7. TRENDS AND AGING

7.1 Overview

This chapter considers modeling of time trends in an initiating-event rate λ or in a probability p. Such a trend might be in terms of calendar time, or in terms of system age. Section 7.2 considers trends in λ when the events are grouped into bins, such as counts of events in calendar years. Section 7.3 considers trends in λ when the individual event times are used as the data. Section 7.4 considers trends in p. These sections all model λ or p as a parametric function of time. The final section of this chapter, Section 7.5, ties together some of the methods that have been presented in different specific applications in this chapter and in Chapter 6.

Modeling a trend normally involves some rather detailed mathematics. A Bayesian analysis must construct a sample from the posterior distribution, and a frequentist analysis must calculate the fitting equations and estimate the uncertainties in any unknown parameters. The viewpoint taken here is that the computer software will do those calculations. This chapter avoids any equations that the user will not need. Instead, this chapter presents the various approaches that are possible, and shows how to interpret the computer output and translate it, if necessary, into the quantities needed for PRA.

It turns out that the Bayesian analysis is simpler to present than the frequentist analysis. Therefore, it is given first in Chapter 7.

Some of this material is drawn from an INEEL report by Atwood (1995), and indirectly from more advanced sources. The INEEL report gives additional examples, including definitions of somewhat more complex models and discussion of pitfalls in constructing such models.

7.2 Binned Poisson Data

7.2.1 Examples

Two examples are given here. The first example was introduced as Example 6.5, unplanned demands for the HPCI system during 1987-1993. Table 6.8, summarizing that example, is repeated here for convenience as Table 7.1.

The second example, Example 7.1, groups events not by year of occurrence, but by age of the reactor.

Table 7.1 HPCI demands and reactor-critical-years (from Example 6.5).

Calendar year	HPCI demands	Reactor-critical-years
1987	16	14.63
1988	10	14.15
1989	7	15.75
1990	13	17.77
1991	9	17.11
1992	6	17.19
1993	2	17.34

Example 7.1 Thermal-fatigue leak events, by plant age.

Thermal-fatigue leaks in PWR stainless-steel primary-coolant-system piping are tabulated by Shah et al. (1998).

Age (years from initial criticality)	Number of leaks	Reactor-years
0.0 – 5.0	2	1052
5.0 – 10.0	1	982.5
10.0 – 15.0	4	756.9
15.0 – 20.0	4	442.4
20.0 – 25.0	2	230.9
25.0 – 30.0	0	43.9

The PWR plants considered here include all the western-designed PWRs, 217 reactors in all, from initial criticality until May 31, 1998, or until decommissioning. For details, see Shah et. al (1998). Plant age is summarized in 5-year bins, shown in the first column of the table. Other bins, such as 1-year bins or 10-year bins could have been constructed.

For each thermal-fatigue leak event, the age of the plant at the time of the event was calculated. The number of such events for each age bin is given in the second column of the table.

To count the reactor-years for each age, the number of reactors that experienced 1 year, 2 years, etc. were totaled. For example, Three Mile Island 2 had its initial criticality in December 1978, and was shut down in March 1979. It was counted as contributing 1/4 of a reactor-year (= 3 months) to the first bin (age 0.0 – 5.0). At the other extreme, Yankee Rowe operated from July 1961 to September 1991. It was counted as contributing 5 reactor-years in each of the six age bins. These counts of reactor-years are totaled in the third column of the table in Example 7.1.

The two examples have identical form. Each bin corresponds to a time, which can be coded numerically. In Example 6.5, each bin corresponds to a year, and could run from 1987 through 1993, or from 87 to 93, or from 0 to 6, or even −3 to +3. Any coding works, as long as it is numerical. Denote the coded time for bin i by t_i. The units of t_i are calendar years, and the different possible codes correspond to different definitions of year 0. In Example 7.1, the bins are identified most simply by the midpoints of the age bins: 2.5, 7.5, 12.5, etc. The units of t_i are age in years. They could also be identified by other numerical codes, corresponding to the smallest age in the bin or the largest age in the bin or some other code.

In both examples, each bin has a calendar time or age, an event count, and an exposure time. When considering possible trends in calendar time, we must distinguish between calendar time and exposure time, paying attention to a distinction that was not so important in the earlier chapters. In Chapter 6, exposure time was denoted by t_i, but that symbol has been given a different meaning here. Therefore, denote the event count and the exposure time for bin i by x_i and s_i, respectively. Suppose that λ has units of events per reactor-year. Then the units of exposure time must be reactor-years. During any short time interval from t to $t + \Delta t$, the exposure time Δs equals Δt times the number of reactors operating during the time period.

To avoid redundancy, this section will use only the Example 6.5 for illustrating the methods, although either example could be used.

7.2.2 Model

7.2.2.1 General Model

The assumed model is an extension of the model for a Poisson process given in Section 2.2.2. The following assumptions are made. These are a simplification of the slightly weaker assumptions given by Thompson (1981).

1. The probability that an event will occur in any specified interval with short exposure time approaches zero as the exposure time approaches zero.

2. Exactly simultaneous events do not occur.

3. Occurrences of events in disjoint time periods are statistically independent.

This model is a **nonhomogeneous Poisson process (NHPP)**. The model in Section 2.2 is a **homogeneous Poisson process (HPP)**, a special case of the model given here. Consider now a single operating system, so that exposure time equals elapsed calendar time. In the HPP, the probability of an event in the time interval (t, $t + \Delta t$) is approximately $\lambda\Delta t$. In the NHPP, λ is not constant, but is a function of t. The function $\lambda(t)$ is called the time-dependent event occurrence rate. Some authors call it the **Poisson intensity function**.

Four special cases are mentioned here.

- The HPP has $\lambda(t)$ = a constant > 0.
- The **loglinear** model has $\ln\lambda(t) = a + bt$, or equivalently, $\lambda(t) = Ae^{bt}$, with $A = e^a$. This is also called the **exponential event rate** model. Here, a and b are unknown parameters, which must be estimated from the data.
- The **Weibull process**, or **power-law event rate** model has $\lambda(t) = (b/c)(t/c)^{b-1}$, or equivalently $\lambda(t) = At^{b-1}$. Both b and c are unknown parameters (with $b > 0$, $c > 0$), to be estimated from the data. This model can be rewritten as $\ln\lambda(t) = \ln(A) + (b-1)\ln(t)$,
  which is a linear function of $\ln(t)$. Several parameterizations are found in the literature.
- An **extended power-law** process has $\lambda(t) = At^{b-1} + \lambda_0$, for $\lambda_0 \geq 0$. The Swedish I-Book (Pörn et al. 1996) uses a Bayesian analysis of

this model. The method is developed by Pörn (1990) and briefly explained by Pörn et al. (1993) and Lehtinen et al. (1997).

An occurrence rate must be nonnegative. Note that all four of these models have this requirement built in — they force $\lambda(t)$ to be positive for all t. The number of possible models is infinite, because any nonnegative function of t gives a different model.

In the first bulleted case, $\lambda(t)$ is constant, and in the other three cases $\lambda(t)$ is monotone in t, either increasing forever or decreasing forever. An event frequency has a **bathtub** shape if $\lambda(t)$ is decreasing when t is small, then roughly constant, and finally increasing when t is large. Models of a bathtub shape require more parameters. Therefore, bathtub curves are commonly used to describe models qualitatively, but have not been implemented in widely-used quantitative models.

As noted, the power-law model can be expressed as a loglinear model in $\ln(t)$, as long as t stays away from zero. Therefore, the illustrations of this section will use the loglinear model. Readers can translate this material into the power-law model by replacing t with $\ln(t)$ if they desire. The extended power-law model will not be considered further here.

It is helpful to explore the relation between this chapter's models, with non-constant λ, and the constant-λ model of the earlier chapters. Consider again a *single* system, such as a single HPCI system in Example 6.5, or a single PWR in Example 7.1. Assume that the event of interest can occur at the system, that is, the reactor is critical for Example 6.5, and the reactor is operating in Example 7.1. Then exposure time for the system equals elapsed calendar time. Under the three assumptions given above, the number of events in the system until time t is random, having a Poisson distribution with mean

$$\int_0^t \lambda(u)du \quad . \tag{7.1}$$

This agrees with the result given in Section 2.2 for the homogeneous case. In that case, $\lambda(u)$ is the constant λ, and the integral equals λt. That is the mean in Equation 2.1. For an interval that does not necessarily start at zero, the mean number of events in the time interval is the integral over the interval. If $\lambda(u)$ is not strongly curved in the interval, the integral is approximately the value of λ at the midpoint times the length of the interval.

When multiple systems are observed simultaneously, the total number of events is again a Poisson random variable, and the mean count is the sum of the means for the individual systems. This fact will be used to account for the exposure times in the examples.

The loglinear and (simple) power-law models given above are now discussed in detail.

7.2.2.2 Loglinear Model

An occurrence rate must have units of events per exposure time unit. Therefore, in the loglinear model, A has units of events per exposure time unit, and b has the inverse units of the time axis. In Example 6.5, A has units 1/reactor-critical-year, and b has units 1/calendar-year. In Example 7.1, A has units 1/reactor-calendar-year, and b has units 1/year-of-age.

The loglinear model is illustrated here, for $a = 0.1$ and $b = +1.0$, 0.0, and -1.0. Figure 7.1 shows $\ln\lambda(t)$ as a function of t, and Figure 7.2 shows $\lambda(t)$ as a function of t.

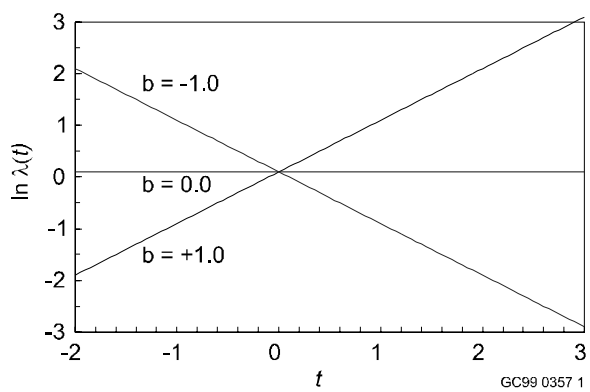

Figure 1 Loglinear model, $\ln\lambda(t) = a + bt$, for $a = 0.1$ and three possible values of b. The vertical axis shows $\ln\lambda(t)$.

The interpretation of b is the slope of $\ln\lambda(t)$. If b is negative, $\ln\lambda(t)$ is decreasing, and therefore the event occurrence rate $\lambda(t)$ is also decreasing. If b is positive, $\lambda(t)$ is increasing, and if $b = 0$, $\lambda(t)$ is constant. Tests for trend will be formulated as tests of whether $b = 0$.

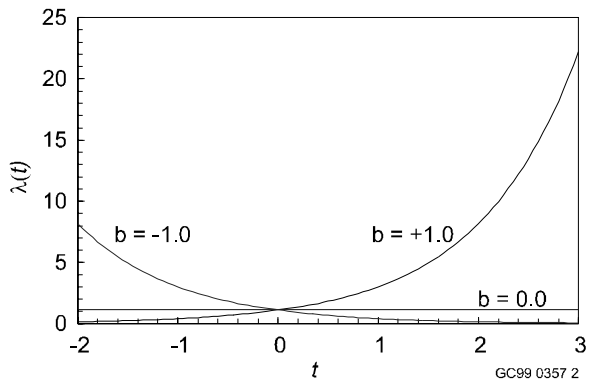

Figure 2 Same model as in previous figure, showing $\lambda(t)$ instead of $\ln\lambda(t)$.

The interpretation of a is the intercept of $\ln\lambda(t)$, that is, the value of $\ln\lambda(t)$ at $t = 0$. The meaning of a depends on how time is coded. In Example 6.5, with HPCI demands, if t runs from 1987 to 1993, the value $t = 0$ corresponds to about 2000 years ago, and a is the value of $\ln\lambda(t)$ at that time. This is an enormous extrapolation, and a can be estimated only with great uncertainty. Coding t as running from 87 to 93 involves less extrapolation, because now $t = 0$ corresponds to the year 1900, only some 90 years before the data. Other possibilities are to let t run from 0 to 6, or from -3 to $+3$. These coding schemes involve no extrapolation at all, because 0 is included in the range of the observed data.

In theory, it makes no difference which coding system is used. The different codings for t and the different meanings of a compensate for each other. For any particular time, such as 1992 or 1996, the different coding systems give exactly the same estimate of λ at that time. In practice, however, no computed value is exact, and roundoff errors can accumulate. Use of large extrapolations can introduce large errors. Well-programmed software should protect against this problem automatically, no matter how the times are entered. Nevertheless, the analyst can do no harm by choosing a coding system with zero reasonably close to the data.

7.2.2.3 Power-Law Model

In both parameterizations given above for the power-law model, b is a unitless shape parameter. As will be seen below, $b - 1$ is the slope of $\ln\lambda(t)$ as a function of $\ln(t)$. In the first parameterization, c is a scale parameter with units of t. It does not have a simple interpretation. In the second parameterization, A has strange units, but a simple interpretation as the numerical value of $\lambda(t)$ at $t = 1$.

Figures 7.3 and 7.4 illustrate the power-law model for $A = 1.0$, and $b = 0.5$, 1.0, 2.0, and 3.0.

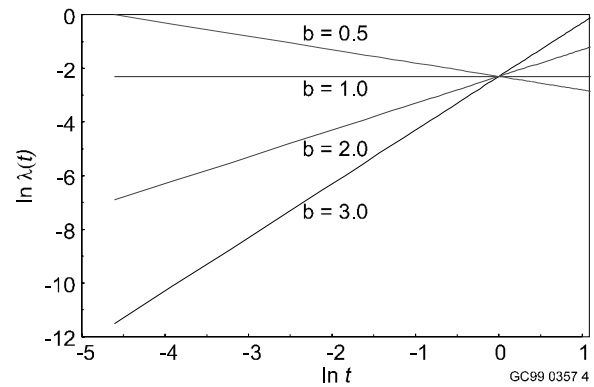

Figure 3 Power-law model, showing $\ln\lambda(t)$ as a linear function of $\ln(t)$, with $A = 1.0$ and several values of b.

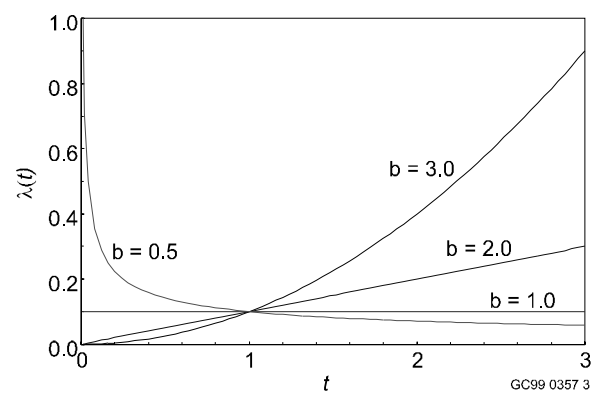

Figure 4 Same model as in previous figure, with $\lambda(t)$ shown as function of t.

In Figure 7.3, the parameter $b - 1$ is the slope of $\ln\lambda(t)$ as a function of $\ln(t)$. In Figure 7.4, b is a shape parameter, defining the shape of the curve. In either figure, the interpretation of A is the numerical value of $\lambda(t)$ at $t = 1$.

This model requires $t \geq 0$, so a coding system should be chosen so that all the observed times correspond to nonnegative values of t. Also, if the occurrence rate is

decreasing, the modeled occurrence rate becomes infinite at $t = 0$.

The loglinear and power-law models are widely used, but they are chosen for their simplicity and convenience, not their theoretical validity. Any model must be checked for goodness of fit. Moreover, no model should be extrapolated far into the future — even if some convenient algebraic formula fits a trend well in the past, that is no guarantee that the data will continue to follow that formula in the future.

7.2.3 Bayesian Estimation with Loglinear Model

The bins must be small enough that $\lambda(t)$ is approximately a straight line function within each bin, not strongly curved within the bin. Denote the midpoint of the ith bin by t_i. Then the expected number of events in the bin is well approximated by $\lambda(t_i)s_i$, where s_i is the exposure time for the bin. The method is to fit the observed Poisson counts to $\lambda(t_i)s_i$, while assuming that $\lambda(t_i)$ has some parametric form.

The data are Poisson counts, but it is useful to recognize what happens as the observed values become large. As this happens, the form of the likelihood function approaches the form of a normal density in the two variables a and b. This is a general fact for large data sets that is exploited by advanced statistics texts, such as Cox and Hinkley (1974, Section 10.6). Therefore, with large data sets the **conjugate** prior is normal: if a and b are given a normal prior distribution, the posterior will be approximately normal, with very good approximation as the data set becomes large. The corresponding **noninformative** prior for a and b is the limiting case as the variance approaches infinity, which is a constant density.

For the work here, it will be assumed that a computer program produces a sample from the posterior distribution. The theory sketched above then leads us to the conclusion that the posterior distributions not only *appear* normal, they really *are* normal, or very close to normal.

A convenient software package is BUGS (1995), Bayesian inference Using Gibbs Sampling. The Windows version is called WinBUGS. It is also described in Section 8.2.3, and documented by Spiegelhalter et al. 1995. It is currently available for free download at

http://www.mrc-bsu.cam.ac.uk/bugs/ .

WinBUGS allows the use of a graphical description of the model, called a "directed graph," although the text script may be easier to use. Example 6.5 will be analyzed using WinBUGS here, assuming a loglinear model.

With the data of Example 6.5 (Table 7.1), BUGS was used to model $\lambda(i) = \exp(a + bi)$, for i from 1 to 7. Then $X(i)$ was modeled as having a Poisson distribution with mean $\lambda(i)\times s(i)$. Finally, a and b were given very diffuse prior normal distributions. Figure 7.5 is the logical diagram showing the relations.

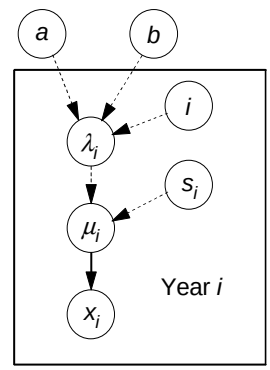

Figure 5 Directed graph for analysis of Poisson trend in Example 6.5.

In Figure 7.5, deterministic relations are shown by dashed arrows, and stochastic relations (random number generation) are shown by solid arrows.

Figure 7.6 shows the BUGS script that was used.

One idiosyncrasy of BUGS is that it parameterizes the normal distribution in terms of the precision $\tau = 1/\sigma^2$. The reasons are explained in Section 6.7.1.2.1. Therefore, a precision of 0.0001 in the script corresponds to a standard deviation of 100. That gives a very diffuse distribution.

The script was executed with 4 separate starting values of a and b, generating 4 **chains** of values, each 10000 elements long. The first 200 elements of each chain were discarded, and the remaining 39200 elements were used to estimate the posterior

distributions. Table 7.2 summarizes the posterior distributions for *a* and *b*. When interpreting these summaries, be aware that *a* and *b* are not independently distributed.

```
model
{
   for (i in 1:N) {
      lambda[i] <- exp(a + i*b)
      mu[i] <- lambda[i]*s[i]
      x[i] ~ dpois(mu[i])
   }
   a ~ dnorm(0.0, 0.0001)
   b ~ dnorm(0.0, 0.0001)
}
```

Figure 6 BUGS script for analyzing data of Example 6.6.

Table 7.2 Posterior statistics for *a* and *b*, for loglinear model of Example 6.5.

	a	*b*
mean	0.264	−0.237
median	0.270	−0.237
st. dev.	0.251	0.067
5th percentile	−0.157	−0.348
95th percentile	0.662	−0.129

Even though the numbers are not necessarily accurate to three places, the table shows that the mean and median are nearly equal in each case, and the 5th and 95th percentiles are approximately 1.645 standard deviations from the mean. That is, the distributions appear to be approximately normal. (BUGS supplies graphical estimates of the densities, which also appear normal.) This is consistent with the theory mentioned earlier.

The mean of *b* is negative, and 3.5 standard deviations away from the mean. This is very strong evidence of a downward trend. The posterior belief in a flat or rising trend is only 2.3E−4. (This is $\Phi(-3.5)$, from Table C.1.)

The posterior distribution of λ is lognormal in any year. It is shown in Figure 7.7. The median is plotted as a solid line and the 5th and 95th percentiles are shown as dashed lines. The simple point estimates

$$\hat{\lambda} = x_i \,/\, s_i$$

are plotted as dots.

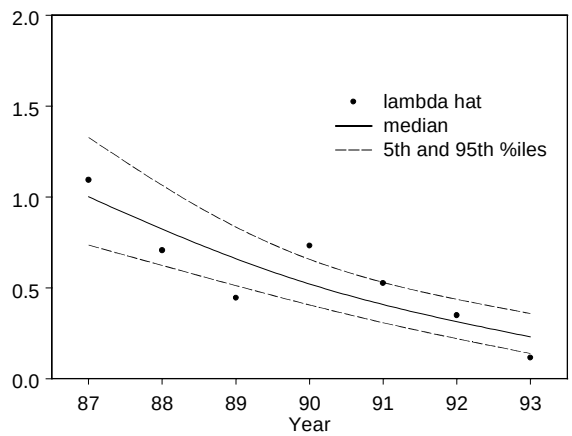

Figure 7 Posterior distribution of λ, assuming exponential trend in Example 6.5.

The correct interpretation of such figures is subtle. For *any particular year*, the value of $\lambda(t) = \exp(a + bt)$ is between the two dotted lines with 90% probability. However, for two distinct times t_1 and t_2, a pair (a, b) that puts $\lambda(t_1)$ between the lines may put $\lambda(t_2)$ outside the dotted lines. Therefore, the entire curve

$$\{ (t, \lambda(t)) \mid -\infty < t < \infty \}$$

does not fall between the two dotted lines with 90% probability. A 90% region for the entire curve would need to be wider than the band shown in Figure 7.7. This issue is revisited for frequentist estimation in Section 7.2.4.5.

7.2.4 Frequentist Estimation with Loglinear Model

The frequentist method has several variations, which have been implemented in various software packages. They are presented here, applied to the example of Table 7.1, and the results are compared to each other and to the Bayesian results.

Assume a loglinear model, $\ln\lambda(t) = a + bt$. Different statistical software packages use different terminology, summarized here.

- **General linear model**: the mean of the observable random variable is a linear function of unknown parameters. This is NOT useful for the present problem. It is mentioned only to point out the possible confusion with the general*ized* linear model below.
- **Loglinear model**: the logarithm of the mean of the observable random variable is a linear function of unknown parameters. This is exactly the model considered in this section.
- **Generalized linear model**: a transformation of the mean of the observable random variable is a linear function of unknown parameters. This includes the loglinear model as a special case, when the transformation is chosen to be the logarithm.

7.2.4.1 Point Estimation

Analysis of the loglinear model finds the maximum likelihood estimates (MLEs) of a and b, based on the Poisson counts. The discussion below will sometimes call this the **Poisson-maximum-likelihood** method. The software user does not need to understand the details, beyond knowing that the solution is by numerical iteration, not by algebraic solution of an equation.

This approach is applied here to Example 6.5. No calculations are given for Example 7.1 because they would be very similar, differing only by using age instead of calendar year.

A loglinear model was fitted to the HPCI demand data of Table 7.1 (Example 6.5). When the years were coded as 87 through 93, the estimates of a and b were 20.5305 and −0.2355. The second number is the slope of $\ln\lambda(t)$, and the first is the fitted value of $\ln\lambda(t)$ when $t = 0$, that is, in the year 1900. (Of course, no HPCI systems existed in 1900, but the model does not know that and fits a value.) When, instead, the years are coded from 0 to 6, the slope is the same, but the intercept parameter is different, because now the value $t = 0$ is the year 1987. The estimate of a, the intercept, is 0.0389, the fitted value of $\ln\lambda(t)$ for 1987.

The fitted value of $\lambda(t)$ is the same, whichever coding method is used. The fitted values are shown in Figure 7.8, overlaid on Figure 6.14. Each point and vertical confidence interval is based on data from a single year, but the fitted trend uses all the data.

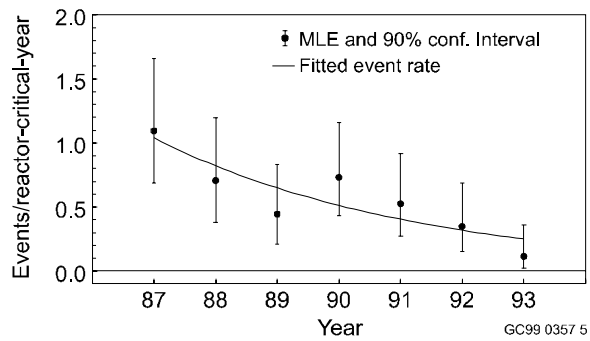

Figure 8 Frequency of unplanned HPCI demands, from Figure 6.19, with exponentially decreasing fitted trend line overlaid.

7.2.4.2 Confidence Intervals for *a* and *b*

With the point estimates $\hat{a}$ and $\hat{b}$, almost all software will also report standard errors, estimates of the standard deviations of the estimators. The estimators are assumed to be approximately normally distributed. This is a good approximation if the number of observations in each bin is not too small. (One old rule of thumb is that the count in most bins be at least 5. This is based on the observation that each bin's Poisson distribution is approximately normal if the mean is 5 or more. This old rule of thumb is sufficient, but perhaps unnecessarily conservative.)

A $100(1 - \alpha)\%$ confidence interval for b is

$$\hat{b} \pm z_{1-a/2} se(b)$$

where $\hat{b}$ is the estimate, and $z_{1-\alpha/2}$ is the $1 - \alpha/2$ quantile of the normal distribution. For example, for a 90% confidence interval, α equals 0.1 and the 0.95 quantile is 1.645. The term $se(b)$ is the corresponding standard error of b, the estimated standard deviation of the estimator. This is sometimes written st.dev.($\hat{b}$).

As mentioned in Appendix B, tests for hypotheses about b are intimately related to confidence intervals for b. The hypothesis

H_0: $b = b_0$

is rejected in favor of the hypothesis

H_1: $b \neq b_0$

at significance level α if and only if the $100(1 - \alpha)\%$ confidence interval does not contain b_0. In particular, the hypothesis

H_0: $b = 0$,

the hypothesis of no trend, is rejected at level 0.10 if the 90% confidence interval for b is entirely on one side of 0. The hypothesis is rejected at level 0.05 if the 95% confidence interval is entirely on one side of 0, and so forth. Most software packages print out a significance level at which H_0: $b = 0$ is rejected, the p-value for the trend.

A confidence interval for a is constructed in a similar way, but is normally much less interesting. Who cares what value a has? That parameter is just the intercept at some arbitrarily coded time with $t = 0$. The parameter a is of interest only because it can be used to construct confidence intervals for $\ln\lambda(t) = a + bt$. Nevertheless, software packages typically print a significance level at which the hypothesis $a = 0$ is rejected.

7.2.4.3 Test for Presence of Trend

Let two hypotheses be defined by:

H_0: $\lambda(t)$ is constant
H_1: $\lambda(t) = \exp(a + bt)$, $b \neq 0$

The loglinear model is used as an illustration for the alternative hypothesis, but any other specific model could be used, as long as it is not constant.

If we assume $\lambda(t) = \exp(a + bt)$, then the null hypothesis H_0 is false if and only if b is nonzero. Therefore, the test of H_0 is the same as a test that $b = 0$, the test given above.

Section 6.2.3.2.2 gave tests for the presence of a trend, illustrated with Example 6.5. The only test it gave with binned data was the chi-squared test of:

H_0: $\lambda(t)$ is constant
H_1: $\lambda(t)$ is not constant

Section 6.2.3.2.2 commented that the test is not very powerful, because it considers such a broad class of possibilities as the alternative hypothesis.

In Example 6.5, the chi-squared test rejected the hypothesis of constant λ with p-value 0.009. The present test of $b = 0$ rejects this hypothesis with p-value 0.0004. Although both tests reject H_0, the test based on the loglinear model finds stronger evidence against constant λ than the chi-squared test did. In an example with a less clear trend, the test based on $b = 0$ might find a statistically significant trend when the chi-squared test did not.

Suppose that λ were not constant, but went up and down in an irregular way with no persistent increasing or decreasing trend. The chi-squared test might discover this, but the test based on b would not discover the nonconstancy of λ — more precisely, the test of $b = 0$ might "discover" the nonconstancy because the random data might appear to indicate a trend in spite of the true non-trending pattern of λ, but this would only be an accident. The test to use depends on the alternatives that the analyst regards as credible. A test that focuses on those alternatives will be more powerful than a test that is designed for different alternatives.

7.2.4.4 Confidence Interval for $\lambda(t)$ at Fixed t

Most software packages also can find approximate confidence intervals for $\ln\lambda(t)$ at particular values of t. It is worthwhile understanding the approach, because the software output may require modification to coincide with the analyst's needs. The idea is that the MLEs $\hat{a}$ and $\hat{b}$ are approximately normally distributed. The software finds an approximate $100(1 - \alpha)\%$ confidence interval for $\ln\lambda(t)$ as

$$\hat{a} + \hat{b}t \pm [z_{1-\alpha/2} \times se(a + bt)] \qquad (7.2)$$

where $z_{1-\alpha/2}$ is as defined earlier, and $se(a + bt)$ is the standard error, the estimated standard deviation of $\hat{a} + \hat{b}t$. The standard error depends on the value of t. It is found by the software — it cannot be found in a naive way from the standard errors of a and b, because the MLEs $\hat{a}$ and $\hat{b}$ are correlated, not independent.

Expression 7.2 is a confidence interval for the logarithm of $\lambda(t)$. The confidence interval for $\lambda(t)$ itself is found by taking the exponential of the two bounds in Expression 7.2.

Understanding this algebraic form may be useful. For example, suppose that the software insists on giving only a 95% confidence interval for $\lambda(t)$, and the analyst desires a 90% interval instead. The following modification can be made. The α corresponding to a 95% confidence interval is 0.05. First, take logarithms of the reported upper and lower confidence limits for $\lambda(t)$. Use these two values, and the form of Expression 7.2, to find

$$z_{0.975} \times se(a + bt) \ .$$

This follows from the fact that a 95% confidence interval corresponds to $1 - \alpha/2 = 0.975$. Using $z_{0.975}$ and $z_{0.95}$ from a table of the normal distribution, calculate the value of

$$z_{0.95} \times se(a + bt) \ .$$

From this, calculate the 90% confidence interval for $\lambda(t)$,

$$\exp[\hat{a} + \hat{b}t \pm z_{0.95} se(a + bt)] \ . \qquad (7.3)$$

7.2.4.5 Simultaneous Confidence Band at All *t*

The above confidence interval is for a particular t. For many values of t, many such confidence intervals could be calculated. Each is a valid confidence interval, but they are not simultaneously valid. This is a subtle point. To appreciate it, recall the interpretation of the 90% confidence interval for $\lambda(t)$ for some particular time t_1:

Pr[confidence interval for $\lambda(t_1)$ contains true occurrence rate at time t_1] = 0.90. (7.4)

Here, the data set is thought of as random. If many data sets could be generated from the same set of years, each data set would allow the calculation of a confidence interval for $\lambda(t_1)$, and 90% of these confidence intervals would contain the true occurrence rate.

A similar confidence statement applies to each time. The simultaneous statement would involve

Pr[confidence interval for $\lambda(t_1)$ contains true occurrence rate at time t_1 AND
confidence interval for $\lambda(t_2)$ contains true occurrence rate at time t_2 AND
so forth]. (7.5)

This probability is hard to quantify, because the intervals are all calculated from the same data set, and thus are correlated. However, Expression 7.5 is certainly smaller than 0.90, because the event in square brackets in Expression 7.5 is more restrictive that the event in brackets in Equation 7.4.

This problem is familiar in the context of least squares fitting. For example, Neter and Wasserman (1974) discuss it, and attribute the solution to Working, Hotelling, and Scheffé. A simple adaptation to the present setting is sketched by Atwood (1995, App. B-7). The simultaneous confidence band is obtained by replacing $z_{0.95}$ in Expression 7.3 by

$$[\chi^2_{0.90}(r)]^{1/2} \ ,$$

where the expression in square brackets is the 90th percentile of a chi-squared distribution with r degrees of freedom. Here, r is the number of unknown parameters, 2 in this example.

The simultaneous 90% confidence band is about 30% wider than the band of 90% confidence intervals, because

$$[\chi^2_{0.90}(r)]^{1/2} = 2.15,$$

which is about 30% larger than

$$z_{0.95} = 1.645$$

when $r = 2$.

Figure 7.9 again uses the HPCI unplanned-demand data of Example 6.5. The annual estimated event frequencies are shown, along with the fitted frequency, the simultaneous 90% confidence band on the frequency, and the band constructed from the individual 90% confidence intervals.

Simultaneous confidence bands typically are not calculated by software packages. They can be calculated by the user, however, based on the formulas

above and the information produced by the software package.

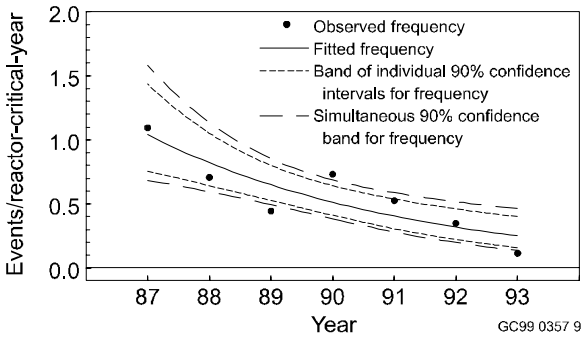

Figure 9 Simultaneous 90% confidence band and band formed by individual 90% confidence intervals for Poisson event rate, $\lambda(t)$.

Which should be presented to users, the simultaneous band or the band of individual confidence intervals? If the user's interest is in a single time, such as the most recent time, then clearly the confidence interval at that time is of greatest interest. If, on the other hand, the user will look at the entire plot, for example to judge the existence of a trend, then the simultaneous confidence band is a better indication of the uncertainty in the estimated line. To satisfy both types of users, the graph could show the simultaneous confidence band and the confidence interval at the time of greatest interest. Figure 7.10 shows such a plot, assuming that a user would be most interested in the most recent time, 1993.

7.2.4.6 Alternative Using Least-Squares Software

Since the model assumes

$$\ln\lambda(t) = a + bt ,$$

one might decide simply to use least-squares software as follows. First, estimate λ based for each bin, based on only the data for that bin;

$$\hat{\lambda}_i = x_i / s_i$$

Then fit $\ln \hat{\lambda}_i$ to $a + bt_i$ by least squares. In principle, this works. In practice, the method has several twists in the road, described next.

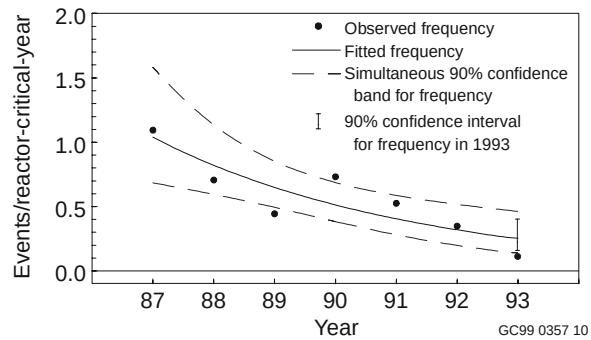

Figure 10 Simultaneous 90% confidence band for $\lambda(t)$, and 90% confidence interval for one frequency of special interest, $\lambda(93)$.

First, if the observed count is zero in any bin, the MLE $\hat{\lambda}$ will be zero for that bin, and the logarithm will be undefined. This is the case for the final bin of Example 7.1. The following ways around this have been proposed.

- Instead of estimating λ_i by x_i/s_i, use $(x_i + ½)/s_i$. This is equivalent to replacing the MLE by the posterior mean based on the Jeffreys noninformative prior.
- Estimate λ_i by the posterior mean based on the constrained noninformative prior. In this case, the constraint could be that the prior mean equals the observed overall mean $\Sigma_i x_i/\Sigma_i s_i$, or equals a modification to guarantee a positive number, $(\Sigma_i x_i + ½)/\Sigma_i s_i$.

Such ways tend to reduce the trend slightly, because they add a constant to all the failure counts, slightly flattening out any original differences.

The second point that must be considered is that the variance of X_i/s_i is not constant. For simplicity, this issue will be explained for the case with no zero counts, and with $\ln\lambda_i$ estimated by $\ln(\hat{\lambda}_i) = \ln(x_i/s_i)$. The variance of $\ln(X_i/s_i)$ is approximately the **relative variance** of X_i/s_i, defined as $\mathrm{var}(X_i/s_i)/E^2(X_i/s_i)$. This is $1/E(X_i) = 1/(\lambda_i s_i)$ if X_i has a Poisson$(\lambda_i s_i)$ distribution. Ordinary least-squares fitting has some optimality properties if the variance is constant. Otherwise, it is more efficient to use weighted least squares, with weights inversely proportional to the variances of the observations.

Unfortunately, the variances depend on the λ_i values, which are unknown. Therefore, the following **iteratively reweighted least-squares** method can be used. Begin by assuming that λ is constant, and fit $\ln(x_i/s_i)$ to a straight line with weighted least squares, and weights s_i. Calculate the resulting estimates of λ_i, $\hat{\lambda}_i = \exp(\hat{a} + \hat{b}t_i)$. Then refit the data to a straight line, using weighted least squares, and weights $\hat{\lambda}_i s_i$. Repeat this process until the estimates stabilize.

The final point is that least-squares fitting typically assumes that the data are approximately normally distributed around the straight line. In the present context, this means that $\ln(X_i/s_i)$ is assumed to be approximately normally distributed. This assumption is acceptable, unless the mean count is close to zero. However, the variance of the normal distribution is then estimated from the scatter around the fitted line. This differs from the typical treatment of Poisson data, where the mean determines the variance.

A 90% confidence interval for $\lambda(t)$ at a particular t is given by

$$\exp\{\hat{a} + \hat{b}t \pm [t_{0.95}(d) \times se(a + bt)]\} \qquad (7.6)$$

where $t_{0.95}(d)$ is the 95th percentile of Student's t distribution with d degrees of freedom. The software will report the value of d. It is the number of bins minus the number of estimated parameters, 7 – 2 in Example 6.5. The form of this equation is very similar to the form of Equation 7.3, although the estimates and standard deviation are calculated somewhat differently.

A simultaneous 90% confidence band has the same form, but the multiplier $t_{0.95}(d)$ is replaced by

$$[2F_{0.90}(r, d)]^{1/2},$$

where $F_{0.90}(r, d)$ is the 90th percentile of the F distribution with r and d degrees of freedom. This modification of Equation 7.6 is analogous to that for Equation 7.3 to get a simultaneous confidence band.

7.2.5 Comparison of Methods

The three frequentist methods are compared here. Following that comparison, the Bayesian method is compared to the frequentist methods, first for the current example and then in general.

Figure 7.11, from Atwood (1995), shows results from three frequentist analyses of the data of Table 7.1. As can be seen in the figure, the fitted lines are similar for all three analyses. The unweighted-least-squares method gives an unnecessarily wide confidence band. This shows the inefficiency of unweighted least squares when the variance is not constant. The Poisson-maximum-likelihood approach has a slightly narrower confidence band than does the weighted-least-squares approach. The reason is that the least-squares approach introduces an additional parameter that must be estimated, the variance around the line. The pure Poisson approach calculates the variance from the fitted mean. The price of estimating an extra parameter is a larger multiplier, $[2F_{0.90}(2, 5)]^{1/2} = 2.75$ instead of $[\chi^2_{0.90}(2)]^{1/2} = 2.15$.

One other reason for the difference between the two plots on the right of the figure might be that they use different estimators of the variance. In this example, however, the scatter around the line agrees almost perfectly with the scatter predicted by the Poisson assumptions, so the two calculations of the variance agree. This is coincidence, but it eliminates a possible distraction in comparing the two methods.

In summary, all three calculations are valid, and yield similar fitted lines. The method of unweighted least squares uses the data in a somewhat inefficient way, however, and therefore produces an unrealistically wide confidence band. The method of iteratively reweighted least squares provides a narrower confidence band, and the loglinear model provides the narrowest band of all. None of the calculations is exact: the least-squares method treats ln(count/time) as approximately normal, and the Poisson-maximum-likelihood method treats the parameter estimators as approximately normal. However, the analysis based on the Poisson-maximum-likelihood method is preferable (if the Poisson assumption is correct!), because it gives the tightest confidence band, and reweighted least squares is second best. It may be that extra sources of variation, or "overdispersion," have entered the data, variation beyond that caused by the Poisson distribution. If so, reweighted least squares would be best and the Poisson-maximum-likelihood method would produce an unrealistically narrow band.

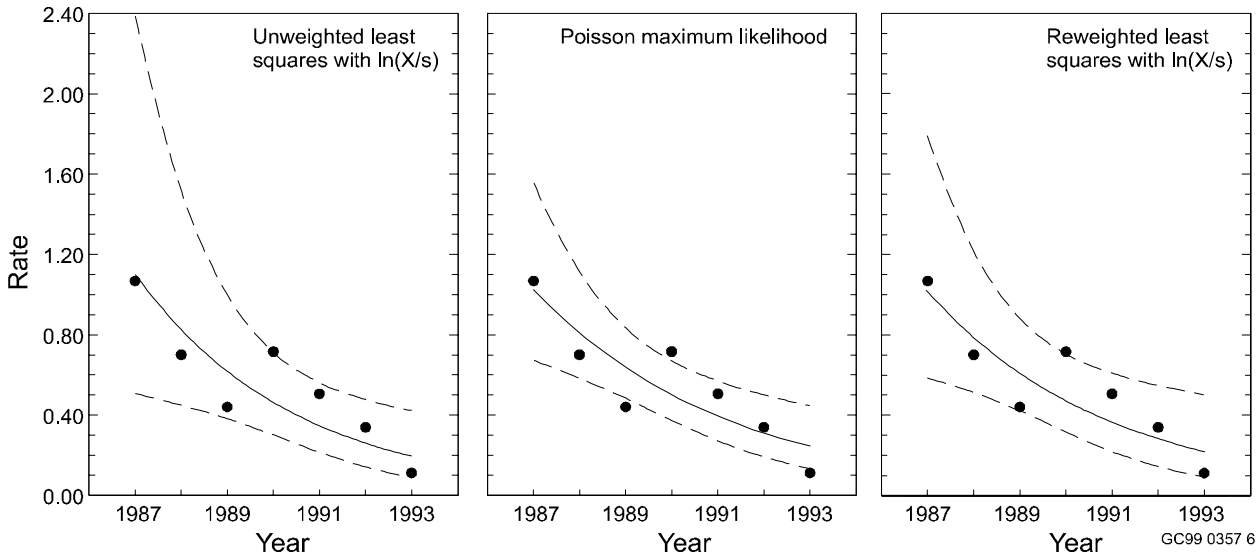

Figure 11 Fitted Poisson event occurrence rate and simultaneous 90% confidence band, based on three ways of fitting the HPCI unplanned-demands data.

Now the frequentist methods are compared with the Bayesian method, first for the particular example, and then in general.

The corresponding figure from the Bayesian analysis is Figure 7.7. Careful examination of the figures shows the following.

- The posterior median in Figure 7.7 is close to the fitted line (the MLE) in Figure 7.9, which is the middle panel of Figure 7.11.
- The 90% credible band in Figure 7.7 shows a band that is valid for any one time, but not simultaneously for all time. It is close to the inner band in Figure 7.9, which also is valid for any one time but not simultaneously. These bands are somewhat narrower than the simultaneous bands of Figure 7.11.

The following comments apply in general, not just to the example.

- Frequentist estimation relies on approximate normality of the estimators, and therefore does not work well with data having few observed events. Bayesian estimation obtains normal posteriors when the data set is large, but does not fail entirely when the data set is small – it merely obtains a different, non-normal, posterior.
- Most frequentist software packages for analyzing trends include calculations for investigating the goodness of fit of the model, as will be seen in Section 7.2.6. Current Bayesian software may neglect this issue of model validation.

7.2.6 Model Validation

The three assumptions for a nonhomogeneous Poisson process are given at the beginning of Section 7.2.2. The first assumption is difficult to test from data. The second, dealing with common cause failures, has been addressed in Chapter 6. The third assumption is that event counts in disjoint intervals are independent. This was addressed using a test of serial correlation in Chapter 6 when λ is constant, but the analogue is too complicated to consider here. When the data are collected into bins, as assumed here, serial dependence can result in an unusually high or low event count in a single bin. This can be discovered by goodness-of-fit tests, considered below.

The final assumption made when fitting a trend model is the form of the model. Methods to examine the goodness of fit will be illustrated with the loglinear model, following the precedent of the rest of this section.

7.2.6.1 Graphical Check for Goodness of Fit

The natural graphical check is to compare the observed values to the fitted line. This is illustrated by Figure 7.8. In that figure, the 90% confidence interval for each year overlaps the fitted trend line. Because no year deviates strongly from the overall trend, the data appear consistent with the assumption of an exponential trend. Even if one 90% interval had failed to overlap the fitted trend line, one would not necessarily conclude that the exponential-trend assumption is violated. The reason for not being concerned about a single failure to overlap is that some 90% confidence intervals are expected to miss the true value. In the long run, as many as 10% of the intervals may fail to contain the true value. In the short run, one miss in seven is 14%, so one miss in the graph is not alarming.

The above discussion is written from a statistical viewpoint. An engineering viewpoint may reveal more. For example, if the estimates for the individual bins (plotted as dots in this section) drop very suddenly, it may be that the mechanism has changed. If the time bins correspond to plant age, the frequent early events may correspond to a learning period. Such conjectured causes should be investigated, and confirmed or rejected based on more detailed study of the events. As is typical, a statistical analysis only puts up road signs, pointing to interesting subjects for engineering investigations.

The statistical analysis is illustrated with an example here. The examples given earlier in this chapter could be used, but Example 2.1 is more interesting when investigating lack of fit.

Example 2.1 stated that a particular plant had 34 unplanned reactor trips while at power in 1987-1995. Table 7.3 gives the dates of those initiating events. This data set is a portion of the database used by Poloski et al. 1999a. This particular plant had its initial criticality on 1/3/87 and its commercial start on 5/2/87.

These events were grouped by calendar year. Because reactor trips only occur when the reactor is at power, the relevant normalizing time is critical time, given in Table 7.4 as critical years.

The now-familiar picture is given in Figure 7.12. This figure shows that the first observed value, for 1987, is well above the fitted line, and the second observed value, for 1988, is well below the fitted line. In fact, the assumed model seems to be forcing data with an L-shaped trend into an exponentially decreasing trend.

Table 7.3 Dates of initiating events at one plant, 1987-1995. (from Example 2.1)

01/21/87	04/03/87	06/17/87	11/08/87	02/07/89	07/15/92
01/22/87	04/12/87	06/21/87	03/09/88	02/22/89	07/17/92
02/27/87	04/14/87	06/22/87	10/14/88	03/14/89	10/12/95
03/11/87	04/21/87	07/09/87	10/30/88	10/09/89	11/05/95
03/13/87	04/22/87	08/04/87	01/16/89	06/03/91	
03/31/87	05/24/87	11/07/87	02/06/89	07/12/92	

Table 7.4 Initiating events and reactor-critical-years.

Calendar year	Initiating events	Reactor-critical-years
1987	19	0.70936
1988	3	0.75172
1989	6	0.79482
1990	0	0.89596
1991	1	0.81529
1992	3	0.75123
1993	0	0.99696
1994	0	0.82735
1995	2	0.83760

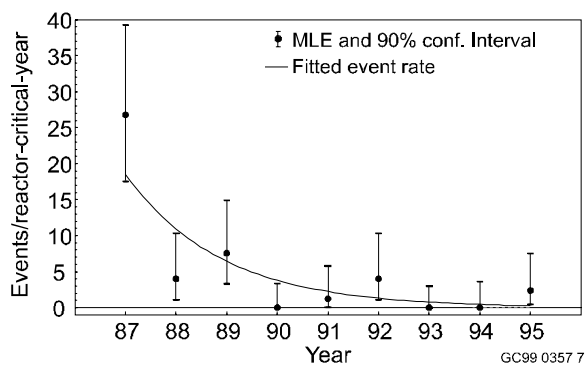

Figure 12 Annual frequency of initiating events, with fitted exponentially decreasing $\lambda(t)$ and simultaneous 90% confidence band on $\lambda(t)$.

It appears that the plant had a learning period, during which initiating events were quite frequent, followed by a period with a much smaller frequency. Examination of Table 7.4 shows that the learning period seems to have lasted until the summer of 1987 (three events in June, one each in July and August, and only infrequent events after that). It is not certain that the explanation is "learning" in the usual sense, but it is clear that the event frequency dropped suddenly about six months after the initial criticality.

Incidentally, a plot based on Bayesian calculations would show the same general information as Figure 7.12. Replace the MLE fitted line by the posterior median, and replace each confidence interval by a credible interval for λ based on a noninformative prior and one year's data. For an example of such a Bayesian plot, see Figure 7.18 in Section 7.4.3.

In ordinary least-squares fitting, it is standard to plot **residuals**, where each residual is defined as the observed value minus the fitted value. Under the assumed model, the residuals do not all have the same variance, so sometimes the **standardized residuals** are plotted, where a standardized residual is the residual divided by its theoretical standard deviation.

In the present context, the ith count, X_i, is assumed to be Poisson with mean $s_i\lambda(t_i)$. The ith residual, denoted r_i, is

$$r_i = x_i - s_i\hat{\lambda}(t_i) .$$

The variance of a Poisson random variable equals the mean. Therefore, the standardized residual is

$$r_i / \sqrt{s_i\hat{\lambda}(t_i)} .$$

In the context of binned data, these are also sometimes called the **Pearson residuals** or the **chi-squared residuals**, because the sum of the squared Pearson residuals is equal to the Pearson chi-squared statistic. A plot of the these residuals against time may be helpful.

Figure 7.13 plots the standardized residuals against calendar year, for the example of Tables 7.3 and 7.4. This plot shows something that may not be evident from Figure 7.12. The largest value corresponds to 1995, not 1987. This reflects the fact that in Figure 7.12, the 1995 confidence interval is farthest from the fitted line, in relative terms.

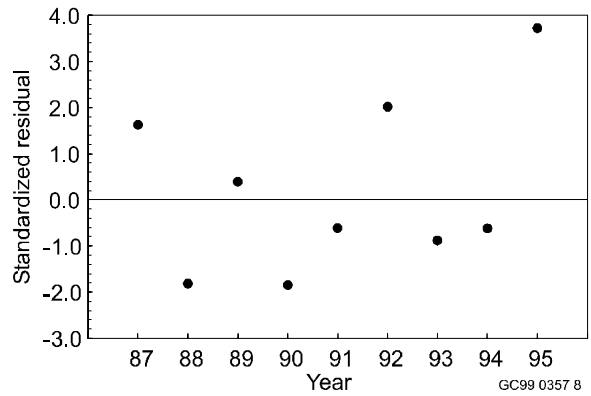

Figure 13 Standardized residuals, also called the Pearson chi-squared residuals, for data of Figure 7.12.

An informative plot for this example is the simple cumulative plot, introduced in Chapter 6 (see Figure 6.15). This plot would normally be used to check on whether the event occurrence rate is constant. The slope is the event rate, and a nonconstant slope corresponds to a departure from a straight line.

The cumulative event count is plotted against event date in Figure 7.14. In this example, the plot shows a clear nonconstant slope, and moreover, the form of the nonconstancy is shown: a very large rate (slope) during the first year, followed by a somewhat smaller rate, and then a very small rate in the last years. Two comments must be made. First, this figure requires the unbinned data, which may not be available in every problem. Second, the use of calendar time instead of critical time on the horizontal axis may distort the figure somewhat.

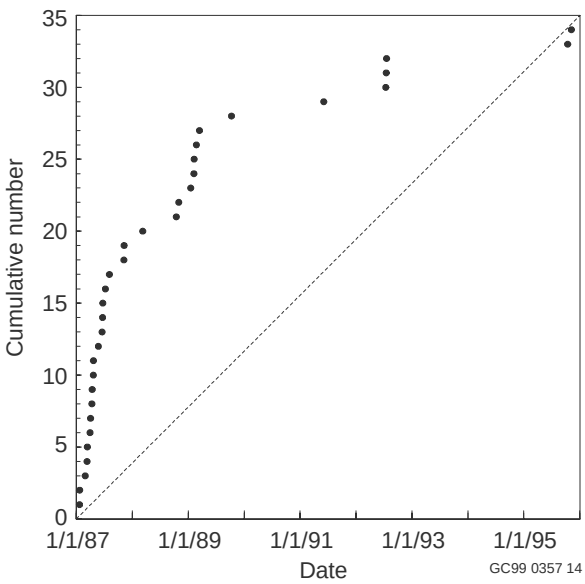

Figure 14 Cumulative count of initiating events, for data of Table 7.3 (Example 2.1).

If lack of fit is present, the analyst should try to identify the causes of the lack of fit. One frequent cause of lack of fit is systematic variation — the assumed form of the model is incorrect. In the examples of this section, systematic variation means that $\ln\lambda(t)$ is not of the form $a + bt$. This was revealed by Figures 7.12 and 7.14. Another possible cause is extra-Poisson variance, additional sources of variation that are not accounted for in the Poisson model. Figure 7.13 may show this. The residual for 1995 is surprisingly large. Table 7.3 and Figure 7.14 both show that two events occurred in relatively quick succession in 1995, and that three events occurred in quick succession in 1992. If any of these events were dependent on each other, such dependence would exaggerate the normal random variation in the counts from year to year.

One must be careful about how to correct lack of fit. In this example, it is reasonable to delete the early history of the plant, the part corresponding to the learning period. This would flatten the fitted line for the rest of Figure 7.12, making it lower in 1988 and higher in 1995. Thus, dropping early data would make the late data fit better. One would have to perform the analysis to know whether this proposed solution will completely remove the lack of fit in the late years or only reduce it.

The above approach does not consist of throwing away data. Instead, it divides the data into two relatively homogeneous sets, which never should have been pooled. The set after the end of the learning period can be analyzed as described above. The other set, during the learning period, can also be analyzed. If it is thrown away, that is only because no one chooses to analyze it.

If dividing the data into homogeneous pieces does not correct the lack of fit, another option is to construct a more complex model. For example, λ could be modeled as a function of more variables than just t. Or one could postulate a random count with larger variance than the Poisson variance. Such topics are beyond the scope of this handbook.

7.2.6.2 Statistical Test for Goodness of Fit

7.2.6.2.1 Test Based on Poisson Maximum Likelihood

When Poisson maximum likelihood is used to fit the trend to the data, some software packages give two measures of goodness of fit, the Pearson chi-squared statistic and the **deviance**. The Pearson chi-squared statistic, denoted X^2, is the sum of squares of the Pearson residuals. The deviance is based on the theory of generalized linear models. It is defined as

$$D = 2\sum x_i\{\ln(x_i) - \ln[s_i\,\hat{\lambda}(t_i)]\}\ .$$

For more details, see Atwood (1995) or books on the generalized linear model.

The assumed model is

$$H_0\colon \lambda(t) = a + bt$$

for some (unknown) constants a and b. If this model is correct, and if the number of observations per bin is large, both X^2 and D have approximately a chi-squared distribution, with degrees of freedom equal to the number of bins minus the number of unknown parameters. In fact, the two statistics are asymptotically equal. For small samples, on the other hand, the two are not necessarily nearly equal to each other, nor is their distribution approximately chi-squared. The distribution of X^2 typically approaches the chi-square distribution faster than the distribution of D does.

These two statistics can be used to test whether H_0 is true. Four situations can arise in practice.

- Both X^2 and D are in the upper tail of the chi-squared distribution, larger than, say, the 95th percentile. This is evidence of lack of fit to the model, H_0. Report the p-value based on X^2. For example, if X^2 is at the 98th percentile of the chi-squared distribution, report a p-value of 0.02. Investigate the data to try to discover the reason for the lack of fit.
- Both X^2 and D are in the middle of the chi-squared distribution, say between the 5th and 95th percentiles. Then the model appears to fit adequately.
- Both X^2 and D are in the lower tail of the chi-squared distribution. This is an indication of overfit, with too complex a model to be justified by the data. Although such a situation will probably not arise with the two-parameter models of this chapter, it can arise when the model contains many factors, such as component age, manufacturer, environment, etc.
- X^2 and D are so different from each other that they give conflicting conclusions. That is, one statistic is in the upper tail of the chi-squared distribution and the other is in the lower tail, or one is in a tail and the other is in the middle. This can indicate one of two possibilities. (1) The data set may be too small to allow an accurate assessment of the goodness of fit. The problem often can be remedied by pooling the data to some extent. For example, it is possible to fit a loglinear model using nine one-year bins from Table 7.4. If X^2 and D conflict, try pooling the data into two-year bins, and so forth. (2) H_0 may be false in a way that one statistic detects and the other does not. X^2 and D are asymptotically equal only if H_0 is true, not if H_0 is false. In this case, if it is really important to decide whether H_0 should be rejected, one could try simulating data from the fitted model, and seeing what fraction of the simulated data sets produce a simulated X^2 or D as large as actually observed. That fraction would approximate the exact p-value, without relying on the asymptotic chi-squared approximation.

For the HPCI unplanned demand data in Table 7.1, the loglinear model seems to fit well. The values of X^2 and D are 4.90 and 5.12, respectively. These are both in the middle of a chi-squared distribution. The degrees of freedom, 5, equals the number of bins, 7, minus the number of unknown parameters, a and b.

For the initiating-event data of Table 7.4, X^2 and D are 28.93 and 24.17. These are both far out in the right tail of a chi-squared distribution with 7 degrees of freedom. This is very strong evidence against the loglinear model.

The results of these statistical tests are consistent with the conclusions based on the graphs.

7.2.6.2.2 Test Based on Weighted Least-Squares Fit

Consider fitting a function of the form

$$y = a + bt$$

based on observations y_i at times t_i. In the present context, y equals $\ln(x_i / s_i)$. In Section 7.2.4.6, the weighted sum of squares

$$\sum w_i[y_i - (a + bt_i)]^2$$

was minimized, with the weights equal to the inverses of the estimated variances of $Y_i = \ln(X_i / s_i)$. If the model assumptions are correct, and if the Y_is are approximately normally distributed, the weighted sum of squares has approximately a chi-squared distribution. The degrees of freedom d is the number of bins minus the number of unknown parameters. The degrees of freedom are 5 in the HPCI-demand example and 7 in the initiating-events example. (Purists will note that the chi-squared distribution applies if the weights are fixed in advance, not derived from the random data. This slight departure from theory is commonly ignored.)

If the weighted sum of squares is in the right tail of the chi-squared distribution, such as beyond the 95th percentile, this is evidence of lack of fit. As mentioned above, one common cause of lack of fit is systematic variation — the assumed form of the model is incor-

rect. In the examples of this section, that means that $\ln\lambda(t)$ is not of the form $a + bt$. Another possible cause is extra-Poisson variance, additional sources of variation that are not accounted for in the Poisson model. To gain insight as to which contributors to lack of fit seem to be present, examine plots similar to Figures 7.12 and 7.13.

7.3 Unbinned Poisson Data

Example 6.6 and Table 7.3 are typical examples of the type of data considered here. That is, the exact event times are used. The corresponding summary tables of counts, given in Tables 7.1 and 7.4, are not used.

In principle, the exact event times contain more information than the summaries of counts in bins. The counts can be calculated from the exact event times, but the exact event times cannot be retrieved from the count totals. Therefore, in principle, better procedures can be squeezed out of the exact event times. However, software based on binned event counts is more widely available. Also, little information is lost by grouping the event times into bins, unless the bins are very large and coarse. Therefore, use of binned data is usually the most practical method for the data analyst.

7.3.1 Bayesian Analysis

When the exact event times are used, it is difficult to write out the likelihood. In principle, one must write the likelihood of each event time, conditional on the previous times, and finally the probability of no events after the final observed event until the end of the observation period. (See Section 3.3 of Cox and Lewis 1966.) Formulas are given for several cases by Atwood (1992), but they are not intuitive.

Binning the data, as in Section 7.2, is much simpler. Moreover, the bins may be made as numerous and small as the analyst desires. Many of the bins would then have event counts of zero, but that is allowed. This approach would capture virtually all of the information in the data. In practice there is little advantage in constructing very fine bins, but the analyst who was intent of squeezing every last bit of information from the data could do it.

7.3.2 Frequentist Analysis

Frequentist analysis is also simpler when the data are binned, although Atwood (1992) works out the formulas for the MLEs and approximate confidence intervals for several cases that use the exact event times. The unified notation in that article does not make the expression immediately obvious for any particular model. Typically, the MLEs must be found through numerical iteration rather than directly from algebraic formulas.

The only relatively easy case is the power-law model when the process is observed from time zero. Typically, the data collection begins at some time in the middle of operation, but in those rare cases when the data collection starts at time zero and ends at some time τ, the MLEs of the parameters in the two parameterizations of the power-law model are:

$$\hat{b} = n \Big/ \sum_{i=1}^{n} \ln(\tau / t_i)$$

$$\hat{c} = \tau / n^{1/\hat{b}}$$

$$\hat{A} = n\hat{b} / \tau^{\hat{b}} \ .$$

Here n is the number of events, t_i is the time of the ith event, and τ is the final time in the data-observation period. These formulas can be obtained by translation into the notation of Section 7.2.2.1 of formulas in Bain and Engelhardt (1991, Chap. 9, Eq. 13) or Atwood (1992, Section 6.1). Those references also consider confidence intervals and goodness-of-fit tests.

The simplest approach is to bin the data and use the methods of Section 7.2. Remember that the bins must not be too fine; a conservative rule of thumb says that most of the bins should have expected event counts of 5 or more.

7.4 Binomial Data

This section parallels Section 7.2 closely. Only the formulas are different, because the section deals with failures on demand instead of events in time. For this reason, some of the topics are given a cursory treatment. These topics are completely analogous to the

material in Section 7.2, where a fuller description can be found.

7.4.1 Examples

Example 6.10 can be used. This example consisted of 63 demands for the HPCI system during seven years, and 12 failures. It is convenient to combine the data into bins, such as calendar months, calendar years, etc. Such binning summarizes the data in a compact way. For goodness-of-fit tests, discussed in Section 7.4.6, binning is not merely convenient – it is required. If the bins are too fine (too few failures and successes expected in each bin) then the goodness-of-fit statistics X^2 and D will be inconsistent with each other, and neither will have an approximate chi-squared distribution under H_0. On the other hand, the bins must not be too coarse. As in Section 7.2 denote the midpoint of the ith bin by t_i. The bins must be small enough so that the expected number of failures in the bin can be approximated by the number of demands in the bin times $p(t_i)$.
Therefore, the data from Example 6.10 are summarized by calendar year in Table 6.14, which is repeated here as Table 7.5.

Table 7.5 HPCI failures on demand, by year (from Example 6.10).

Calendar year	Failures	Demands
1987	4	16
1988	2	10
1989	1	7
1990	3	13
1991	2	9
1992	0	6
1993	0	2

In this example the bins correspond to calendar years. Other examples could be constructed in which the bins correspond to ages, so that p would be modeled as a function of age rather than calendar time.

7.4.2 Model

7.4.2.1 General Model

The model is the same as that in Sections 2.3.1 and 6.3, except now the probability p depends on time, t. Thus, the model assumptions are:

1. The outcome of a demand at time t is a failure with some probability $p(t)$, and a success with probability $1 - p(t)$.

2. Occurrences of failures on different demands are statistically independent.

The number of demands and their times are assumed to be fixed, and the outcome on each demand is assumed to be random.

7.4.2.2 Logit Model

By far the most commonly used functional form for $p(t)$ is the **logit** model. In this model the logit transform of $p(t)$, defined as $\ln\{ p(t)/[1 - p(t)] \}$, is a function of unknown parameters. Fitting such a model to data is sometimes called **logistic regression**. The model that will be used in this section is

$$\ln\{ p(t)/[1 - p(t)] \} = a + bt .$$

This function of p has a name, the **logit** function. (This function was also encountered in Section 6.3.2.5.2 and in Appendix A.7.9, where the logistic-normal distribution is introduced.) Therefore, the above equation could be written

$$\text{logit}[p(t)] = a + bt .$$

Like the loglinear model for $\lambda(t)$, the logit model for $p(t)$ is a model in the class of generalized linear models.

A frequently used relation is that

$$y = \text{logit}[p(t)] \equiv \ln\{ p(t)/[1 - p(t)] \}$$

is equivalent to

$$p(t) = \text{logit}^{-1}(y) \equiv e^y/(1 + e^y) , \qquad (7.7)$$

denoting the inverse function of the logit by logit^{-1}. Figure 7.15 shows logit[$p(t)$] as a function of t, and Figure 7.16 shows $p(t)$ itself as a function of t. Notice that in Figure 7.16 the value of $p(t)$ stays between 0.0 and 1.0, as it should.

The parameters have simple interpretations: a is the value of the logit of p when $t = 0$, and b is the slope of the logit of p.

7.4.2.3 Loglinear Model

If p is small, then logit(p) is close to ln(p), and the logit model could be approximated by

$$\ln[p(t)] = a + bt \ .$$

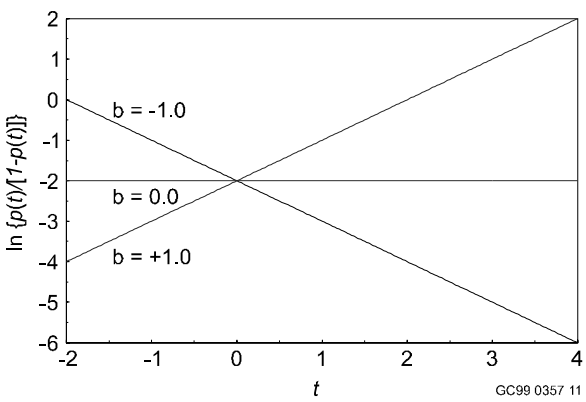

Figure 15 Plot of ln{ $p(t)/[1 - p(t)]$ } = $a + bt$, with $a = -2$ and three values of b.

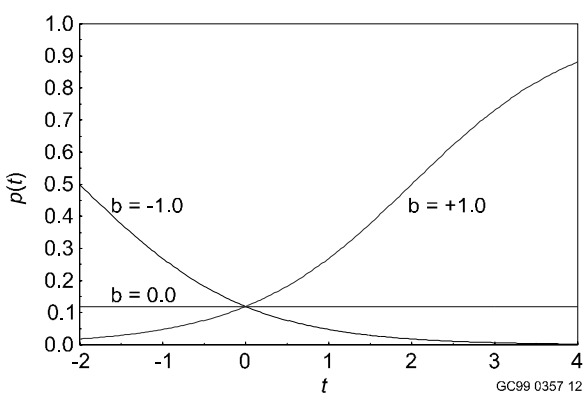

Figure 16 Plot of $p(t)$ as a function of t, corresponding to Figure 7.15.

This is a loglinear model, just as in Section 7.2.2.2. Software programs for analyzing a generalized linear model always include the logit model as one special case, and software programs for logistic regression are based on the logit model. However, if for some reason the analyst has software that only allows for the log-transformation, not the logit-transformation, that software is probably adequate as long as p is small.

7.4.3 Bayesian Estimation with Logit Model

The large-sample theory mentioned in Section 7.2.3 applies here as well. As the data set becomes large (many demands and failures) the form of the likelihood function approaches the form of a normal density in the two variables a and b. Therefore, with large data sets the **conjugate** prior is normal: if a and b are given a normal prior distribution, the posterior will be approximately normal, with very good approximation as the data set becomes large. The corresponding **noninformative** prior for a and b is the limiting case as the variance approaches infinity, which is a constant density.

For the work here, it will be assumed that a computer program produces a sample from the posterior distribution. The theory sketched above will then lead us to the conclusions that the posterior distributions not only *appear* normal, they really *are* normal, or very close to normal.

A convenient, and free, software package is BUGS (1995), Bayesian inference Using Gibbs Sampling. This was also used in Section 7.2.3, where it is described in more detail. Example 6.10 is analyzed here using the Windows version, WinBUGS, assuming a loglinear model.

Using the data of Example 6.10 (Table 7.5), BUGS was used to model $p(i) = \exp(a + bi)$, for i from 1 to 7. Then $X(i)$ was modeled as having a binomial($p(i)$, $d(i)$) distribution, where $d(i)$ is the number of demands in year i. Finally, a and b were given very diffuse prior normal distributions.

Figure 7.17 shows the BUGS script that was used to analyze the data.

```
model
{
   for (i in 1:N) {
      p[i] <- exp(a + i*b)/(1 + exp(a+i*b))
      x[i] ~ dbin(p[i], d[i])
   }
   a ~ dnorm(0.0, 0.0001)
   b ~ dnorm(0.0, 0.0001)
}
```

Figure 17 BUGS script for analyzing data of Example 6.10.

This uses the Equation 7.7 for expressing p in terms of the normally distributed quantity $a + ib$. Note the way BUGS happens to parameterize distributions, putting p before d in the list of binomial parameters, and parameterizing the normal distribution in terms of the precision $\tau = 1/\sigma^2$. The reasons are explained in 6.7.1.2.1. Therefore, a precision of 0.0001 in the script corresponds to a standard deviation of 100. That gives a very diffuse distribution.

The script was executed with 4 separate starting values of a and b, generating 4 **chains** of values, each 10000 elements long. The first 500 elements of each chain were discarded, and the remaining 38000 elements were used to estimate the posterior distributions. Table 7.6 summarizes the posterior distributions for a and b. When interpreting these summaries, be aware that a and b are not independently distributed.

Table 7.6 Posterior statistics for *a* and *b*, for loglinear model of Example 6.5.

	a	*b*
mean	−0.8838	−0.2085
median	−0.8654	−0.204
st. dev.	0.6477	0.1961
5th percentile	−1.981	−0.5395
95th percentile	0.1471	0.1063

The table shows that the mean and median are nearly equal in each case, and the 5th and 95th percentiles are approximately 1.645 standard deviations from the mean. This approximation is poorest for a, with $-0.8838 + 1.645\times0.6477 = 0.182$, somewhat larger than the reported 95th percentile. That is, the distributions appear to be approximately normal, as anticipated by the theory mentioned earlier, but the approximation is not perfect in the tails.

The mean of b is negative, but not strongly so, only 1.06 standard deviations below 0. Treating b as normally distributed, a table of the normal distribution shows that b is negative with probability 0.86 and positive with probability 0.14. Therefore, we are not really sure that the trend is downward.

The posterior distribution of p is approximately logistic-normal in any year. It is shown in Figure 7.18. The median is plotted as a solid line and the 5th and 95th percentiles are shown as dashed lines. This is analogous to Figure 7.7 for λ.

Figure 7.7 also plotted the MLEs, based on each year's data, as dots. To illustrate the possibilities of variation, Figure 7.18 is constructed somewhat differently. The total data set has 12 failures in 63 demands. Therefore we constructed the constrained noninformative prior with mean 12/63. Interpolation of Table C.8 shows that this prior is approximately beta(0.324, 1.376). For each year of data, this prior was updated to obtain the posterior for that year; the 90% credible interval was plotted as a vertical line, with a dot showing the posterior mean.

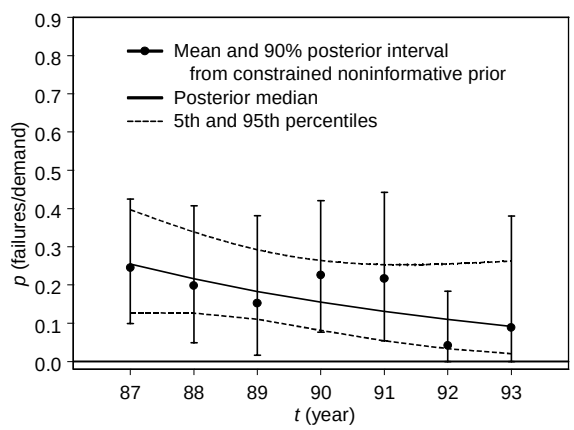

Figure 18 Posterior trend line for p with 90% credible band, for data of Table 7.5. In addition, annual estimates and 90% credible intervals are shown, based on constrained noninformative prior.

All the intervals overlap the fitted trend line. This is graphical evidence that the logit model fits the data well. (Many other models might also fit this sparse data set well.)

7.4.4 Frequentist Estimation with Logit Model

7.4.4.1 Point Estimation

The model is analyzed by finding the MLEs of a and b, based on binomial counts. The discussion below will sometimes call this the **binomial-maximum-likelihood** method. The software finds the estimates by numerical iteration.

When this model is fitted to the data of Table 7.5, a fitted trend is found, which can be overlaid on Figure 6.26. It is not shown here, but will be displayed with a simultaneous confidence band in Section 7.4.4.4.

7.4.4.2 Confidence Intervals for *a* and *b*

As in Section 7.2.4.2, almost all software for estimating a and b reports standard errors, estimates of the standard deviations of the maximum likelihood estimators. The estimators are assumed to be approximately normally distributed, which is valid unless the sample size is small.

Therefore, as in Section 7.2.4.2, a $100(1 - \alpha)\%$ confidence interval for b is

$$\hat{b} \pm z_{1-\alpha/2} se(b)$$

where $\hat{b}$ is the estimate, and $z_{1-\alpha/2}$ is the $1 - \alpha/2$ quantile of the normal distribution. The term $se(b)$ is the standard error of b, the estimated standard deviation of the estimator.

As in Section 7.2.4.2, the hypothesis

H_0: $b = 0$,

the hypothesis of no trend, is rejected at level 0.10 if the 90% confidence interval for b is entirely on one side of 0. The hypothesis is rejected at level 0.05 if the 95% confidence interval is entirely on one side of 0, and so forth. Most software packages print out a significance level at which the hypothesis that $b = 0$ is rejected, the p-value for the trend.

7.4.4.3 Test for Presence of Trend

Consider the two hypotheses defined by:

H_0: $p(t)$ is constant
H_1: $p(t) = \text{logit}^{-1}(a + bt)$, $b \neq 0$

Note, the null hypothesis H_0 is true if $p(t) = \text{logit}^{-1}(a + bt)$ and b is zero. Therefore, with this choice of an alternative hypothesis, the test of H_0 is the same as a test that $b = 0$, the test given above based on a confidence interval for b.

This is different from the tests of Chapter 6. Section 6.3.3.2.2 uses a different alternative hypothesis,

H_1: $p(t)$ is not constant .

It also uses a different test, the chi-squared test. Section 6.3.3.2.2 commented that the test is not very powerful against the alternative of a trend in p, because it considers such a broad class of possibilities as the alternative hypothesis.

Consider Example 6.10, with the HPCI failures during unplanned demands, as summarized in Table 7.5. In Section 6.3.3.2.2, the chi-squared test rejected the hypothesis of constant p with p-value 0.77. That is, the test found no evidence of nonconstant p. The present test of $b = 0$ rejects this hypothesis with p-value 0.30. That is, this test still does not reject the hypothesis of constant p. However, it notices the slightly falling values of p in Figure 6.33, and therefore sees somewhat stronger evidence against constant p than the chi-squared test did. Incidentally, the test based on $\hat{b}$ and the Wilcoxon-Mann-Whitney test for trend (Section 6.3.3.2.2) reach very similar conclusions.

7.4.4.4 Confidence Intervals and Confidence Bands

Confidence intervals for $p(t)$ at a particular t and simultaneous confidence bands valid for all t both are based on the approximate normality of the MLEs $\hat{a}$

and $\hat{b}$. The software finds an approximate 100(1 − α)% confidence interval for logit[$p(t)$] as

$$\hat{a} + \hat{b}t \pm [z_{1-\alpha/2} \times se(a + bt)] \qquad (7.8)$$

where, as before, $z_{1-\alpha/2}$ is the 100(1 − α/2) percentile of the standard normal distribution, and $se(a + bt)$ is the standard error, the estimated standard deviation of $\hat{a} + \hat{b}t$. The standard error depends on the value of t, and accounts for the fact that the MLEs $\hat{a}$ and $\hat{b}$ are statistically correlated, not independent. The confidence interval for $p(t)$ itself is found by inverting the logit function. If L and U are the lower and upper confidence bounds for logit[$p(t)$], then

$$\text{logit}^{-1}(L) = e^L/(1 + e^L) \text{ and}$$
$$\text{logit}^{-1}(U) = e^U/(1 + e^U) \qquad (7.9)$$

are the corresponding confidence bounds for $p(t)$. Manipulation of Equation 7.8 allows the analyst to convert from one degree of confidence to another, say from 90% to 99%, by using different percentiles of the normal distribution and the single standard error found by the software.

As discussed in Section 7.2.4.5, a confidence interval is valid at one t, and the band constructed from the individual confidence intervals is not simultaneously valid for all t. A simultaneous 100(1 − α)% confidence band for logit[$p(t)$] is found by replacing $z_{1-\alpha/2}$ in Equation 7.8 by

$$[\chi^2_{1-\alpha}(r)]^{1/2} ,$$

with r equal to the number of estimated parameters, 2 in Equation 7.8. This is exactly the value that was used in Section 7.2.4.5. The only difference is that there the confidence band was for ln$\lambda(t)$ and here it is for logit[$p(t)$]. The confidence band for $\lambda(t)$ was found by inverting the logarithm function, that is, by taking an exponential. The confidence band for $p(t)$ is found by inverting the logit function: if L and U are now used to denote the lower and upper edges of the simultaneous confidence band for logit[$p(t)$] at some t, the corresponding points on the confidence band for $p(t)$ are given by Equation 7.9.

Figure 7.19 shows the data from Table 7.5, plotted as in Figure 6.33 but now with the fitted trend line and the simultaneous 90% confidence band overlaid. The confidence band can easily contain a horizontal line; this is consistent with the fact that the hypothesis of constant p cannot be rejected.

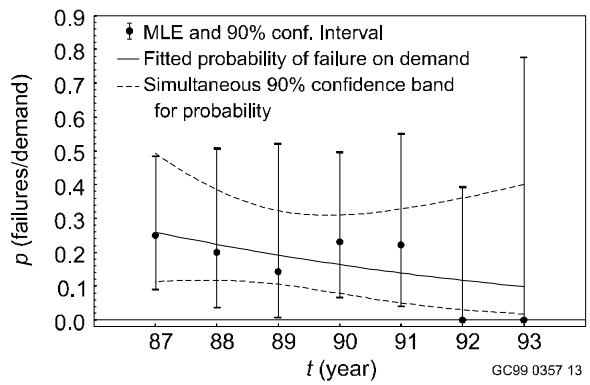

Figure 19 Annual estimates of p, fitted trend line and 90% confidence band for $p(t)$, for data of Table 7.5.

7.4.4.5 Alternative Using Least-Squares Software

The model assumes that logit[$p(t)$] = $a + bt$. Therefore, as in Section 7.2.4.6, one might decide simply to use least-squares software as follows. First, estimate p for each bin, based on only the data for that bin:

$$\hat{p}_i = x_i / n_i .$$

Then fit logit($\hat{p}_i$) to $a + bt_i$ by least squares. The same problems that were mentioned in Section 7.2.4.6 must be addressed here.

First, if any observed failure count, x_i, equals either 0 or the demand count n_i, the MLE $\hat{p}$ will be 0.0 or 1.0 for that bin, and the logit will be undefined. In the data of Table 7.5, this happens for the final two years. The following ways around this have been proposed, analogues of proposals in Section 7.2.4.6.

1. Instead of estimating p_i by x_i/n_i, use $(x_i + ½)/(n_i + 1)$. This is equivalent to replacing the MLE by the posterior mean based on the Jeffreys noninformative prior.
2. Estimate p_i by the posterior mean based on the constrained noninformative prior. In this case, the constraint could be that the prior

mean equals the observed overall mean $\Sigma_i x_i / \Sigma_i n_i$, or equals a modification to guarantee a positive number, $(\Sigma_i x_i + ½)/(\Sigma_i n_i + 1)$.

Such ways tend to reduce the trend slightly, because they add a constant to all the failure and demand counts, slightly flattening out any original differences.

The second point that must be considered is that the variance of the estimator of p is not constant. Therefore, the following iteratively reweighted least-squares method can be used. Assume that p in the ith bin is estimated by $(X_i + \alpha)/(n_i + \alpha + \beta)$. If the simple MLE is used, then α and β are both zero. If method 1 above is used, then $\alpha = ½$ and $\beta = ½$. If method 2 above is used, then α and β must be found from Table C.8. Neter and Wasserman (1974, Eq. 9.51) state that the asymptotic variance of logit(MLE of p_i) is

$$1/[n_i\, p_i(1 - p_i)] \; .$$

The method given here is a generalization when α and β are not both zero, setting the weight w_i to the inverse of the asymptotic variance of the estimator.

Begin by assuming that p is constant, and let $\hat{p}_i$ be some simple estimate of p, the same for all i. Fit $\text{logit}[(x_i+\alpha)/(n_i+\delta)]$ to a straight line with weighted least squares, and weights

$$w_i = \frac{(n_i \hat{p}_i + \alpha)^2 [n_i(1 - \hat{p}_i) + \beta]^2}{n_i \hat{p}_i (1 - \hat{p}_i)(n_i + \alpha + \beta)^2} \quad .$$

Calculate the resulting estimates of p_i,

$$\hat{p}_i = \text{logit}^{-1}(\hat{a} + \hat{b} t_i) \; .$$

Recalculate the weights with these estimates, and refit the data to a straight line using weighted least squares. Repeat this process until the estimates stabilize.

The third and final point is that least-squares fitting typically assumes that the data are approximately normally distributed around the straight line. In the present context, this means that logit($\hat{p}_i$) is assumed to be approximately normally distributed. This assumption is acceptable if the number of failures in each bin is not close to zero or to the number of demands. However, the variance of the normal distribution is then estimated from the scatter around the fitted line. This differs from typical treatment of binomial data, where the mean determines the variance.

A 90% confidence interval for $p(t)$ at a particular t is given by

$$\text{logit}^{-1}\{\hat{a} + \hat{b}t \pm [t_{0.95}(d) \times se(a + bt)]\} \qquad (7.10)$$

where $t_{0.95}(d)$ is the 95th percentile of Student's t distribution with d degrees of freedom, just as with $\lambda(t)$ in Section 7.2.4.6. The software will report the value of d. It is the number of bins minus the number of estimated parameters, 7 – 2 in the example of Table 7.5. The form of this equation is very similar to other equations in Sections 7.2 and 7.4, although the estimates and standard deviation are calculated somewhat differently.

A simultaneous 90% confidence band has the same form, but the multiplier $t_{0.95}(d)$ is replaced by

$$[2F_{0.90}(r, d)]^{½} \; ,$$

where $F_{0.90}(r, d)$ is the 90th percentile of the F distribution with r and d degrees of freedom.

7.4.5 Comparison of Bayesian and Frequentist Estimation with Logit Model

When the Bayesian analysis uses very diffuse priors, the conclusions of the two analyses will be numerically similar.

The posterior median in Figure 7.18 is very close to the fitted line (the MLE) in Figure 7.19. The 90% credible band in Figure 7.18 is narrower than the simultaneous confidence band in Figure 7.19, probably because the simultaneous confidence band is based on an inequality. The vertical lines in Figure 7.18, representing credible intervals for p based on individual years of data, are generally close to the confidence intervals in Figure 7.19, except for the years with little data, 1992 and 1993. For those two years, the confidence intervals are quite wide, but

the credible intervals are shorter, under the influence of the prior mean of 0.19.

Frequentist estimation must address the issue of interval-at-a-particular-time versus simultaneous-bands. That issue does not arise for Bayesians, for whom all belief is simultaneously valid.

Frequentist estimation relies on approximate normality of the estimators, and therefore does not work well with data having few observed events. Bayesian estimation obtains normal posteriors when the data set is large, but does not fail entirely when the data set is small – it merely obtains a different, non-normal, posterior.

Most frequentist software packages for analyzing trends include calculations for investigating the goodness of fit of the model, as will be seen in Section 7.4.6. Current Bayesian software may neglect this issue of model validation.

7.4.6 Model Validation

The two assumptions for a time-dependent binomial process are given at the beginning of Section 7.4.2. The first assumption is difficult to test from data. The other assumption is that outcomes on distinct demands are independent. One kind of dependence is serial dependence. Positive serial dependence means that failures tend to be followed by more failures, for example if a failure is misdiagnosed the first time, or if a single cause results in a number of failures before it is corrected. Negative serial dependence means that failures tend to be followed by successes, for example if the major cause of failure is wearout, at which time a new component is installed (without any failures from installation problems.)

Positive serial dependence results in failures tending to cluster together, with relatively long gaps between failures. When the data are collected into bins, this can translate into unusually high or low event counts in individual bins. This can be discovered by goodness-of-fit tests, considered below. However, it is impossible to decide, from the failure counts alone, whether the outcomes are serially correlated or whether p is going up and down. The cause can be determined only by an investigation to discover the failure mechanisms.

A negative serial dependence results in less-than-expected variation in the event counts. A goodness-of-fit test will report a p-value near 1.0, indicating surprisingly good fit, too good to be believable.

The final assumption made when fitting a trend model is the form of the model. Goodness-of-fit tests are designed for testing this assumption. In fact, a goodness-of-fit test is an all-purpose test for the various assumptions, although it is not good at deciding which assumption may be violated.

7.4.6.1 Graphical Check for Goodness of Fit

The natural graphical check is to compare the observed values to the fitted line.

Figure 7.18 and 7.19 each show such a plot for the data of Table 7.5 (Example 6.10). Either figure may be used. In Figure 7.18, each credible interval overlaps the fitted trend line, and in Figure 7.19, the 90% confidence interval for each year overlaps the fitted trend line. Because no year deviates strongly from the overall trend, the data appear consistent with the assumption of the logit model.

The discussion at the end of Section 7.2.6.1 applies here as well, concerning interpretation of a few intervals' failure to overlap the fitted line, and concerning the need for an engineering assessment of any strange patterns.

As with the loglinear model for $\lambda(t)$, the residuals and standardized residuals can be plotted for $p(t)$. Software may report these as the "raw residuals" and the "Pearson chi-squared residuals," respectively.

In the present context, the ith count, X_i, is assumed to be binomial with mean $n_i p(t_i)$. The ith residual, denoted r_i, is

$$r_i = x_i - n_i \hat{p}(t_i) \ .$$

The variance of a binomial(n, p) random variable equals $np(1 - p)$. Therefore, the standardized residual is

$$r_i \,/\, \sqrt{n_i \hat{p}(t_i)[1 - \hat{p}(t_i)]} \ .$$

These are also sometimes called the Pearson residuals or chi-squared residuals. A plot of the standardized

residuals against time may be helpful, as in Section 7.2.6.1.

Also, a simple cumulative plot may be informative, as it was in Section 7.2.6.1. In the present example, the cumulative plot is given in Figure 6.34, which shows no pattern of interest. In other data sets, such a plot might not only show nonconstancy in p, but it might suggest the form of the nonconstancy.

The cumulative plot shows failures per demand, revealing changes in p as a function of the demand count. However, if the rate of demands is changing, as it is in the present example, the plot can give a distorted picture if p is regarded as a function of calendar time. When p is constant, the issue of distortion is irrelevant — a constant p is constant, whether it is regarded as a function of demand count or of calendar time.

Just as in Section 7.2, lack of fit may be caused by systematic variation or by extra-binomial variance, additional sources of variation that are not accounted for in the binomial model. See the discussion in Section 7.2.6.1.

7.4.6.2 Statistical Test for Goodness of Fit

7.4.6.2.1 Test Based on Binomial Maximum Likelihood

Just as in Section 7.2, software packages that use binomial maximum likelihood may give two measures of goodness of fit, the Pearson chi-squared statistic, denoted X^2, and the deviance, denoted D. The discussion in Section 7.2.6.2.1 applies here as well.

For the HPCI failure data in Table 7.5, the logit model seems to fit well. The values of X^2 and D are 2.12 and 2.91, respectively. These are both in the middle of a chi-squared distribution with five degrees of freedom. The degrees of freedom, 5, equals the number of bins, 7, minus the number of unknown parameters, a and b. The p-value for lack of fit, based on X^2, is 0.84, indicating very good fit.

7.4.6.2.2 Test Based on Weighted Least-Squares Fit

Consider fitting a function of the form

$$y = a + bt$$

based on observations y_i at times t_i. In the present context, y equals logit(x_i/n_i). As in Section 7.2, if the model assumptions are correct, and if the Y_is are approximately normally distributed, the weighted sum of squares has approximately a chi-squared distribution. The degrees of freedom d is the number of bins minus the number of unknown parameters. The number of degrees of freedom is 5 in the HPCI-failure example. The discussion in Section 7.2.6.2.2 applies here as well.

7.5 Discussion

This section ties together some specific methods given in Chapters 6 and 7, showing the unifying formulations. Readers who are happy simply using the earlier recipes may skip this section.

7.5.1 Generalized Linear Models

Some software packages that implement the loglinear and logit models do so in the framework of the generalized linear model. Such models are described in a highly mathematical way by McCullagh and Nelder (1989), and in an introductory way in Appendix B-2 of Atwood (1995). This model has several elements, a **random component**, a **systematic component**, and a **link**, a function relating the random and the systematic components.

- The random component consists of some independent observations $Y_1, \ldots, Y_m$. This is thought of as an m-dimensional vector, $\mathbf{Y}$. The examples of this chapter have been the normalized Poisson event count, $Y_i = X_i/s_i$, and the fraction of binomial failures on demand $Y_i = X_i/n_i$.
- The systematic component is an m-dimensional vector $\boldsymbol{\eta}$, with the ith element η_i related to explanatory variables, and to unknown parameters in a linear way. The example of this chapter has been $\eta_i = a + bt_i$, where a and b are unknown parameters and the explanatory variable t_i is the calendar time or age for Y_i.
- The link is a function g with

$$\eta_i = g[E(Y_i)] \ . \qquad (7.11)$$

In this chapter, the links have been the log function for Poisson data and the logit function for binomial data.

Thus the two examples of this chapter have been the Poisson example, with

$$\ln[E(X_i/s_i)] = \ln[\lambda(t_i)] = a + bt_i ,$$

and the binomial example, with

$$\text{logit}[E(X_i/n_i)] = \text{logit}[p(t_i)] = a + bt_i .$$

This is the terminology used by many statistical software packages. The analyst must specify the distribution for the random component, the form of the systematic component in terms of the unknown parameters and the explanatory variables, and the link function.

Other software packages use a slight variant of this terminology. This is a generalized linear model with an **offset**, replacing Equation 7.11 by

$$g[E(Y_i)] = \eta_i + \text{offset}_i . \tag{7.12}$$

In the Poisson example, let Y_i be the Poisson count itself, X_i, not the normalized count X_i/s_i. Then the expected value of Y_i is $s_i\lambda(t_i)$, with $\ln\lambda(t_i)$ modeled as $a + bt_i$. To satisfy Equation 7.12, let the offset term be $\ln(s_i)$. Then we have:

$$\begin{aligned} g[E(Y_i)] &= \ln[E(X_i)] \\ &= \ln[s_i\lambda(t_i)] \\ &= \ln[\lambda(t_i)] + \ln(s_i) \\ &= a + bt_i + \ln(s_i) \\ &= \eta_i + \text{offset}_i . \end{aligned}$$

In this version of the model, the software package requires the analyst to specify the distribution of the random component, the form of the systematic component, the link function, and the offset. The disadvantage of this formulation is the extra term that must be specified, the offset. The advantage is that the distribution of X_i is Poisson, whereas the distribution of X_i/s_i is hard to specify because it does not have a standard name.

Much more elaborate models can be constructed in this general framework, by adding more explanatory variables. For example, both calendar time and age of the individual component could be treated together in one model as explanatory variables. The explanatory variables do not even have to be continuous. Manufacturer, system, and plant could be used as discrete explanatory variables. The possibilities are limited only by the availability of data. However, such models go beyond the limited scope of this handbook.

7.5.2 The Many Appearances of the Chi-Squared Test

In Chapter 6, the Pearson chi-squared test was used to test whether λ or p was constant. In Chapter 7, λ or p is assumed to be nonconstant, yet the chi-squared test is used anyway. Also, the chi-squared test was used in Section 6.7.2.3.2 to test whether durations had an assumed distributional form.

To see the unity in this apparent diversity, note first that in every case the Pearson chi-squared statistic has the form

$$X^2 = \Sigma_i \,(\text{observed}_i - \text{expected}_i)^2 / \text{expected}_i .$$

To clarify possible confusion one must think about the hypothesis being tested. The big general theorem states that when H_0 is true, X^2 has approximately a chi-squared distribution, and the degrees of freedom is the number of unknown parameters under H_1 minus the number of unknown parameters under H_0. The approximation is valid when the degrees of freedom stays constant and the size of the data set becomes large. This is now applied to the specific cases in this handbook.

Consider first Poisson data with event rate λ. In Chapter 6, the null and alternative hypotheses were:

H_0: λ is constant
H_1: λ is not constant .

The data were partitioned into c cells. These cells may have corresponded to different sources of data, such as different plants, or they may have resulted from binning the data, for example corresponding to c years. If H_0 is true there is one unknown parameter, the single value of λ. If, instead, H_1 is true, there are c unknown parameters, the values of λ in the different cells. Therefore, by the big general theorem, when H_0 is true

X^2 is approximately chi-squared with $c - 1$ degrees of freedom. This is the result stated in Section 6.2.3.1.2.

Consider now the corresponding case with binomial data. The data fall into a $2 \times J$ contingency table. The value J corresponds to the J sources of data or J bins, and the two rows correspond to failures and successes. The null and alternative hypotheses are:

H_0: p is constant
H_1: p is not constant .

If H_0 is true, there is one unknown parameter, p. If, instead, H_1 is true there are J unknown parameters, corresponding to the J bins or data sources. Therefore, the big general theorem says that the degrees of freedom for X^2 is $J - 1$, just as stated in Section 6.3.3.1.

Consider now the setting of Section 7.2, with $\lambda(t)$ modeled as $a + bt$. The time axis is divided into c bins. To test the goodness of fit, the hypotheses are:

H_0: $\lambda(t) = a + bt$,
H_1: λ for each bin is arbitrary

Under H_0 there are two unknown parameters, a and b. Under H_1, the number of unknown parameters is the number of bins, since each can correspond to a different λ. Therefore, the big general theorem says that the degrees of freedom, when testing H_0, is $c - 2$. This agrees with Section 7.2.6.2. Recall that the chi-squared distribution is approximate, for large data sets. The deviance was used as a back-up check, to help ensure that the data set was large enough.

The treatment of p in Section 7.4.6.2 is exactly parallel to that of λ.

Finally, the chi-square test was used in Section 6.7.2.3.2 to test whether durations follow a distribution of an assumed form. A duration was denoted as T. In the examples of Section 6.7, the assumed distribution of T had two unknown parameters. Possibilities included the lognormal(μ, σ) distribution and the gamma(α, β). The hypotheses in this setting were

H_0: T has a distribution of the assumed form.
H_1: T has some other distribution.

The data were partitioned into c bins, so that every observed value of T falls into one bin. Only the counts in the bins are used, not the individual duration times. If H_0 is true, there are two unknown parameters (μ and σ, or α and β, or whatever the two parameters of the assumed distribution are).

If H_1 is true, there are $c - 1$ parameters. These parameters are Pr(T falls in bin 1), Pr(T falls in bin 2), etc. There are c bins, but only $c - 1$ parameters, because the probabilities must sum to 1.0. Thus, any $c - 1$ parameters determine the final one.

The big general theorem should say that the degrees of freedom are $(c - 1) - 2$. However a subtle complication arises. An assumption of the big theorem is that the two unknown parameters of the distribution are estimated using the maximum likelihood estimated based on the bin counts. In this setting, however, it is far easier to estimate those parameters from the raw data. Therefore, as stated in Section 6.7.2.4.2, the degrees of freedom fall somewhere between $c - 3$ and $c - 1$.

To summarize this section, the Pearson chi-squared test has many applications. To avoid confusion, the analyst must clearly specify the null hypothesis being tested and the alternative hypothesis.

7.5.3 Nonparametric Estimators of $\lambda(t)$

The nonparametric estimators of a density in Section 6.7.3 can also be used to estimate a time-dependent event-occurrence rate $\lambda(t)$. In each case, the data consist of a number of times, durations in Section 6.7.3 and event times in the present case. The only difference is the scale: a density integrates to 1.0, and an occurrence rate does not. To use an estimator from Section 6.7.3 in the occurrence-rate setting, multiply the estimate by the total number of observed events. This is the correct scale correction.

The estimators in Section 6.7.3.1.1 showed a problem at 0, estimating a positive density to the left of zero even though a duration time cannot be negative. This problem was corrected by reflecting the data around 0, and initially using a data set that contained both the true data and the mirror images (a value at $-t$ for every value of t). Such a problem also occurs with estimation of $\lambda(t)$. If data are collected in a time period from τ_0 to

τ_1, simple kernel estimates will have this problem at both ends. To correct this problem, reflect the data at each end, so that artificial data has been constructed beyond τ_0 on the left and beyond τ_1 on the right. Construct the density estimate based on this augmented data set. Then truncate the density — set it to zero outside the observation interval, and multiply it by 3 so that it again integrates to 1.0. Finally, convert it from a density estimate to an estimate of the Poisson intensity by multiplying it by the number of observed events. When interpreting the resulting graphed function, be aware that the estimate will be flat at the two ends, by the way the estimate was constructed. The slope of the line at the two ends cannot be used to draw inferences about whether the Poisson intensity is changing.

This method was applied to the data of Table 7.3 (Example 2.1), unplanned scrams at a new reactor. The normal kernel was used. The 34 event dates were converted to consecutive days, and the standard deviation was 874 days. The formula

$$h = 1.06\, \sigma\, n^{-1/5}$$

resulted in h = 457 days. This is only a very preliminary suggested value for two reasons: σ was estimated, not known, and the true shape of the intensity function is not close to normal. Because it is wise to undersmooth rather than oversmooth, a value h = 400 days was used.

The estimated Poisson intensity function, $\lambda(t)$, is shown in Figure 7.20. The calculations were performed in terms of calendar days, and converted to reactor-critical-years by assuming 0.82 critical years per calendar year. This is the average of the values shown in the right column of Table 7.4.

The curve in Figure 7.20 can be compared with the exponential curve in Figure 7.12. The curve using the kernel estimator follows the ups and downs of the data more closely than does the simple exponential curve. However, the kernel-estimator curve at the

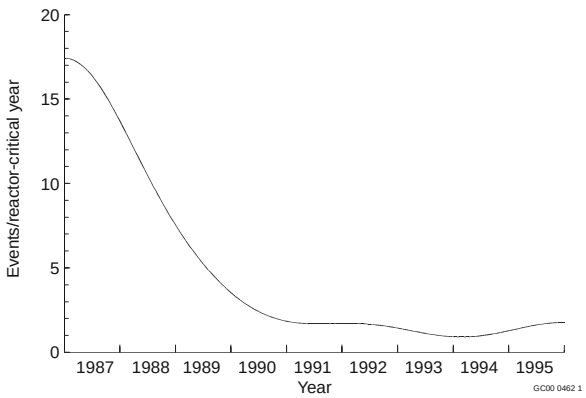

Figure 20 Estimate of $\lambda(t)$ using standard normal kernel, with bandwidth h = 400.

beginning of 1987 is only 17.4, substantially below the simple maximum likelihood estimate of 26.8, based on the 1987 data only. Two factors contribute to this. First, the bandwidth of h = 400 days is apparently too wide. The learning period at the plant lasted less than one year, so a smaller bandwidth would better reflect the rapid learning that was taking place. Second, the conversion from calendar time to reactor-critical-years used the average value for nine years, 0.82. In fact, the first calendar year had only 0.71 critical years. Therefore, the estimate during the first year should be about 15% larger than shown.

Figure 7.21 shows the kernel estimator with a smaller bandwidth, h = 200 days. It follows the rapid drop in the scram frequency at the beginning of the plant's history much more closely. It also is more sensitive to small, perhaps random, clusters of events later in the plant's history. The constant conversion rate of 0.82 reactor-critical-years per calendar year has been used, with the same effect as in the previous figure.

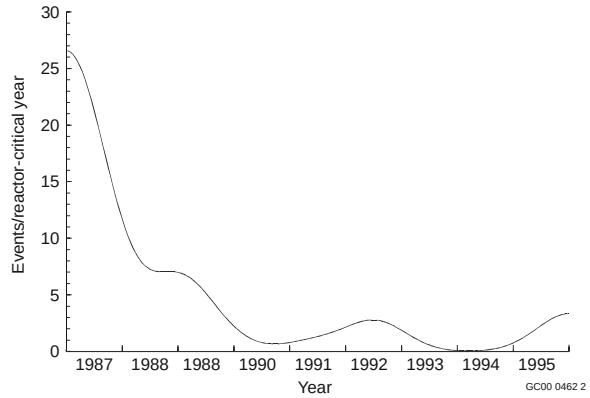

Figure 21 Estimate of $\lambda(t)$ using standard normal kernel, with bandwidth h = 200.

This example has illustrated both some advantages and some difficulties of nonparametric estimation of a Poisson intensity function. The same advantages and difficulties were seen for nonparametric density estimation. See Section 6.7.3 for more discussion of these issues.