}$$

The uniform kernel gives equal weight to all deaths in the interval $t - b$ to $t + b$, whereas the other two kernels give progressively heavier weight to points close to t.

The kernel-smoothed hazard rate estimator is defined for all time points $t > 0$. For time points t for which $b \le t \le t_D - b$, the kernel-smoothed estimator of $h(t)$ based on the kernel $K()$ is given by

$$\hat{h}(t) = b^{-1}\sum_{i=1}^{D} K\left(\frac{t - t_i}{b}\right)\Delta\tilde{H}(t_i). \tag{6.2.4}$$

The variance of $\hat{h}(t)$ is estimated by the quantity

$$\sigma^2[\hat{h}(t)] = b^{-2}\sum_{i=1}^{D} K\left(\frac{t - t_i}{b}\right)^2 \Delta\hat{V}[\tilde{H}(t_i)]. \tag{6.2.5}$$

When t is smaller than b, the symmetric kernels described in (6.2.1)–(6.2.3) are not appropriate because no event times less than 0 are observable. In this region, the use of an asymmetric kernel is suggested. Let $q = t/b$. We define a modified kernel which accounts for the restricted range of the data. Following Gasser and Müller (1979) these modified kernels, for the uniform kernel (6.2.1), are expressed by

$$K_q(x) = \frac{4(1 + q^3)}{(1 + q)^4} + \frac{6(1 - q)}{(1 + q)^3}x, \quad \text{for } -1 \le x \le q, \tag{6.2.6}$$

for the Epanechnikov kernel (6.2.2),

$$K_q(x) = K(x)(\alpha_E + \beta_E x), \quad \text{for } -1 \le x \le q, \tag{6.2.7}$$

where

$$\alpha_E = \frac{64(2 - 4q + 6q^2 - 3q^3)}{(1 + q)^4(19 - 18q + 3q^2)}$$

and

$$\beta_E = \frac{240(1 - q)^2}{(1 + q)^4(19 - 18q + 3q^2)},$$

and for the biweight kernel (6.2.3),

$$K_q(x) = K(x)(\alpha_{BW} + \beta_{BW}x), \quad \text{for } -1 \le x \le q, \tag{6.2.8}$$

where

$$\alpha_{BW} = \frac{64(8 - 24q + 48q^2 - 45q^3 + 15q^4)}{(1+q)^5(81 - 168q + 126q^2 - 40q^3 + 5q^4)}$$

and

$$\beta_{BW} = \frac{1120(1-q)^3}{(1+q)^5(81 - 168q + 126q^2 - 40q^3 + 5q^4)}.$$

For time points in the right-hand tail ($t_D - b < t < t_D$) let $q = (t_D - t)/b$. The asymmetric kernel $K_q(x)$ in (6.2.6)–(6.2.8) is used with x replaced by $-x$. The estimated, smoothed, hazard rate and its variance are given by (6.2.4) and (6.2.5), respectively, using the kernel K_q.

Confidence intervals or confidence bands for the hazard rate, based on the smoothed hazard rate estimate, can be constructed similarly to those for the cumulative hazard rate discussed in Chapter 4. For example a $(1-\alpha)\times 100\%$ pointwise confidence interval for the hazard rate, based on a log transformation, is expressed as

$$\hat{h}(t) \exp\left[\pm \frac{Z_{1-\alpha/2}\sigma(\hat{h}(t))}{\hat{h}(t)}\right].$$

Some care in interpreting this interval must be taken because the estimator $\hat{h}(t)$ may be quite biased (See Practical Note 1).

EXAMPLE 6.1 We shall find the smoothed hazard rate estimates, in the three disease categories, for the disease-free survival times of bone marrow transplant patients discussed in section 1.3. To illustrate the calculations, consider the group of ALL patients. In Example 4.2 the Nelson–Aalen estimator of the cumulative hazard rate of the disease-free survival time was found (see Table 4.3). For illustrative purposes, we shall use the Epanechnikov kernel with a bandwidth of 100 days. An estimate of $h(t)$ over the first two years (730 days) after transplant is desired.

Table 6.1 shows some of the calculations needed to construct the estimate. First, consider the estimate at $t = 150$ days. Here, t is in the interval b to $t_D - b$ (662–100), so that the symmetric kernel (6.2.2) is used. The estimate of the hazard rate is given by $\hat{h}(150) = [0.0270\times 0.0731 + 0.0278\times 0.3168 + 0.0286\times 0.4428 + 0.0294\times 0.5913 + 0.0303\times 0.6113 + 0.0313\times 0.6239 + 0.0322\times 0.6300 + 0.0667\times 0.6912 + 0.0357\times 0.7169 + 0.0370\times 0.7137 + 0.0385\times 0.6177 + 0.0400\times 0.6048 + 0.0435\times 0.2700]/100 = 0.00257$. Similar calculations, using (6.2.6), yield an estimated standard error of $\sigma(\hat{h}(150)) = 0.00073$.

TABLE 6.1
Weights Used in Smoothing the Nelson–Aalen Estimator for the ALL Group

t_i	$\Delta\hat{H}(t_i)$	$\Delta\hat{V}[\hat{H}(t_i)]$	$\frac{150-t_i}{100}$	$K\left(\frac{150-t_i}{100}\right)$	$\frac{50-t_i}{100}$	$K\left(\frac{50-t_i}{100}\right)$	$\frac{600-t_i}{100}$	$K\left(\frac{600-t_i}{100}\right)$
1	0.0263	0.00069	1.49	0.0000	0.49	1.0618	5.99	0.0000
55	0.0270	0.00073	0.95	0.0731	−0.05	0.9485	5.45	0.0000
74	0.0278	0.00077	0.76	0.3168	−0.24	0.7482	5.26	0.0000
86	0.0286	0.00082	0.64	0.4428	−0.36	0.6047	5.14	0.0000
104	0.0294	0.00087	0.46	0.5913	−0.54	0.3867	4.96	0.0000
107	0.0303	0.00091	0.43	0.6113	−0.57	0.3518	4.93	0.0000
109	0.0313	0.00099	0.41	0.6239	−0.59	0.3290	4.91	0.0000
110	0.0322	0.00103	0.40	0.6300	−0.60	0.3177	4.90	0.0000
122	0.0667	0.00222	0.28	0.6912	−0.72	0.1913	4.78	0.0000
129	0.0357	0.00128	0.21	0.7169	−0.79	0.1275	4.71	0.0000
172	0.0370	0.00138	−0.22	0.7137	−1.22	0.0000	4.28	0.0000
192	0.0385	0.00147	−0.42	0.6177	−1.42	0.0000	4.08	0.0000
194	0.0400	0.00161	−0.44	0.6048	−1.44	0.0000	4.06	0.0000
230	0.0435	0.00188	−0.80	0.2700	−1.80	0.0000	3.70	0.0000
276	0.0454	0.00207	−1.26	0.0000	−2.26	0.0000	3.24	0.0000
332	0.0476	0.00228	−1.82	0.0000	−2.82	0.0000	2.68	0.0000
383	0.0500	0.00247	−2.33	0.0000	−3.33	0.0000	2.17	0.0000
418	0.0527	0.00277	−2.68	0.0000	−3.68	0.0000	1.82	0.0000
468	0.0555	0.00310	−3.18	0.0000	−4.18	0.0000	1.32	0.0000
487	0.0589	0.00345	−3.37	0.0000	−4.37	0.0000	1.13	0.0000
526	0.0625	0.00391	−3.76	0.0000	−4.76	0.0000	0.74	0.2492
609	0.0714	0.00511	−4.59	0.0000	−5.59	0.0000	−0.09	0.8918
662	0.0769	0.00592	−5.12	0.0000	−6.12	0.0000	−0.62	0.6904

At $t = 50$ days, the asymmetric kernel (6.2.7) is used with $q = 50/100 = 0.5$. We have $\alpha_E = 64(2 - 4\times 0.5 + 6\times 0.5^2 - 3\times 0.5^3)/[(1+0.5)^4(19 - 18\times 0.5 + 3\times 0.5^2)] = 1.323$ and $\beta_E = 240(1-0.5)^2/[(1+.5)^4(19 - 18\times 0.5 + 3\times 0.5^2)] = 1.102$. Thus $K_{0.5}(-0.05) = 0.75[1.323 + 1.102(-0.05)]\times(1 - 0.05^2) = 0.9485$. Applying formulas (6.2.4) and (6.2.5) yields $\hat{h}(50) = 0.0015$ and $\sigma[\hat{h}(50)] = 0.00052$. Note that the tail adjustment using this kernel gives a higher weight to estimates of $\Delta\hat{H}$ smaller than 50 to compensate for the fact that we can not observe any estimates in the range −50 to 0.

At $t = 600$ days we make the upper tail correction. Here $q = (662 - 600)/100 = 0.62$, which yields $\alpha_E = 1.148$ and $\beta_E = 0.560$. Only deaths in the range 500–662 have a nonzero value of the kernel. For $t_i = 609$ days ($x = -0.09$) the weight is $K(-0.09) = 0.75[1.148 + 0.560(0.09)](1 - 0.09^2) = 0.8918$. Note that, because we are estimating h in the right-hand tail, we have replaced −0.09 by 0.09. Applying (6.2.4) and (6.2.5) yields $\hat{h}(600) = 0.0013$ and $\sigma[\hat{h}(600)] = 0.00084$.

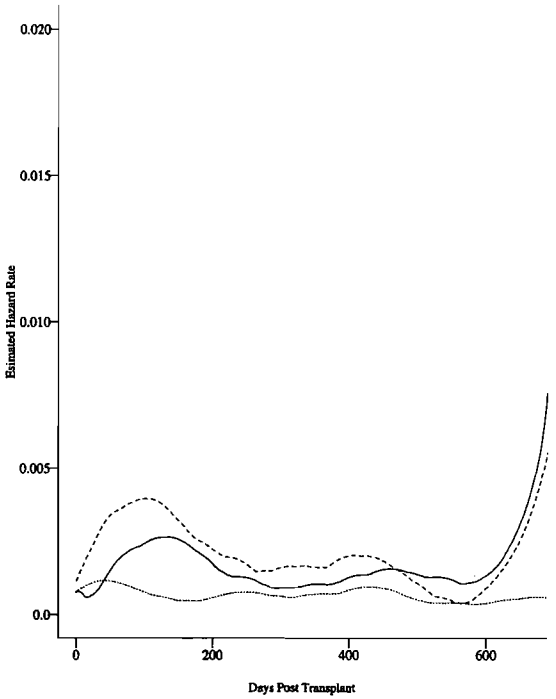

Figure 6.1 *Smoothed estimates of the hazard rates for bone marrow transplant patients based on the Epanechnikov kernel with a bandwidth of 100 days. ALL (———); AML-Low risk (———); AML-High risk (— — —).*

Figure 6.1 shows the estimated hazard rates for the three disease groups, indicating that the risk of relapse or death increases in the first 150 days after transplant after which the hazard rate decreases. The initial peak is higher for AML high-risk patients. The estimated hazard rates again confirm the impression that AML low-risk patients have the lowest rate of relapse or death.

EXAMPLE 6.2

We shall illustrate the effects of changing the bandwidth and the choice of kernel on the kidney transplant data in section 1.7. Here, we shall ignore the age and race of the patient at the time of transplant. The estimate of the hazard rate constructed serves as the unadjusted mortality rate for these transplant patients.

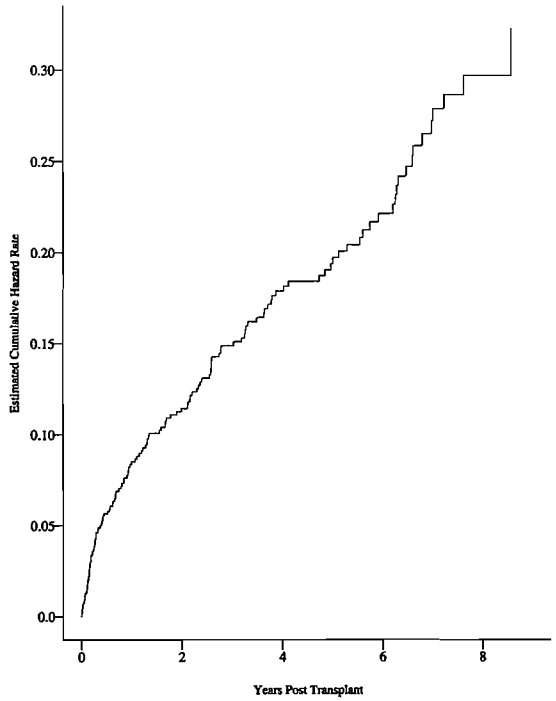

Figure 6.2 *Estimated cumulative hazard rate for kidney transplant patients*

Figure 6.2 shows the Nelson–Aalen estimate of the cumulative hazard rate on which the smoothed hazard rate estimator is based. Figure 6.3 shows the estimated hazard rate based on a bandwidth of 1 year for the uniform, Epanechnikov, and biweight kernels. Note that the kernels provide different degrees of smoothness. The biweight kernel is the smoothest, whereas the uniform kernel is rather jagged, typical of the performance of these kernels.

Figure 6.4 shows the effects of changing the bandwidth on the estimate of $h(t)$. In this figure, based on the Epanechnikov kernel, we see that increasing the bandwidth provides smoother estimates of the hazard rate. This increase in smoothness is at the expense of an increase in the bias of the estimate (see Practical Note 1).

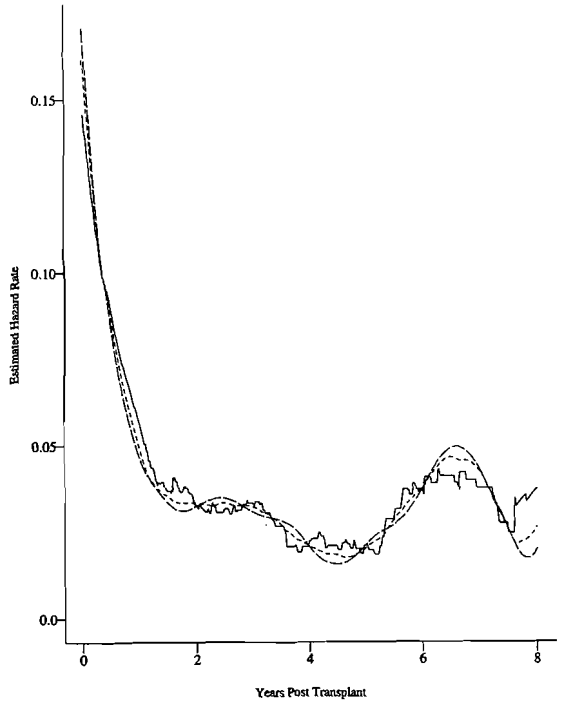

Figure 6.3 *Effects of changing the kernel on the smoothed hazard rate estimates for kidney transplant patients using a bandwidth of 1 year. Uniform kernel (——); Epanechnikov kernel (——) Biweight kernel (— — —)*

One problem in using kernel smoothing to obtain an estimate of the hazard rate is the selection of the proper bandwidth. One way to pick a good bandwidth is to use a cross-validation technique for determining the bandwidth that minimizes some measure of how well the estimator performs. One such measure is the mean integrated squared error (MISE) of $\hat{h}$ over the range τ_L to τ_U defined by

$$MISE(b) = E\int_{\tau_L}^{\tau_U}[\hat{h}(u) - h(u)]^2 du$$
$$= E\int_{\tau_L}^{\tau_U}\hat{h}^2(u)du - 2E\int_{\tau_L}^{\tau_U}\hat{h}(u)h(u)du + E\int_{\tau_L}^{\tau_U}h^2(u)du.$$

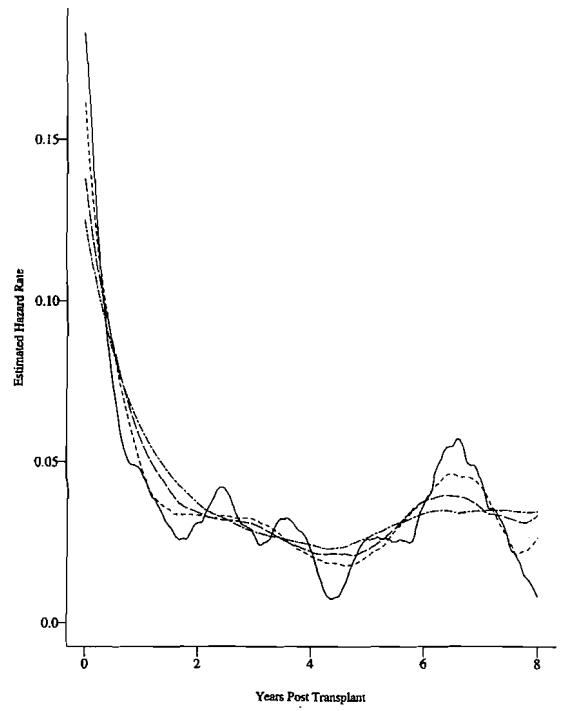

Figure 6.4 *Effects of changing the bandwidth on the smoothed hazard rate estimates for kidney transplant patients using the Epanechnikov kernel. bandwidth = 0.5 years (——) bandwidth = 1.0 years (——) bandwidth = 1.5 years (— — —) bandwidth = 2.0 years (— · — · —)*

This function depends both on the kernel used to estimate h and on the bandwidth b. Note that, although the last term depends on the unknown hazard rate, it is independent of the choice of the kernel and the bandwidth and can be ignored when finding the best value of b. The first term can be estimated by $\int_{\tau_L}^{\tau_U}\hat{h}^2(u)du$. If we evaluate $\hat{h}$ at a grid of points $\tau_L = u_1 < \cdots < u_M = \tau_U$, then, an approximation to this integral by the trapezoid rule is $\sum_{i=1}^{M-1}(\frac{u_{i+1}-u_i}{2})[\hat{h}^2(u_i) + \hat{h}^2(u_{i+1})]$. The second term can be estimated by a cross-validation estimate suggested by Ramlau–Hansen (1983a and b). This estimate is $b^{-1}\sum_{i\neq j}K(\frac{t_i-t_j}{b})\Delta\tilde{H}(t_i)\Delta\tilde{H}(t_j)$, where the sum is over the event times between τ_L and τ_U. Thus, to find the best value of b which minimizes the MISE for a fixed kernel, we

find b which minimizes the function

$$g(b) = \sum_{i=1}^{M-1} \left(\frac{u_{i+1} - u_i}{2} \right) [\hat{h}^2(u_i) + \hat{h}^2(u_{i+1})] - 2b^{-1} \sum_{i \neq j} K\left(\frac{t_i - t_j}{b} \right) \Delta\tilde{H}(t_i)\Delta\tilde{H}(t_j).$$

EXAMPLE 6.2 *(continued)* To find the best bandwidth for the kidney transplant patients, in Figure 6.5 we show a plot of b versus $g(b)$ for the three kernels with $\tau_L = 0$, $\tau_U = 6$ years. This figure is based on a grid of 100 equally spaced values for b over the range 0.01–1.00. The optimal values of b are 0.17 for the uniform kernel, 0.20 for the Epanechnikov kernel and

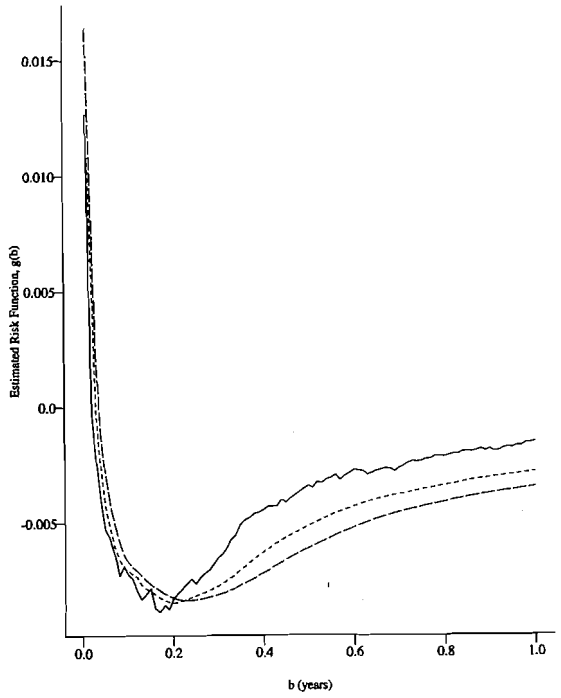

Figure 6.5 *Estimated risk function, $g(b)$, for use in determination of the best bandwidth for the kidney transplant data. Uniform kernel (———); Epanechnikov kernel (——) Biweight kernel (— — —).*

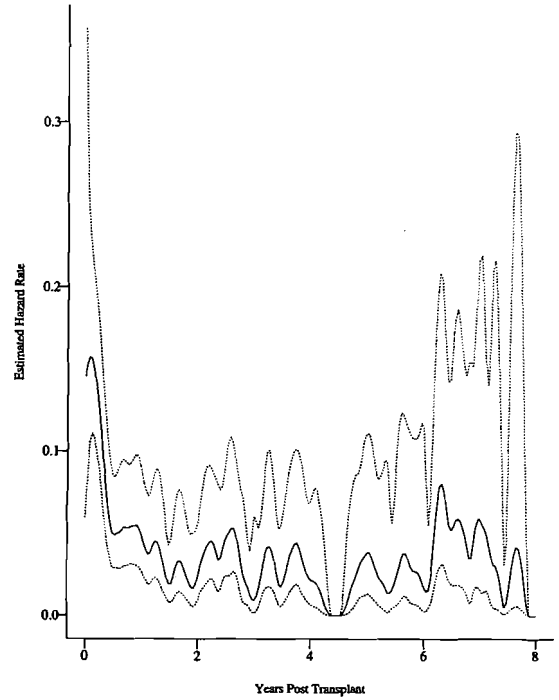

Figure 6.6 *Smoothed estimate of the hazard rate (———) and 95% confidence interval (——) for the time to death following a kidney transplant based on the biweight kernel and the best bandwidth.*

0.23 for the biweight kernel. Figure 6.6 shows the estimated hazard rate and a 95% pointwise confidence interval based on the biweight kernel with this optimal bandwidth.

EXAMPLE 6.1 *(continued)* The cross validation technique yields optimal bandwidths, based on the Epanechnikov kernel, of 161 days for the ALL group, 50 days for the AML low-risk group, and 112 days for the AML high-risk group for the estimates of the hazard rate over the range 0–730 days. Figure 6.7 shows the estimated hazard rates using these values of b.

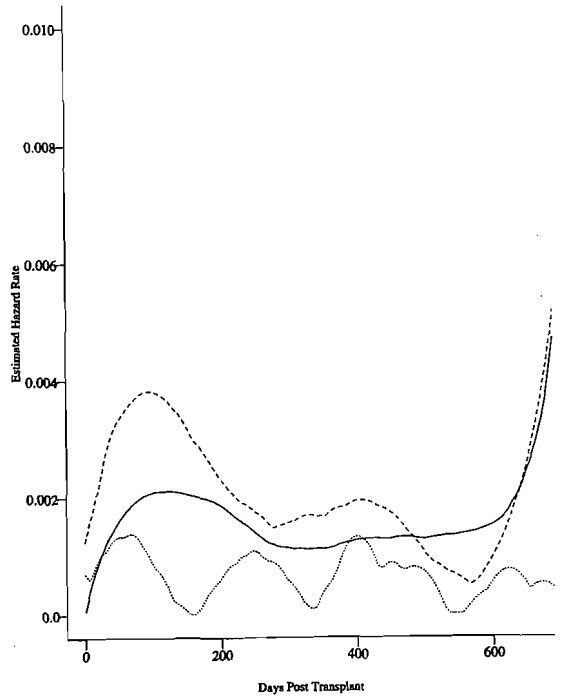

Figure 6.7 *Smoothed estimates of the hazard rates for bone marrow transplant patients based on the Epanechnikov kernel using optimal bandwidths. AML-Low risk (——) AML-High risk (— — —) ALL (———)*

Practical Notes

1. One must be very careful in interpreting the kernel-smoothed estimates constructed by these techniques. What these statistics are estimating is not the hazard rate $h(t)$, but rather a smoothed version of the hazard rate $h^*(t)$. This quantity is defined by $h^*(t) = b^{-1}\int K(\frac{t-u}{b})h(u)du$. It depends on both the bandwidth b and the kernel used in estimation. The confidence interval formula is, in fact, a confidence interval for h^*.
2. All that is required to apply the techniques in this section is an estimator of the cumulative hazard rate and its variance. Hence, these techniques apply equally well to right-censored or left-truncated data.

3. The smoothed estimator of the hazard rate was first introduced in Ramlau–Hansen (1983a and b). A detailed discussion of the large-sample properties of this estimator can be found in Andersen et al. (1993). A good general survey of smoothing techniques is found in Izenman (1991).

Theoretical Notes

1. The mean integrated squared error (MISE) measures $E\{\int_{\tau_L}^{\tau_U}[\hat{h}(u) - h(u)]^2 du\}$. This quantity is asymptotically approximately equal to the sum of a "bias" term, $\int\{h^*(u) - h(u)\}^2 du$ and a "variance" term $\int E\{[\hat{h}(u) - h^*(u)]^2\}du$. A small bandwidth produces a small bias term, but a large variance term, whereas the reverse holds for a large bandwidth. The optimal bandwidth is a trade-off between the two terms.
2. The bias of the smoothed hazard rate estimator, for large n, is approximately, $0.5b^2h''(t)k^*$, where h'' is the second derivative of h and $k^* = \int_{-1}^{1} s^2K(s)ds$.

6.3 Estimation of Excess Mortality

In some applications of survival analysis techniques, it is of interest to compare the mortality experience of a group of individuals to a known standard survival curve. The reference survival curve, which may be different for each individual in the sample, could be drawn from published mortality tables or other population-based mortality studies. Two simple models have been proposed to provide an inference on how the study population's mortality differs from that in the reference population.

Suppose we have data on n individuals. Let $\theta_j(t)$ be the reference hazard rate for the jth individual in the study. This known reference hazard rate typically depends on the characteristics of the jth patient, such as race, sex, age, etc. The first model for excess mortality, commonly known as the relative mortality model, assumes that the hazard rate at time t for the jth patient under study conditions, $h_j(t)$, is a multiple, $\beta(t)$, of the reference hazard rate for this individual, that is,

$$h_j(t) = \beta(t)\theta_j(t), \quad j = 1, 2, \ldots, n. \tag{6.3.1}$$

Here, if $\beta(t)$ is greater than 1, then, individuals in the study group are experiencing the event of interest at a faster rate than comparable

individuals in the reference population. Let $B(t) = \int_0^t \beta(u)du$ be the cumulative relative excess mortality.

The data available for estimating $B(t)$, for each individual, consists of study times and death indicators. For the jth individual, let $Y_j(t)$ be 1 if the individual is at risk at time t and 0, otherwise. Note that this definition of $Y_j(t)$ allows for left-truncated and right-censored data. Define the function $Q(t) = \sum_{j=1}^{n} \theta_j(t) Y_j(t)$. To allow for ties in the data, let $t_1 < t_2 < \cdots < t_D$ be the times at which the events occur and d_i the number of events observed at time t_i. The estimator of $B(t)$ is given by

$$\hat{B}(t) = \sum_{t_i \le t} \frac{d_i}{Q(t_i)}. \tag{6.3.2}$$

An estimator of the variance of $\hat{B}(t)$ is given by

$$\hat{V}[\hat{B}(t)] = \sum_{t_i \le t} \frac{d_i}{Q(t_i)^2}. \tag{6.3.3}$$

The statistic $\hat{B}(t)$ has a large-sample normal distribution so that confidence intervals or confidence bands for the cumulative relative mortality can be constructed by replacing the Nelson–Aalen estimator and its variance by $\hat{B}(t)$ and its variance in the appropriate formulas in sections 4.4 and 4.5. A crude estimator of the relative risk function $\beta(t)$ is given by the slope of the estimated cumulative relative mortality estimator. An improved estimator of $\hat{B}(t)$ can be found by a kernel smoothing of $\hat{B}(t)$ similar to that developed for the estimated cumulative hazard rate discussed in the previous section.

EXAMPLE 6.3 To illustrate the estimation of the relative mortality function consider the data on the 26 psychiatric patients in Iowa described in section 1.15. We shall use the 1959–1961 Iowa State life tables (US Dept. of Health and Human Services (1959)) as the reference population. This life table in Table 6.2 is based on the 1960 census and the average number of deaths in the period 1959–1961 and provides the population survival functions $S()$ for males and females. For the population hazard rates, we assume that the hazard rates are constant over each one year interval reported in the table, so that the hazard rate at age a is $\lambda(a) = -\ln[S(a)] - \{-\ln[S(a+1)\}$. Table 6.2 shows values of the estimated hazard rates for males and female for $a = 18, 19, \ldots, 77$.

The time scale used in this example is the time on study for each patient. A patient who enters the study at age a has $\theta_i(t)$ found by using the hazard rate in the $(a+t)$th row of Table 6.2. For example, the female who entered the study at age 36 has $\theta(1) = \lambda_F(36+1) = 0.00130$, $\theta(2) = \lambda_F(38) = 0.00140$, etc. Table 6.3 shows the estimate of $B(t)$

TABLE 6.2
1960 Iowa Standard Mortality

	Males				
Age	*Survival Function*	*Hazard Rate*	*Age*	*Survival Function*	*Hazard Rate*
18–19	0.96394	0.00154	48–49	0.89596	0.00694
19–20	0.96246	0.00164	49–50	0.88976	0.00751
20–21	0.96088	0.00176	50–51	0.88310	0.00810
21–22	0.95919	0.00188	51–52	0.87598	0.00877
22–23	0.95739	0.00190	52–53	0.86833	0.00956
23–24	0.95557	0.00185	53–54	0.86007	0.01052
24–25	0.95380	0.00173	54–55	0.85107	0.01159
25–26	0.95215	0.00158	55–56	0.84126	0.01278
26–27	0.95065	0.00145	56–57	0.83058	0.01402
27–28	0.94927	0.00137	57–58	0.81902	0.01536
28–29	0.94797	0.00134	58–59	0.80654	0.01683
29–30	0.94670	0.00136	59–60	0.79308	0.01844
30–31	0.94541	0.00141	60–61	0.77859	0.02013
31–32	0.94408	0.00146	61–62	0.76307	0.02195
32–33	0.94270	0.00153	62–63	0.74650	0.02386
33–34	0.94126	0.00159	63–64	0.72890	0.02586
34–35	0.93976	0.00170	64–65	0.71029	0.02795
35–36	0.93816	0.00181	65–66	0.69071	0.03020
36–37	0.93646	0.00198	66–67	0.67016	0.03262
37–38	0.93461	0.00215	67–68	0.64865	0.03521
38–39	0.93260	0.00235	68–69	0.62621	0.03800
39–40	0.93041	0.00258	69–70	0.60286	0.04102
40–41	0.92801	0.00284	70–71	0.57863	0.04424
41–42	0.92538	0.00312	71–72	0.55359	0.04773
42–43	0.92250	0.00350	72–73	0.52779	0.05175
43–44	0.91928	0.00397	73–74	0.50117	0.05646
44–45	0.91564	0.00450	74–75	0.47366	0.06188
45–46	0.91153	0.00511	75–76	0.44524	0.06795
46–47	0.90688	0.00575	76–77	0.41599	0.07454
47–48	0.90168	0.00636	77–78	0.38611	0.08181

and its standard error. Figure 6.8 shows the estimated value of $B(t)$ and a 95% pointwise confidence interval for $B(t)$ based on the log-transformed confidence interval formula for the cumulative hazard rate. (See Practical Note 1 in section 4.3. Here we use Eq. 4.3.5 and replace $\hat{H}(t_0)$ by $\hat{B}(t)$ and $\sigma_H(t_0)$ by the standard error of $\hat{B}(t)$.)

The slope of $\hat{B}(t)$ in Figure 6.8 provides a crude estimate of $\beta(t)$. Here, we see that, in the first two years of observation, psychiatric patients were 20–30 times more likely to die than comparable individuals

TABLE 6.2
1960 Iowa Standard Mortality

	Females				
Age	*Survival Function*	*Hazard Rate*	*Age*	*Survival Function*	*Hazard Rate*
18–19	0.97372	0.00057	48–49	0.93827	0.00352
19–20	0.97317	0.00056	49–50	0.93497	0.00381
20–21	0.97263	0.00055	50–51	0.93141	0.00414
21–22	0.97210	0.00054	51–52	0.92756	0.00448
22–23	0.97158	0.00054	52–53	0.92341	0.00481
23–24	0.97106	0.00056	53–54	0.91898	0.00509
24–25	0.97052	0.00059	54–55	0.91431	0.00536
25–26	0.96995	0.00062	55–56	0.90942	0.00565
26–27	0.96935	0.00065	56–57	0.90430	0.00600
27–28	0.96872	0.00069	57–58	0.89889	0.00653
28–29	0.96805	0.00072	58–59	0.89304	0.00724
29–30	0.96735	0.00075	59–60	0.88660	0.00812
30–31	0.96662	0.00079	60–61	0.87943	0.00912
31–32	0.96586	0.00084	61–62	0.87145	0.01020
32–33	0.96505	0.00088	62–63	0.86261	0.01132
33–34	0.96420	0.00095	63–64	0.85290	0.01251
34–35	0.96328	0.00103	64–65	0.84230	0.01376
35–36	0.96229	0.00110	65–66	0.83079	0.01515
36–37	0.96123	0.00121	66–67	0.81830	0.01671
37–38	0.96007	0.00130	67–68	0.80474	0.01846
38–39	0.95882	0.00140	68–69	0.79002	0.02040
39–40	0.95748	0.00152	69–70	0.77407	0.02259
40–41	0.95603	0.00162	70–71	0.75678	0.02494
41–42	0.95448	0.00176	71–72	0.73814	0.02754
42–43	0.95280	0.00193	72–73	0.71809	0.03067
43–44	0.95096	0.00216	73–74	0.69640	0.03446
44–45	0.94891	0.00240	74–75	0.67281	0.03890
45–46	0.94664	0.00268	75–76	0.64714	0.04376
46–47	0.94411	0.00296	76–77	0.61943	0.04902
47–48	0.94132	0.00325	77–78	0.58980	0.05499

in the standard population. In years 3–40, the patients were between 2–5 times more likely to die.

A second model, which can be used for comparing the study population to a reference population is the excess or additive mortality model. Here, we assume that the hazard rate at time t for the jth individual under study is a sum of the population mortality rate $\theta_j(t)$ and an excess mortality function $\alpha(t)$. The function $\alpha(t)$, which is assumed to be

TABLE 6.3
Computation of Cumulative Relative Mortality for 26 Psychiatric Patients

t_i	d_i	$Q(t_i)$	$\hat{B}(t)$	$\hat{V}[\hat{B}(t)]$	$\sqrt{\hat{V}[\hat{B}(t)]}$
1	2	0.05932	33.72	568.44	23.84
2	1	0.04964	53.86	974.20	31.21
11	1	0.08524	65.59	1111.84	33.34
14	1	0.10278	75.32	1206.51	34.73
22	2	0.19232	85.72	1260.58	35.50
24	1	0.19571	90.83	1286.69	35.87
25	1	0.18990	96.10	1314.42	36.25
26	1	0.18447	101.52	1343.81	36.66
28	1	0.19428	106.67	1370.30	37.02
32	1	0.18562	112.05	1399.32	37.41
35	1	0.16755	118.02	1434.94	37.88
40	1	0.04902	138.42	1851.16	43.03

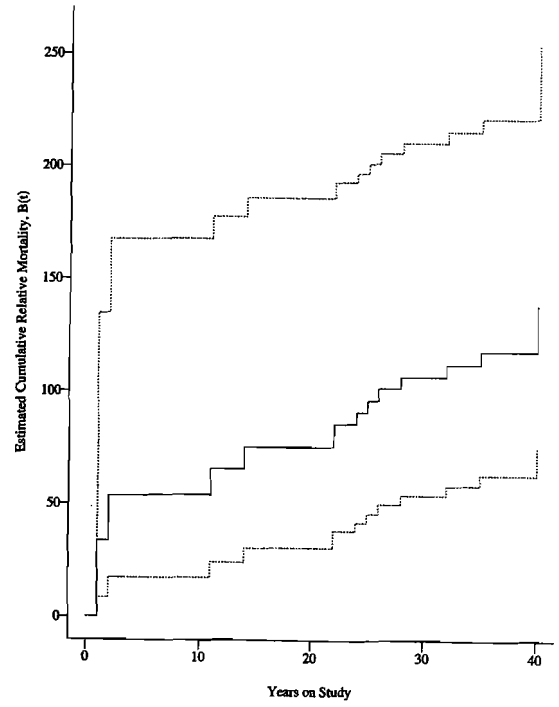

Figure 6.8 *Estimated cumulative relative mortality (solid line) and 95% pointwise confidence interval (dashed line) for Iowa psychiatric patients*

the same for all individuals in the study group, can be positive when study patients are dying faster than those in the reference population or be negative when the study group has a better survival rate than the reference population. The model is

$$h_j(t) = \alpha(t) + \theta_j(t), \quad j = 1, \ldots, n. \tag{6.3.4}$$

As in the case of the multiplicative model, direct estimation of $\alpha()$ is difficult. Instead, we estimated the cumulative excess mortality function $A(t) = \int_0^t \alpha(u)du$. The estimator of $A(t)$ is constructed from the difference of the observed hazard rate, estimated by the ordinary Nelson–Aalen estimator (see section 4.2) $\tilde{H}(t)$ and an "expected" cumulative hazard rate $\Theta(t)$ based on the reference hazard rates. The expected cumulative hazard rate is a weighted average of the reference cumulative hazard rates at each time, where the weights are based on the fraction of individuals at risk at time t, that is,

$$\Theta(t) = \sum_{j=1}^{n} \int_0^t \theta_j(u) \frac{Y_j(u)}{Y(u)} du, \tag{6.3.5}$$

where $Y(t) = \sum_{j=1}^{n} Y_j(t)$ is the number at risk at time t.

The estimated excess mortality is given by

$$\hat{A}(t) = \sum_{t_i \le t} \frac{d_i}{Y(t)} - \Theta(t). \tag{6.3.6}$$

The estimated variance of the cumulative excess mortality function is given by the variance of the Nelson–Aalen estimator, namely,

$$\hat{V}[\hat{A}(t)] = \sum_{t_i \le t} \frac{d_i}{Y(t)^2}. \tag{6.3.7}$$

As for the relative mortality model, confidence intervals and confidence bands for $A(t)$ can be computed using the techniques in sections 4.3 and 4.4, and smoothed estimates of $\alpha(t)$ can be constructed using the methods of the previous section.

The $\hat{A}(t)$ may be either negative or positive. It will be decreasing and negative for times smaller than the smallest death time. With this caution in mind, one may use these estimates to construct "corrected" survival curves. The Kaplan–Meier estimator, $\hat{S}(t)$, provides an estimate of the observed or uncorrected survival curve. The survival curve, $S^*(t) = \exp[-\Theta(t)]$, provides an estimate of the expected survival curve if the reference mortality model is the same as the study population. The ratio of these two survival functions, $S^C(t) = \hat{S}(t)/S^*(t)$, is taken as a "corrected" survival function estimate for the study population. Care must be taken in using this curve because the ratio of the two curves may be greater than one (especially for small t) and the corrected survival curve need not be nonincreasing.

EXAMPLE 6.3 *(continued)* To estimate the expected hazard rate using the standard Iowa mortality data, we first compute $\Theta(t)$. Here, we assume, again, that the hazard rates are constant over each age interval of unit length which simplifies computations. At one year after entry into the study, $\Theta(1) = \sum_{j=1}^{n} \lambda_S(a_j)/26$, where a_j is the age of the jth individual at entry into the study and $\lambda_S()$ is the value of the hazard rate from Table 6.2 for the patient's sex. For an integer age $t > 1$, $\Theta(t) = \Theta(t-1) + \sum_t \lambda_S(a_j + t - 1)/Y(t)$, where the sum is over all patients under observation in the interval $[t-1, t)$. For noninteger times, $\Theta(t)$ is found by linear interpolation.

Table 6.4 shows the results of the computations. Figure 6.9 shows the observed cumulative hazard rate $[\tilde{H}(t)]$, the expected cumulative hazard rate $[\Theta(t)]$ and the cumulative excess mortality $[\hat{A}(t)]$. Notice that the expected cumulative hazard function is a smooth function of the number of years on study, whereas the Nelson–Aalen estimator is a step function with jumps at the observed death times. The excess mortality function has jumps at the death times and is decreasing between the death times. From this figure, we see that a crude estimate of $\alpha(t)$, given by the slope of $\hat{A}(t)$, is a function which is about 0.05 for $t < 2$, about 0 for $2 < t < 21$, and, then, about 0.05 for $t > 21$. After 30 years on study, the cumulative excess mortality is about 0.35, so we estimate that, in a group of 100 patients, we would see 35 more deaths after 30 years than we would expect to see in a standard population. A crude 95% confidence interval for the excess number of deaths after 30 years is $0.3592 \pm 1.96(0.1625)$ or $(0.0407, 0.6777)$. These estimates are a bit imprecise due to the relatively small sample size of this study.

Figure 6.10 depicts the adjusted survival curves for this study. Again, the expected survival function is a smooth curve, and the observed survival curve is a step function. It is of interest here to note that the "corrected" survival curve is not monotone decreasing and, as such, is not strictly a survival curve. A better graphical representation is to plot this function by connecting the points $\hat{S}(t_i)/S^*(t_i)$ only at the death times.

Practical Notes

1. The estimator of relative mortality is a time-varying extension of the standard mortality ratio (SMR) estimator (Breslow, 1975) which assumes a constant relative mortality over time. For this estimator, one computes $E(t) = \int_0^t Q(u)du$, which is thought of as the expected number of deaths before time t. If $\beta(t) = \beta_0$, a constant, then, the

TABLE 6.4
Computation for the Excess Mortality Model

t_i	d_i	$Y(t_i)$	$\tilde{H}(t_i)$	$\Theta(t_i)$	$\hat{A}(t)$	$SE[\hat{A}(t)]$	$\hat{S}(t)$	$S^*(t_i)$	$\frac{\hat{S}(t_i)}{S^*(t_i)}$
1	2	26	0.0769	0.0021	0.0748	0.0544	0.9231	0.9979	0.9250
2	1	24	0.1186	0.0041	0.1145	0.0685	0.8846	0.9959	0.8882
3	0	23	0.1186	0.0059	0.1127	0.0685	0.8846	0.9941	0.8899
4	0	23	0.1186	0.0079	0.1107	0.0685	0.8846	0.9921	0.8917
5	0	23	0.1186	0.0101	0.1085	0.0685	0.8846	0.9900	0.8936
6	0	23	0.1186	0.0124	0.1062	0.0685	0.8846	0.9877	0.8956
7	0	23	0.1186	0.0148	0.1038	0.0685	0.8846	0.9853	0.8978
8	0	23	0.1186	0.0175	0.1011	0.0685	0.8846	0.9827	0.9002
9	0	23	0.1186	0.0203	0.0983	0.0685	0.8846	0.9799	0.9028
10	0	23	0.1186	0.0234	0.0952	0.0685	0.8846	0.9769	0.9056
11	1	23	0.1621	0.0268	0.1353	0.0811	0.8462	0.9736	0.8691
12	0	22	0.1621	0.0303	0.1318	0.0811	0.8462	0.9702	0.8722
13	0	22	0.1621	0.0341	0.1279	0.0811	0.8462	0.9664	0.8755
14	1	22	0.2075	0.0384	0.1691	0.0930	0.8077	0.9623	0.8393
15	0	21	0.2075	0.0428	0.1647	0.0930	0.8077	0.9581	0.8430
16	0	21	0.2075	0.0476	0.1599	0.0930	0.8077	0.9535	0.8471
17	0	21	0.2075	0.0530	0.1546	0.0930	0.8077	0.9484	0.8516
18	0	21	0.2075	0.0588	0.1487	0.0930	0.8077	0.9429	0.8566
19	0	21	0.2075	0.0652	0.1423	0.0930	0.8077	0.9369	0.8621
20	0	21	0.2075	0.0722	0.1353	0.0930	0.8077	0.9303	0.8682
21	0	21	0.2075	0.0799	0.1276	0.0930	0.8077	0.9232	0.8749
22	2	21	0.3028	0.0882	0.2145	0.1148	0.7308	0.9155	0.7982
23	0	19	0.3028	0.0968	0.2060	0.1148	0.7308	0.9078	0.8050
24	1	19	0.3554	0.1062	0.2492	0.1263	0.6923	0.8993	0.7698
25	1	18	0.4109	0.1158	0.2952	0.1380	0.6538	0.8907	0.7341
26	1	17	0.4698	0.1257	0.3441	0.1500	0.6154	0.8819	0.6978
27	0	16	0.4698	0.1358	0.3340	0.1500	0.6154	0.8730	0.7049
28	1	16	0.5323	0.1469	0.3854	0.1625	0.5769	0.8634	0.6682
29	0	15	0.5323	0.1594	0.3729	0.1625	0.5769	0.8527	0.6766
30	0	15	0.5323	0.1731	0.3592	0.1625	0.5769	0.8411	0.6860
31	0	13	0.5323	0.1874	0.3449	0.1625	0.5769	0.8291	0.6958
32	1	11	0.6232	0.2028	0.4204	0.1862	0.5245	0.8164	0.6424
33	0	10	0.6232	0.2207	0.4025	0.1862	0.5245	0.8019	0.6540
34	0	8	0.6232	0.2412	0.3820	0.1862	0.5245	0.7857	0.6676
35	1	7	0.7660	0.2631	0.5029	0.2347	0.4496	0.7687	0.5849
36	0	4	0.7660	0.2848	0.4812	0.2347	0.4496	0.7522	0.5977
37	0	3	0.7660	0.3133	0.4527	0.2347	0.4496	0.7310	0.6150
38	0	2	0.7660	0.3510	0.4150	0.2347	0.4496	0.7040	0.6387
39	0	2	0.7660	0.3926	0.3734	0.2347	0.4496	0.6753	0.6658
40	1	1	1.7660	0.4363	1.3297	1.0272	0.0000	0.6464	0.0000

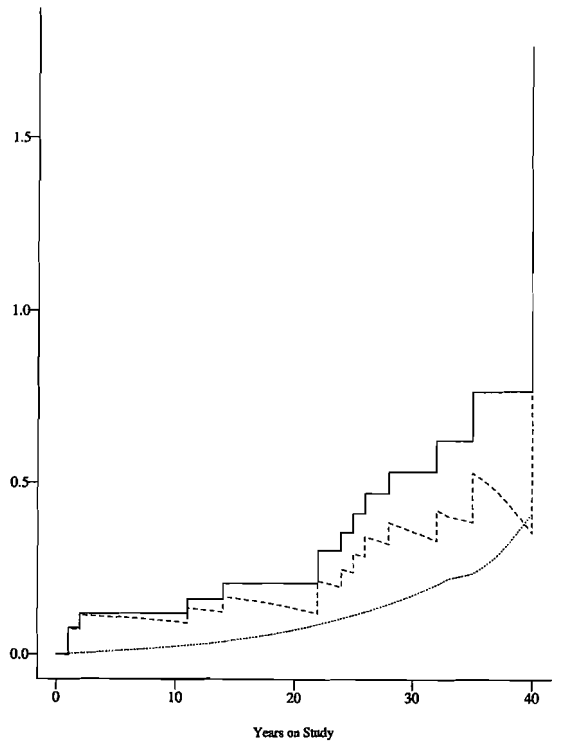

Figure 6.9 *Estimated cumulative excess mortality for Iowa psychiatric patients. Nelson–Aalen estimator (———) Expected cumulative hazard (——) Cumulative excess mortality (— — —)*

maximum likelihood estimator of β_0 is the total number of deaths divided by $E(t_{MAX})$, where t_{MAX} is the largest on study time. The SMR is 100 times this value. If the constant mortality model holds, then, a plot of $\hat{B}(t)$ versus t should be a straight line through the origin. Andersen and Væth (1989) present a test for constant mortality and a version of the total time on test plot which can be used as a graphical check of the assumption of constant relative mortality.

2. An estimator of constant excess mortality $\alpha(t) = \alpha_0$ was proposed by Buckley (1984). For this estimator, let $T(t) = \int_0^t Y(u)du$ be the total time on test at time t, that is, the number of person-years of

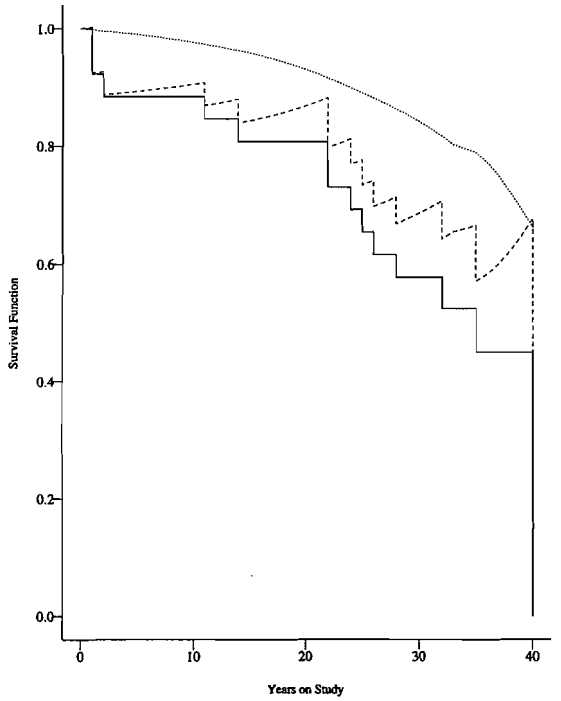

Figure 6.10 *Adjusted survival curves for Iowa psychiatric patients. Observed survival (———) Expected survival (——) Corrected survival (— — —)*

observation prior to time t. At the largest study time, t_{MAX}, $T(t_{MAX})$ is the total years of exposure of all study individuals. The statistic $\frac{D-E(t_{MAX})}{T(t_{MAX})}$ estimates α_0. Here D is the total number of deaths. This estimate is the difference between the occurrence/exposure rate and the expected number of deaths per time on test. Buckley also presents a maximum likelihood estimator that must be found numerically. Again, the constant excess mortality model is reasonable if the plot of $\hat{A}(t)$ versus t is linear. Andersen and Væth (1989) present a formal test.

3. A more general model for excess mortality is a mixed model. Here, $h_j(t) = \beta(t)\theta_j(t) + \alpha(t)$. This model can be fit into an additive regression formulation discussed in Chapter 10.

Theoretical Note

1. Detailed derivation of the estimators for excess and relative mortality are found in Andersen and Væth (1989). These statistics can be derived using a counting process technique as discussed in Andersen et al. (1993).

6.4 Bayesian Nonparametric Methods

An alternative to the classical nonparametric approach to estimating the survival function discussed in Chapters 4 and 5 is to use Bayesian nonparametric methods. In applying these methods, an investigator's a priori belief in the shape of the survival function is combined with the data to provide an estimated survival function. The prior information, which may be based on previous experience with the process under observation or based on expert opinion, is reflected in a prior distribution for the survival function. The sample information is contained in the likelihood function. These two distinct pieces of information are combined by Bayes' theorem to obtain an a posteriori distribution of the survival function which is the distribution of the survival function, given the data.

In the Bayesian approach, the parameters of the model are treated as random variables selected from the prior distribution. This prior distribution, which is a multivariate distribution on the parameters, is selected to reflect the investigator's prior belief in the values of the parameters. Typically, the prior means reflect the investigators best guess, before seeing any data, of the value of the parameters, and the prior variance is a measure of the investigator's uncertainty in his prior means. Often one can think of the prior variance as being inversely proportional to the amount of sample information to be represented by the prior.

In our problem, the parameter of interest is the survival function or, equivalently, the cumulative hazard function. This is to be treated as a random quantity sampled from some stochastic process. Nature picks a sample path from this stochastic process, and this is our survival function. We, then, have data sampled from a population with this survival function which we shall combine with our prior to obtain the distribution of the survival function, given the data.

To obtain an estimate of the survival function, we need to specify a loss function on which to base the decision rule. Analogous to the

simple parametric case, we shall use the squared-error loss function

$$L(S, \hat{S}) = \int_0^\infty [\hat{S}(t) - S(t)]^2 dw(t),$$

where $w(t)$ is a weight function. This loss function is the weighted integrated difference between the true value of the survival function and our estimated value. For this loss function, the value of $\hat{S}$, which minimizes the posterior expected value of $L(S, \hat{S})$, is the posterior mean and the Bayes risk $E[L(S, \hat{S}) \mid \text{DATA}]$ is the posterior variance.

Two classes of prior distributions have been suggested for this problem. Both lead to closed form estimates of the survival function using the squared-error loss function. These priors are chosen because they are conjugate priors for either the survival function or the cumulative hazard function. For a conjugate prior, the prior and posterior distributions are in the same family.

The first prior is for the survival function. For this prior, we assume that the survival function is sampled from a Dirichlet process with a parameter function α. A Dirichlet process, defined on the positive real line, has the property that, for any set of intervals $A_1, \ldots, A_k$, which partition the positive real line, the joint distribution of the prior probabilities $Pr[X \in A_1] = W_1, \ldots, Pr[X \in A_k] = W_k$ has a k dimensional Dirichlet distribution with parameters $[\alpha(A_1), \ldots, \alpha(A_k)]$. This property must hold for any such set of intervals and any k. A k vector $(W_1, \ldots, W_k)$ has a k-dimensional Dirichlet distribution with parameters $(\alpha_1, \ldots, \alpha_k)$ if $W_i = Z_i / \sum_{i=1}^k Z_i$ where the Z_i's are independent gamma random variables with shape parameter α_i. The joint density function of $(W_1, \ldots, W_{k-1})$ is given by

$$f(w_1, \ldots, w_{k-1}) = \frac{\Gamma[\alpha_1 + \cdots + \alpha_k]}{\Gamma[\alpha_1] \cdots \Gamma[\alpha_k]} \left[\prod_{i=1}^{k-1} w_i^{\alpha_i - 1}\right] \left[1 - \sum_{i=1}^{k-1} w_i\right]^{\alpha_k - 1}.$$

The mean of W_i is α_i/α and the variance is $(\alpha - \alpha_i)\alpha_i/(\alpha^2 + \alpha^3)$ where $\alpha = \sum_{i=1}^k \alpha_i$. When $k = 2$ the Dirichlet distribution reduces to the beta distribution with parameters (α_1, α_2).

To assign a prior distribution to the survival function, we assume that $S(t)$ follows a Dirichlet distribution with parameter function α. Typically, we take the parameter function to be of the form $\alpha([t, \infty)) = cS_0(t)$ where $S_0(t)$ is our prior guess at the survival function and c is a measure of how much weight to put on our prior guess. With this prior distribution for $S(t)$, the prior mean is expressed by

$$E[S(t)] = \frac{\alpha(t, \infty)}{\alpha(0, \infty)} = \frac{cS_0(t)}{cS_0(0)} = S_0(t),$$

and the prior variance is given by

$$V[S(t)] = \frac{[\alpha(0, \infty) - \alpha(t, \infty)]\alpha(t, \infty)}{[\alpha(0, \infty)^2 + \alpha(0, \infty)^3]} = \frac{S_0(t)[1 - S_0(t)]}{c + 1}.$$

Note that the prior variance is the equivalent to the sample variance one would have if we had an uncensored sample of size $c + 1$ from a population with a survival function $S_0(t)$. To illustrate what the sample paths of the prior distribution for S look like, we have simulated 10 sample paths for a Dirichlet prior with $S_0(t) = \exp(-0.1t)$ and $c = 5$. These are plotted as dashed lines in Figure 6.11 along with their mean $S_0(t)$, which is plotted as a solid line. Here we see that each sample path is a nonincreasing function with a value of 1 at 0. Note that, although the curves are continuous functions, they are not too smooth in this example. As the value of c increases, the curves will become smoother.

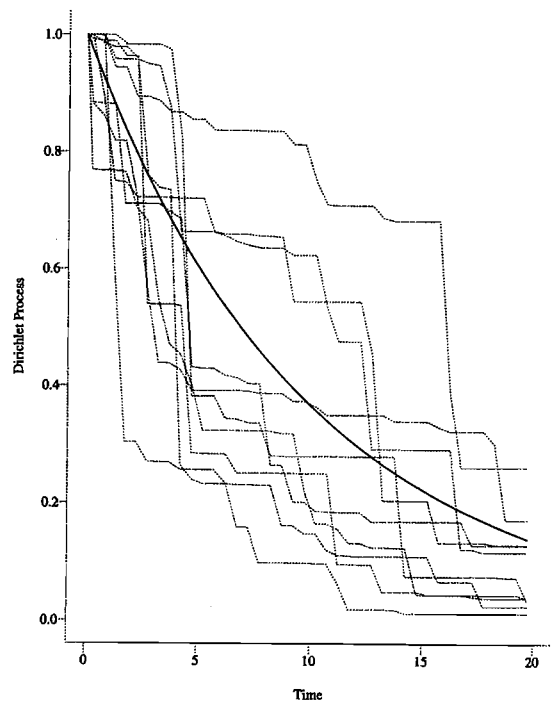

Figure 6.11 *Sample of ten sample paths (dashed lines) and their mean (solid line) for samples from a Dirichlet prior with* $S_0(t) = \exp(-0.1t)$ *and* $c = 5$.

The data we have available to combine with our prior consists of the on study times T_j and the event indicator, δ_j. To simplify calculations let $0 = t_o < t_1 < \cdots < t_M < t_{M+1} = \infty$, denote the M distinct times (censored or uncensored). At time t_i, let Y_i be the number of individuals at risk, d_i the number of deaths and λ_i the number of censored observations. Let Δ_i be 1 if $d_i > 0$ and 0 if $d_i = 0$.

Combining this data with the prior, we find that the posterior distribution of S is also Dirichlet. The parameter of the posterior distribution, α^*, is the original α parameter plus a point mass of one at points where deaths occur. That is, for any interval (a, b),

$$\alpha^*((a, b)) = \alpha((a, b)) + \sum_{j=1}^{n} I[\delta_j > 0,\ a < T_j < b],$$

where $I[\,]$ is the indicator function.

The Bayes estimator of the survival function is

$$\tilde{S}_D(t) = \frac{\alpha(t, \infty) + Y_{i+1}}{\alpha(0, \infty) + n} \prod_{k=1}^{i} \frac{\alpha(t_k, \infty) + Y_{k+1} + \lambda_k}{\alpha(t_k, \infty) + Y_{k+1}} \qquad (6.4.1)$$

$$\text{for } t_i \leq t < t_{i+1},\ i = 0, \ldots, M$$

The Bayes estimator is a continuous function between the distinct death times and has jumps at these death times. For large n this reduces to the Kaplan–Meier estimator, so that the prior information plays no role in the estimate. For small samples, the prior will dominate, and the estimator will be close to the prior guess at S.

A second approach to modeling prior information for survival data is to provide a prior distribution for the cumulative hazard function $H(t) = -\ln[S(t)]$. Here, we shall use a beta process prior. This prior depends on two parameters, $H_0(t)$ and $c(t)$. $H_0(t)$ is a prior guess at the value of the cumulative hazard rate $H(t)$, and $c(t)$ is a measure of how much weight to put on the prior guess at the function $H(t)$ at time t. For this prior, if we let $A_i = [a_{i-1}, a_i)$, $i = 1, \ldots, k$ be a series of nonoverlapping intervals with $0 = a_0 < a_1 < a_2 < \cdots < a_k$, then, a priori, $W_1 = H(a_1) - H(a_0), \ldots W_k = H(a_k) - H(a_{k-1})$ are independent beta random variables with parameters $p_i = c([a_i + a_{i-1}]/2)[H_0(a_i) - H_0(a_{i-1})]$ and $q_i = c([a_i + a_{i-1}]/2)\{1 - [H_0(a_i) - H_0(a_{i-1})]\}$. The prior mean of $H(a_i) - H(a_{i-1})$ is $H_0(a_i) - H_0(a_{i-1})$ and the prior variance is

$$V(W_i) = \frac{\{H_0(a_i) - H_0(a_{i-1})\}\{1 - [H_0(a_i) - H_0(a_{i-1})]\}}{c([a_i + a_{i-1}]/2) + 1}.$$

Here, $c(t)$ can be thought of as the weight to be given to our prior guess at $H_0(a_i) - H_0(a_{i-1})$ at the time $[a_i + a_{i-1}]/2$. The beta process prior is obtained by letting the number of subintervals increase to infinity, so that the interval lengths go to zero. Roughly speaking, $H(t)$ has a beta process if $dH(s)$ has a beta distribution with parameters $c(s)h_0(s)$ and $c(s)[1 - h_0(s)]$, and $dH(s)$ is independent of $dH(u)$ for $u \neq s$. (Here, $dH(s) = [H(s + ds) - H(s)]ds$ for a very small increment of time, and $h_0(t) = dH_0(t)/dt$.)

To illustrate what the sample paths of the prior distribution, for S based on a beta process prior, look like, we have simulated 10 sample paths for a beta process prior with $H_0(t) = 0.1t$ and $c(t) = 5$. These are plotted as dashed lines in Figure 6.12 along with the prior guess at the survival function, $\exp(-0.1t)$. Here we see that each sample path is a nondecreasing function with a value of 1 at 0. As for the Dirichlet process prior, the sample paths are continuous and nondecreasing. As compared to the Dirichlet, the sample paths for the beta process prior are less variable, especially, in the middle section of the curve.

When the data is right-censored with $D(t)$ deaths observed at or prior to time t and $Y(t)$ individuals at risk at time t and a beta process prior is used, then, the posterior distribution of $H(t)$ is a beta process with

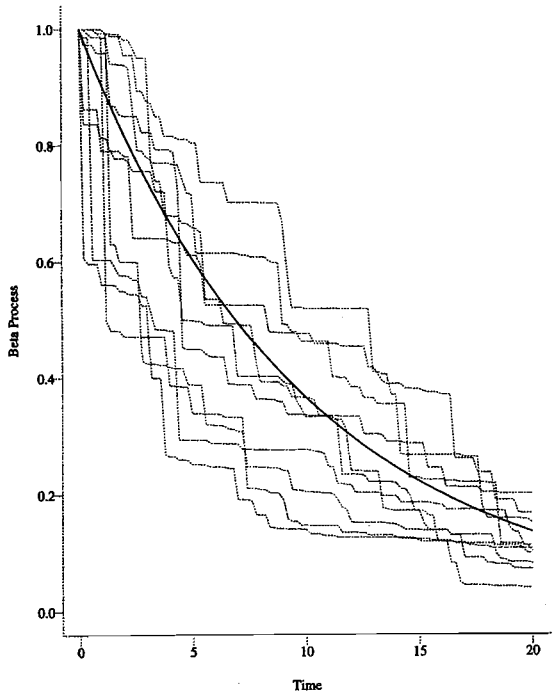

Figure 6.12 *Sample of ten sample paths (dashed lines) and their mean (solid line) for samples from a Beta process prior with* $H_0(t) = 0.1t$ *and* $c(t) = 5$.

parameters $[c(t)A_0(t) + Y(t)D(t)]/(c(t) + Y(t))$ and $c(t) + Y(t)$. Under squared-error loss the Bayes estimator of the survival function is given by

$$\tilde{S}_B(t) = \exp\left\{-\sum_{k=1}^{i}\int_{t_{k-1}}^{t_k}\frac{c(u)h_0(u)}{c(u)+Y_k} - \int_{t_i}^{t}\frac{c(u)h_0(u)}{c(u)+Y_{i+1}}du\right\} \quad (6.4.2)$$

$$\times \prod_{k:t_k\le t}\left[1 - \frac{c(t_k)h_0(t_k)+d_k}{c(t_k)+Y_k}\right]^{\Delta_k}, \quad \text{for } t_i \le t < t_{i+1}.$$

When $c(t)$ is a constant c, this reduces to

$$\tilde{S}_B(t) = \exp\left\{-\sum_{k=1}^{i}\frac{c[H_0(t_k) - H_0(t_{k-1})]}{c+Y_k} - \frac{c[H_0(t) - H_0(t_i)]}{c+Y_{i+1}}\right\}$$

$$\times \prod_{k:t_k\le t}\left[1 - \frac{ch_0(t_k)+d_k}{c+Y_k}\right]^{\Delta_k}, \quad \text{if } t_i \le t < t_{i+1}.$$

The estimator based on the Dirichlet prior has jumps at the death times and is continuous between deaths. Note that, as $c(t) \to 0$ this estimator reduces to the Kaplan–Meier estimator.

EXAMPLE 6.4

We shall illustrate these Bayesian estimators, using the data on remission duration for patients given the drug 6-MP, which was presented in section 1.2. For the Dirichlet prior, we shall use a prior guess at $S_0(t)$ of $\alpha(t,\infty)/\alpha(0,\infty) = e^{-0.1t}$. This prior estimate was chosen so that the a priori mean of the 6-MP group is the same as the control group. Our degree of belief in this prior estimate is that it is worth about $C = 5$ observations, so that $\alpha(0,\infty) = 5$ and $\alpha(t,\infty) = 5e^{-0.1t}$. For the beta process prior, we shall assume the same prior estimate of the survival function and degree of belief in the validity of this guess, so $H_0(t) = 0.1t$ and $c(t) = 5$. Figures 6.11 and 6.12 show samples of sample paths from these two priors.

From the data, we have the following information:

T_i	6	7	9	10	11	13	16	17	19	20	22	23	25	32	34	35
Y_i	21	17	16	15	13	12	11	10	9	8	7	6	5	4	2	1
d_i	3	1	0	1	0	1	1	0	0	0	1	1	0	0	0	0
c_i	1	0	1	1	1	0	0	1	1	1	0	0	1	2	1	1
Δ_i	1	1	0	1	0	1	1	0	0	0	1	1	0	0	0	0.

To illustrate the calculations, first consider a t in the interval $[0, 6)$. For the Dirichlet prior,

$$\tilde{S}_D(t) = \left[\frac{5e^{-0.1t}+21}{5+21}\right] = \left[\frac{5e^{-0.1t}+21}{26}\right],$$

whereas, for the beta process prior,

$$\tilde{S}_B(t) = \exp\left[-\frac{5[0.1(t) - 0.1(0)]}{5+21}\right] = \exp\left(-\frac{0.5t}{26}\right).$$

For a t in the interval $[6, 7)$,

$$\tilde{S}_D(t) = \left[\frac{5e^{-0.1t}+17}{5+21}\right]\left\{\frac{5e^{-0.6}+18}{5e^{-0.6}+17}\right\},$$

whereas, for the beta process prior,

$$\tilde{S}_B(t) = \exp\left\{-\frac{5[0.1(6) - 0.1(0)]}{5+21} - \frac{5[0.1(t) - 0.1(6)]}{5+17}\right\}\left[1 - \frac{5(0.1)+3}{5+21}\right].$$

Figure 6.13 shows the two Bayes estimates, the Kaplan–Meier estimator, and the prior estimate of the survival function. Here, we note that the beta process prior estimate is closer to the prior mean, which is to be expected, because the beta process has sample paths which tend to lie closer to the hypothesized prior guess at the survival function.

The third approach to Bayesian estimation of the survival function is by Monte Carlo Bayesian methods or the Gibbs sampler. This approach is more flexible than the other two approaches. For right-censored data, for which we will describe the procedure, closed form estimates of the survival function are available. For other censoring or truncation schemes, such simple estimates are not available, and the Gibbs sample provides a way of simulating the desired posterior distribution of the survival function. This approach can also be extended to more complicated problems, such as the regression problems discussed in Chapter 8.

To illustrate how this method works, we shall focus on the right-censored data problem. We let $0 < t_1 < \cdots < t_M$ be M time points. Let d_j be the number of deaths in the interval $(t_{j-1}, t_j]$ and λ_j the number of right-censored observations at t_j. Let $P_j = S(t_j)$ be the survival function at time t_j, so the likelihood function is proportional to $\prod_{j=1}^{M}(P_{j-1} - P_j)^{d_j}P_j^{\lambda_j}$. Let $\theta_j = P_{j-1} - P_j$, for $j = 1, \ldots, M$ and $\theta_{M+1} = P_M$. For a prior distribution, we assume that the joint distribution

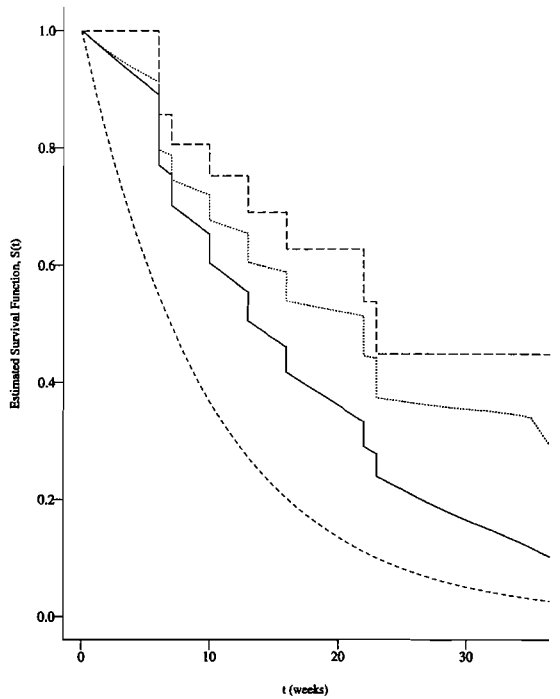

Figure 6.13 *Bayes estimates of the survival function for the 6-MP group. Beta process prior (———) Dirichlet process prior (——) Prior (– – – –) Kaplan-Meier estimate (— — —)*

of the θ's is the Dirichlet distribution with density function

$$\pi(\theta, \ldots, \theta_m) = \text{Constant} \prod_{j=0}^{M+1} (\theta_j)^{\alpha_{j-1}}, \tag{6.4.3}$$

where $\alpha_j = C[S_0(t_{j-1}) - S_0(t_j)]$ for $j = 1, \ldots, M+1$ with $S_0(t_{M+1}) = 0$ and the constant in (6.4.3) is

$$\frac{\Gamma(C)}{\prod_{j=1}^{M+1} \Gamma(\alpha_j)}.$$

The Gibbs sampling approach to Bayesian estimation approximates the posterior distribution via a Monte Carlo simulation. Here, we treat the censored observations as unknown parameters, and we simulate death times for each censored observation. Using these values with the death information, one simulates the parameters θ_j. These new θ's are used to generate new death times for the censored observations, and so forth. Gelfand and Smith (1990) have shown that this procedure converges to a realization of θ drawn from the posterior distribution θ, given the data. This process is repeated a large number of times to obtain a sample from the posterior distribution of θ, given the data which is analyzed to provide the Bayes estimator.

For our censored data problem, a single Gibbs sample is generated as follows. If $\lambda_j > 0$, let $Z_{j+1,j}, \ldots, Z_{M+1,j}$ denote the number of observations out of the λ_j that may have been deaths in the intervals $(t_j, t_{j+1}], \ldots, (t_{M-1}, t_M], (t_M, \infty)$, respectively. Note that $\lambda_j = \sum_{k=j+1}^{M+1} Z_{k,j}$. Suppose that, at the ith iteration, we have a realization of $\theta^i = (\theta_1^i, \theta_2^i, \ldots, \theta_{M+1}^i)$ which sums to 1. We sample $Z_{j+1,j}, \ldots, Z_{M+1,j}$ from a multinomial with sample size λ_j and probabilities

$$p_k = \frac{\theta_k^i}{\sum_{b=j+1}^{M+1} \theta_b^i}.$$

Having sampled the Z's, new θ's are generated from the Dirichlet by first computing

$$R_b^{i+1} = \alpha_b + d_b + \sum_{j=1}^{M} Z_{b,j}$$

and, then, sampling $\theta^{i+1} = (\theta_1^{i+1}, \theta_2^{i+1}, \ldots, \theta_{M+1}^{i+1})$ for a Dirichlet distribution with parameters $(R_1^{i+1}, R_2^{i+1}, \ldots, R_{M+1}^{i+1})$.

The procedure above yields a single realization of θ and R after i steps. Typically i is relatively small, of the order 10 or 20. This process is repeated S times where S is typically of the order 1000–10,000. The posterior estimate of θ_b is, then, given by

$$\tilde{\theta}_b = S^{-1} \sum_{s=1}^{S} \frac{R_{bs}^i}{\sum_{k=1}^{M+1} R_{ks}^i}. \tag{6.4.4}$$

EXAMPLE 6.4 *(continued)* We shall apply the Gibbs sampling approach to the data in Example 6.4. As in that example, we assume, a priori, that $S_0(t) = e^{-0.1t}$ and that our prior belief in the accuracy of our prior guess is $C = 5$ observations. Intervals are formed by taking t_j to be the death and censoring times. For a death time T, we include an "interval" $(T^-, T]$ with a θ_b representing the point mass at time T. (That is, θ_b is the jump in the estimated survival function at an observed death.) The following Table 6.5 shows the 24 intervals needed for this problem and the values of α_j from our prior.

To generate the first Gibbs observation, we generated θ_b^0, $b = 1, \ldots, 24$ from the prior distribution (6.4.3) which is Dirichlet

TABLE 6.5
Estimates Based on Gibbs Sampling

j	$(t_j - 1, t_j]$	d_j	λ_j	α_j	θ^0	*Revised Death Count Iteration 1*	*Posterior Probability (SE)*
1	$(0, 6^-]$	0	0	2.256	0.3378	0	0.0867 (0)
2	$(6^-, 6]$	3	1	0.000	0	3	0.1154 (0)
3	$(6, 7^-]$	0	0	0.261	0.0867	0	0.0105 (0.0001)
4	$(7^-, 7]$	1	0	0.000	0	1	0.0408 (0.0003)
5	$(7, 9]$	0	1	0.450	0.0228	0	0.0182 (0.0002)
6	$(9, 10^-]$	0	0	0.193	0.0001	0	0.0083 (0.0002)
7	$(10^-, 10]$	1	1	0.000	0	1	0.0430 (0.0004)
8	$(10, 11]$	0	1	0.175	0.0428	0	0.0077 (0.0002)
9	$(11, 13^-]$	0	0	0.302	0.0001	0	0.0148 (0.0004)
10	$(13^-, 13]$	1	0	0.000	0	1	0.0500 (0.0007)
11	$(13, 16^-]$	0	0	0.353	0.2673	2	0.0169 (0.0004)
12	$(16^-, 16]$	1	0	0.000	0	1	0.0492 (0.0007)
13	$(16, 17]$	0	1	0.096	0.0000	0	0.0050 (0.0002)
14	$(17, 19]$	0	1	0.166	0.0028	0	0.0091 (0.0004)
15	$(19, 20]$	0	1	0.071	0.0721	1	0.0042 (0.0003)
16	$(20, 22^-]$	0	0	0.123	0.0058	0	0.0080 (0.0004)
17	$(22^-, 22]$	1	0	0.000	0	1	0.0678 (0.0012)
18	$(22, 23^-]$	0	0	0.053	0.0045	0	0.0038 (0.0003)
19	$(23^-, 23]$	1	0	0.000	0	1	0.0662 (0.0381)
20	$(23, 25]$	0	1	0.091	0.0003	0	0.0066 (0.0005)
21	$(25, 32]$	0	2	0.207	0.1570	5	0.0183 (0.0008)
22	$(32, 34]$	0	1	0.037	0.0000	0	0.0072 (0.0008)
23	$(34, 35]$	0	1	0.016	0.0000	0	0.0117 (0.0014)
24	$(35, \infty)$	0	0	0.151	0.0000	4	0.3306 (0.0024)

$(\alpha_1, \ldots, \alpha_{24})$. To generate observations from the Dirichlet distribution, one generates $W_1, \ldots, W_{24}$ as independent gamma random variables with parameters α_b and $\beta = 1$ (i.e., $f(w_b) = w_b^{\alpha_b - 1} \exp\{-w_b\}/\Gamma(\alpha_b)$, and, then, $\theta_b = W_b / \sum W_j$. The first realization of the θ's is included in the table. Using these values, we, then, generate the Z's. For example, we generate $Z_{3,2}, Z_{4,2}, \ldots, Z_{12,2}$ from the appropriate multinomial distribution. In our example, this corresponds to picking an interval $(t_{b-1}, t_b]$ with $b > j$ in which each censored observation at t_j is to be placed and counted as a death. The table includes entries which give the revised death count at t_b of $d_b + \sum_{j=1}^{M} Z_{b,j}$, for the first iteration. These revised death counts are used to update the values of α_b by $Y_b = \alpha_b + d_b + \sum_{j=1}^{M} Z_{b,j}$, also given in Table 6.5. The procedure continues through a total of 10 cycles to produce the Gibbs iterate Y_{b1}^{10}.

This is repeated 1000 times. The final column of the table provides the posterior means of the θ's from (6.4.4) and, for reference, the sample standard errors of the standardized R_{bs}^i which provide some information on the rate of convergence of the algorithm. Notice that the posterior mean estimates in this table are precisely what we would obtain from the Dirichlet process prior, discussed earlier.

Practical Notes

1. The Bayesian estimator of the survival function obtained from a right-censored sample from the Dirichlet process prior model can be extended to other censoring schemes. Johnson and Christensen (1986) developed the estimation procedure for grouped data as found in a life table. Cornfield and Detre (1977) also consider a Bayes estimator of the survival function for life table data which is based on a Dirichlet-like prior.
2. Using the Gibbs sampling approach, additional censoring schemes can be handled quite easily. For example, Kuo and Smith (1992) show how to handle combined right- and left-censored data. This flexibility of the Monte Carlo Bayesian approach is one of the major strengths of the technique.
3. The Gibbs sampling approach presented here generates a Gibbs sample based on a large number of short runs of the algorithm. An alternative is to run a single realization of the algorithm until the successive iterations have the desired posterior distribution and, then, take, as the Gibbs sample, successive θ's generated by the algorithm. The approach suggested here, although requiring a bit more computation time, has the advantage of producing independent replications of the posterior distribution. (See Gelfand and Smith (1990) for a discussion of the merits of the two approaches.)
4. The posterior estimator of the survival function from the Gibbs sample, (6.4.4), is based on the fact that the posterior distribution of θ_b is a mixture of a beta random variable with parameters Y_b and $\sum_{k \neq b} Y_k$. An alternative technique to estimate the posterior distribution of θ_b is to use the empirical distribution function of the simulated values of θ, θ_{bs}^i, $s = 1, \ldots, S$. This would give a posterior estimator of θ_b of the sample mean of S replicates, θ_{bs}^i. To achieve the same precision as found by (6.4.4) for this approach, a larger value of S is required. By this approach, however, one can routinely provide an estimate of any functional of the posterior distribution of θ, by the appropriate functional of the empirical distribution of the simulated θ's.
5. Hjort (1992) discusses how the beta process prior can be used in more complicated censoring schemes and in making adjustments to

the survival function to account for covariates. He provides a Bayes approach to the proportional hazard regression problem discussed in Chapter 8.

Theoretical Notes

1. The Dirichlet process prior estimator of the survival function was first proposed by Ferguson (1973) for uncensored data. Susarla and Van Ryzin (1976) and Ferguson and Phadia (1979) extend the estimation process to right censored data.
2. The beta process prior was introduced in this context by Hjort (1990).
3. Both the Dirichlet and beta process prior estimates converge to the Product-Limit estimator for large samples for any nontrivial prior distribution. By an appropriate choice of the prior distribution, the Product-Limit estimator is a Bayes estimator for any n for both of these priors.
4. If one chooses $c(t) = kS_0(t)$, where $S_0(t) = \exp[-H_0(t)]$ for the weight parameter of the beta process, then, the beta process prior on H is the same as a Dirichlet process prior with parameters $S_0(t)$ and k. Thus, the beta process prior is a more general class of priors than the class of Dirichlet priors.
5. Kuo and Smith (1992) have introduced the use of Monte Carlo Bayesian methods to survival analysis.

6.5 Exercises

6.1 (a) Using the data on the time to relapse of 6-MP patients found in section 1.2, estimate the hazard rate at 12 months using the uniform kernel with a bandwidth of 6 months. Provide the standard error of your estimate.

(b) Compare the estimates obtained in part a to the estimate of $h(12)$ obtained using the Epanechnikov kernel.

(c) Repeat part b using the biweight kernel.

(d) Estimate the hazard rate at 5 months using all three kernels.

6.2 Using the data on the leukemia-free survival times of allogeneic bone marrow transplants in Table 1.4 of Chapter 1 (See Exercise 7 of Chapter 4), estimate the hazard rate at 1, 3, 5, 7, 9, 11, and 13 months using a uniform kernel with a bandwidth of 5 months. Plot your estimates and interpret the shape of the estimated hazard rate.

6.3 (a) Using the data on the infection times of kidney dialysis patients in section 1.4, estimate the hazard rate using a biweight kernel with a bandwidth of 5 months at 3 months for each of the two groups.

(b) Using the same bandwidth and kernel estimate the hazard rate at 10 months in both groups.

6.4 In section 1.7 a study of the death times (in years) and the age (in years) at transplant of 59 black female kidney transplant patients is reported. From this data, compute the patients' age in years at death or at the end of the study. The survival experience of this sample of patients is to be compared to the standard mortality rates of black females found in the 1990 U.S. census using the all-cause mortality for the U.S. population in 1990 found in Table 2.1 of Chapter 2.

(a) Estimate the cumulative relative mortality, $B(t)$, for this group of patients.

(b) Find the standard error of your estimate in part a.

(c) Estimate the excess mortality, $A(t)$, for this group of patients.

(d) Find the standard error of your estimate in part c.

(e) Plot the Kaplan–Meier estimate of the survival function, the expected survival curve, and the corrected survival curve for this group of patients.

6.5 An alternative to autologous bone marrow transplantation for leukemia is chemotherapy. Suppose that it is known that for chemotherapy patients the time from diagnosis to relapse or death has an exponential distribution with survival function hazard rate $\lambda = 0.045$. Assume that this rate is the same for all patients. To compare the survival experience of this reference population to autologous bone marrow transplant patients use the data on autologous transplants in Table 1.4 of Chapter 1 (see Problem 7 of Chapter 4).

(a) Estimate the cumulative relative mortality, $B(t)$, for this group of patients.

(b) Find the standard error of your estimate in part a.

(c) Estimate the excess mortality, $A(t)$, for this group of patients.

(d) Find the standard error of your estimate in part c.

6.6 Table 1.3 of section 1.5 provides data on the time to death (in months) of nine immunoperoxidase-positive breast-cancer patients.

(a) Using a Dirichlet prior for $S(t)$ with $\alpha(t, \infty) = 6\exp(-0.1t^{0.5})$, find the Bayes estimate of the survival function under squared-error loss.

(b) Using a beta prior for $H(t)$ with $q = 6$ and $H_0(t) = 0.1t^{0.5}$ find the Bayes estimate of the survival function under squared-error loss.

(c) Compare the estimates found in parts a and b to the usual Kaplan–Meier estimate of the survival function.

6.7 Table 1.6 of section 1.11 gives data on the times in weeks from diagnosis to death of 28 patients with diploid cancers of the tongue.

(a) Using a Dirichlet prior for $S(t)$ with $\alpha(t, \infty) = 4/(1 + 0.15t^{0.5})$, find the Bayes estimate of the survival function under squared-error loss.

(b) Using a beta prior for $H(t)$ with $q = 4$ and $H_0(t) = \ln(1 + 0.15t^{0.5})$, find the Bayes estimate of the survival function under squared-error loss.

(c) Compare the estimates found in parts a and b to the usual Kaplan–Meier estimate of the survival function.

7 Hypothesis Testing

7.1 Introduction

As we have seen in Chapters 4–6, the Nelson–Aalen estimator of the cumulative hazard rate is a basic quantity in describing the survival experience of a population. In Chapter 4, we used this estimator along with the closely related Product-Limit estimator to make crude comparisons between the disease-free survival curves of bone marrow transplant patients with different types of leukemia, and in section 6.3, we used this statistic as the basis for estimating excess mortality of Iowa psychiatric patients.

In this chapter, we shall focus on hypothesis tests that are based on comparing the Nelson–Aalen estimator, obtained directly from the data, to an expected estimator of the cumulative hazard rate, based on the assumed model under the null hypothesis. Rather than a direct comparison of these two rates, we shall examine tests that look at weighted differences between the observed and expected hazard rates. The weights will allow us to put more emphasis on certain parts of the curves. Different weights will allow us to present tests that are most sensitive to early or late departures from the hypothesized relationship between samples as specified by the null hypothesis.