Chapter

# 10

# Trend Analysis

**Karama Kanoun and Jean-Claude Laprie**
*LAAS-CNRS, Toulouse, France*

## 10.1 Introduction

Generally, software reliability studies are based on the application of reliability growth models to evaluate reliability measures. When performed on a large base of deployed software systems, the results are usually of high relevance (see [Adam84, Kano87] for examples of such studies). However, the utilization of reliability growth models during the early stages of development and validation is much less convincing; when the observed times to failure are in the order of minutes or hours, the predictions based on such data can hardly predict mean times to failure different from minutes or hours, which are far below any acceptable level of reliability! In addition, when a program under validation becomes reliable enough, the times to failure may simply be so large that applying reliability growth models to data collected during the end of validation is impractical due to the (hoped for) scarcity of failure data. On the other hand, in order to become a true engineering exercise, software validation should be guided by quantified considerations relating to its reliability. Statistical tests for trend analysis provide such guides.

This chapter addresses the problem of reliability growth analysis; it shows how reliability trend analyses can help the project manager control the progress of the development activities and appreciate the efficiency of the test programs. Reliability trend changes occur for various reasons. They may be desirable and expected (such as reliability growth due to fault removal) or undesirable (slowing down of testing effectiveness, for example). Timely identification of the latter allows the project manager to make the appropriate decisions in order to avoid problems that may surface later.

We introduce the notions of reliability growth over a given interval and local reliability trend change, allowing a better definition and understanding of the reliability growth phenomena. The already existing trend tests are then revisited using these concepts. We put the emphasis on the way trend tests can be used to help the management of the testing and validation process and on practical results that can be derived from their use. It is shown that, for several circumstances, trend analyses give information of prime importance to the developer. We also discuss their extension to software static analysis (e.g., specification and code inspection or review).

It is worth noting that, generally, most companies are accustomed to trend analysis during software testing (see e.g., [Grad87, Ross87, Vale88]). However, trend analyses are usually applied intuitively and empirically rather than in a quantified and well-defined manner. Moreover, such analyses are commonly restricted to failures reported during software execution. It is undoubtedly important to manage software testing, but equally important to manage the earliest phases of the verification and validation activities (for instance, static analysis through inspections, walk-through or code review) since efficient early static analyses significantly reduce subsequent development cost. We will thus discuss the extension of the trend analyses to data derived from static analysis before software execution (testing).

This chapter focuses on trend analysis. First, emphasis is placed on the characterization of reliability growth via the subadditive property and its graphical interpretation. Then we briefly present the Laplace test, which is a conventional trend test, and outline its relationship with the subadditive property. We then show how trend tests can be used to help manage the validation process before illustration on several data sets from real-life systems. Finally, the last section extends the application of trend tests to trouble reports recorded during static analysis of the software.

## 10.2 Reliability Growth Characterization

Lack of software reliability stems from the presence of faults. It is manifested by failures which are due to fault sensitization (see Chap. 2). Removing faults should result in reliability growth. However, this is not always the case, due to the complexity of the relationship between faults and failures, and therefore between faults and reliability, which was noticed long ago (see e.g., [Litt79a]). Basically, complexity arises from a double uncertainty: the presence of faults and the fault sensitization via the trajectory in the input space of a program.* As a conse-

* By way of example, the data published in [Adam84] concerning nine large software products show that for a program with a mean lifetime of 15 years, only 5 percent of the faults were activated during this period.

quence, reliability trend changes can occur, which may be due to a wide range of phenomena, such as

- *Variation in the utilization environment.* Variation in the testing effort during debugging, change in test sets, addition of new users during the operational life, etc.
- *Dependence between faults.* Some faults can be masked by others, they cannot be activated as long as the latter are not removed [Ohba84].

Reliability decrease may not, and usually does not, mean that the software has more and more faults. It is just an indication that the software exercises more and more failures per unit of time under the corresponding conditions of use. Corrections may reduce the failure input domain, but more faults are activated or faults are more frequently activated due to operational profile changes. However, during fault correction, new faults may also be introduced—regression faults—that are likely to affect the ability of the software to deliver a proper service, depending on the conditions of use. Last but not least, reliability decrease may be due to specification changes, as exemplified by the experimental data reported in [Kenn92].

### 10.2.1 Definitions of reliability growth

A common definition of reliability growth is that the successive interfailure times tend to become larger, i.e., denoting $T_1, T_2, \ldots,$ the sequence of random variables corresponding to interfailure times:

$$T_i \underset{st}{\leq} T_j \qquad \text{for all } i < j \tag{10.1}$$

where $\underset{st}{\leq}$ stands for *stochastically smaller than* (that is, $P\{T_i < v\} \geq P\{T_j < v\}$ for all $v > 0$). Under the stochastic independence assumption, Eq. (10.1) is equivalent to $F_{T_i}(x)$, denoting the cumulative distribution function of $T_i$:

$$F_{T_i}(x) \geq F_{T_j}(x) \qquad \text{for all } i < j \text{ and } x > 0 \tag{10.2}$$

An alternative to the (restrictive) assumption of stochastic independence is to consider that the successive failures are governed by a nonhomogeneous Poisson process (NHPP). Let $N(t)$ be the cumulative number of failures observed during time interval $[0, t]$, $H(t) = E[N(t)]$, its mean value, and $h(t) = dH(t)/dt$ its intensity, i.e., the failure intensity. A natural definition of reliability growth is then that the increase in the expected number of failures tends to become lower, i.e., that $H(t)$ is concave, or equivalently that $h(t)$ is nonincreasing. However, there are several situations where, even though the failure intensity fluctuates locally, reliability growth may take place on average on the considered time interval.* An alternative definition allowing for such local fluctuations is that the expected number of failures in any initial interval (i.e., of the form $[0, t]$) is no less than the expected number of failures in any interval of the same length occurring later (i.e., of the form $[x, x + t]$). The independent increment property of an NHPP enables us to write the latter definition as

$$H(t_1) + H(t_2) \geq H(t_1 + t_2) \qquad \text{for all } t_1, t_2 \geq 0 \text{ and } 0 \leq t_1 + t_2 \leq T \tag{10.3}$$

where inequality is assumed strict for at least one couple $(t_1, t_2)$. When Eq. (10.3) holds, the function is said to be *subadditive* (see e.g., [Holl74]). When Eq. (10.3) is reversed for all $t_1, t_2 \geq 0$ and $0 \leq t_1 + t_2 \leq T$, the function is said to be *superadditive,* indicating reliability decrease on average.

Equation (10.3) is very interesting, since it allows for local fluctuations: locally subintervals of reliability decrease may take place without affecting the nature of the trend over the whole time interval considered. When $h(t)$ is strictly decreasing over $[0, T]$, Eq. (10.3) is verified, but the converse is not true. This is detailed in the next subsection. The case where $h(t)$ is strictly decreasing (respectively, increasing) is usually referred to as *strict* or *monotone* reliability growth (respectively, decrease).

### 10.2.2 Graphical interpretation of the subadditive property

Let $C_t$ denote the portion of the curve representing the mean value function over $[0, t]$ as shown in Fig. 10.1, and $L_t$ be the line joining the two ending points of $C_t$ (i.e., the chord from the origin to point $(t, H(t))$ of $C_t$). Let $\mathcal{A}_H(t)$ denote the difference between (1) the area delimited by $C_t$ and the coordinate axis and (2) the area delimited by $L_t$ and the coordinate axis. With these notations, if $H(t)$ is subadditive over $[0, T]$, then

$$\mathcal{A}_H(t) \geq 0 \qquad \text{for all } t \in [0, T] \tag{10.4}$$

This property can be shown as follows. Let us divide the interval $[0, t]$ in $K$ small time intervals of length $dt$, that is, $t = Kdt$. $K$ may be

* The NHPP assumption (more precisely, the property of independent increments) is essential since a stationary process which is a non-Poisson process may undergo transient oscillations that cannot be distinguished from a trend in a nonstationary Poisson process (see, for instance, [Asch84, Gned69] for renewal processes).

even or odd. Let us consider the even case. In Eq. (10.3), let $t_1$ successively take the values $\{0, dt, 2dt, 3dt, \ldots, (K/2)dt\}$ and $t_2 = t - t_1$. Equation (10.3) successively becomes

$$H(0) + H(Kdt) \geq H(t)$$

$$H(dt) + H((K-1)dt) \geq H(t)$$

$$H(2dt) + H((K-2)dt) \geq H(t)$$

$$\cdots$$

$$H\left(\frac{K}{2}dt\right) + H\left(\frac{K}{2}dt\right) \geq H(t)$$

Summing over the $(K/2) + 1$ inequalities gives

$$\sum_{j=0}^{K} H(jdt) + H\left(\frac{K}{2}dt\right) \geq \left(\frac{K}{2} + 1\right) H(t)$$

Replacing $K$ by $t/dt$ and taking the limit when $dt$ approaches zero lead to

$$\int_0^t H(x)\,dx \geq \frac{t}{2} H(t)$$

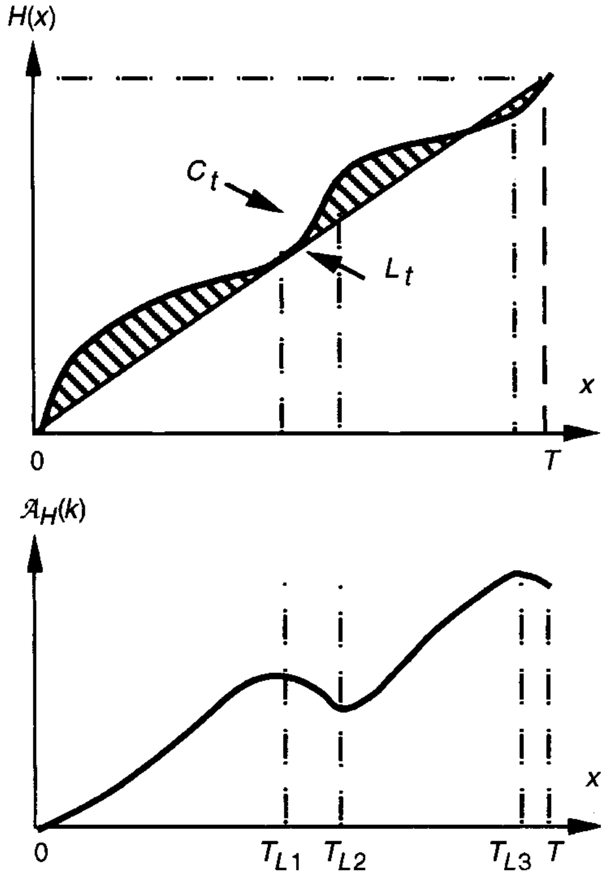


**Figure 10.1** Graphical interpretation of the subadditive property.

The left term corresponds to the area delimited by $C_t$ and the coordinate axis; the right term corresponds to the area between $L_t$ and the coordinate axis.

Equation (10.3) implies Eq. (10.5):

$$\int_0^t H(x)\,dx - \frac{t}{2}H(t) \geq 0 \qquad \text{for all } t \in [0, T] \tag{10.5}$$

That is, $\mathcal{A}_H(t) \geq 0$.

It can also be shown that Eq. (10.5) implies Eq. (10.3), which means that Eqs. (10.3) and (10.5) are equivalent. When $K$ is odd, derivation can be handled in a similar manner.

Throughout this chapter, $\mathcal{A}_H(t)$ is called the *subadditivity factor.*

With this graphical representation, the subadditive property is easily identified. For example, the function considered in Fig. 10.1 is subadditive over $[0, T]$; there is thus reliability growth over the whole time interval.

### 10.2.3 Subadditive property analysis

It is worth noting that, for a subadditive function over $[0, T]$ when $t$ varies from 0 to $T$, the difference between the two areas, $\mathcal{A}_H(t)$, may increase, decrease, or become zero without being negative. The variations of $\mathcal{A}_H(t)$ indicate *local* trend changes.

Let us consider a subadditive function; $\mathcal{A}_H(t)$ is thus positive and increasing at the beginning, and

- Without local trend change, the mean value function is concave, leading to $\mathcal{A}_H(t)$ positive and increasing over the considered interval. Such a situation is illustrated by case A in Fig. 10.2,
- In case of local trend change, the mean value function is no longer concave and the difference between the two areas is not increasing

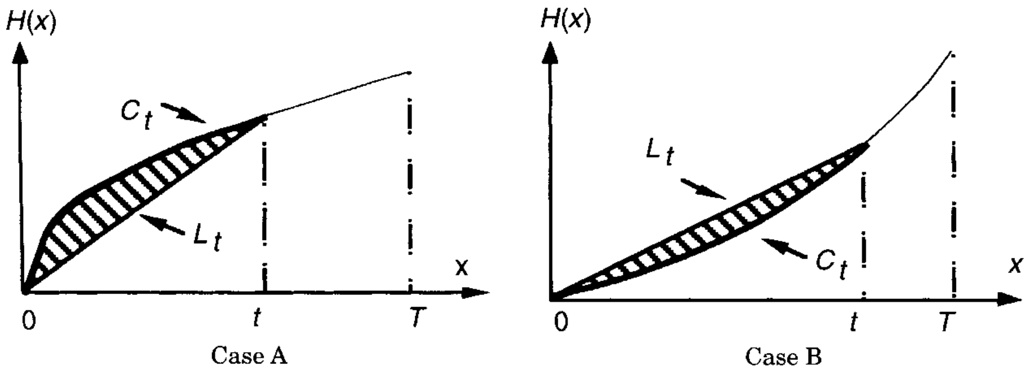


**Figure 10.2** Subadditivity (left) and superadditivity (right) without local trend variation.

over the considered interval. Figure 10.1 gives an example of such a situation. $\mathcal{A}_H(t)$ takes its maxima (minima) when its derivative is null. Let $T_{L1}$ denote the time at which the first maximum takes place. From $T_{L1}$, $\mathcal{A}_H(t)$ is decreasing (denoting local reliability decrease) up to the next point where the derivative is null again (point $T_{L2}$ of Fig. 10.1). From $T_{L2}$, $\mathcal{A}_H(t)$ is increasing again (denoting local reliability growth) and so forth.

In fact, Fig. 10.1 shows a situation with two subintervals of local reliability decrease (namely, intervals $[T_{L1}, T_{L2}]$, and $[T_{L3}, T]$) despite reliability growth on the whole interval $[0, T]$, since the function is subadditive over $[0, T]$.

Let $\mathcal{A}_h(t)$ denote the derivative of $\mathcal{A}_H(t)$ given by

$$\mathcal{A}_h(t) = \frac{d}{dt}\mathcal{A}_{\mathrm{H}}(t) = \frac{d}{dt}\left[\int_0^t H(x)\,dx - \frac{t}{2}H(t)\right] = \frac{1}{2}[H(t) - t \cdot h(t)] \qquad (10.6)$$

As for $\mathcal{A}_H(t)$, a simple graphical interpretation of $\mathcal{A}_h(t)$ leads to the following results: $\mathcal{A}_h(t)$ corresponds to half the difference between (1) the area delimited by $h(t)$, the failure intensity, and the coordinate axis and (2) the area of the rectangle $(t, h(t))$. Local trend change takes place at points $T_L$, which are such that $\mathcal{A}_h(T_L) = 0$ (where both areas are equal).

For a subadditive function, taking the first point of local trend change as the time origin would lead to a superadditive function from this new time origin to the following point of local trend change (since the curve giving the cumulative number of failures is concave over this time interval).

The preceding remarks also hold for a superadditive function. At the beginning, the difference between the two areas is negative and decreasing, and

- Without local reliability fluctuation, the mean value function is convex, leading to a negatively decreasing $\mathcal{A}_H(t)$ (Fig. 10.2, case B).
- In case of local reliability fluctuation, $\mathcal{A}_H(t)$ takes its first minimum at $T_L$ (such as $\mathcal{A}_h(T_L) = 0$). From $T_L$, $\mathcal{A}_H(t)$ is then increasing (indicating local reliability growth) up to the next point of local trend change, and so on.

### 10.2.4 Subadditive property and trend change

There exist more complex cases, however, where the cumulative number of failures is neither subadditive nor superadditive over the considered interval. Since the notion of subadditivity/superadditivity is related to a given interval, a change in the time origin leads to subadditive or superadditive functions over the subintervals of the new considered interval. Two such cases are depicted in Fig. 10.3.

For case C, the function is superadditive before $T_G$, denoting reliability decrease over $[0, T_G]$. $T_G$ corresponds to the point where $\mathcal{A}_H(t)$ changes signs ($\mathcal{A}_H(T_G) = 0$): the function is no longer superadditive. $T_L$ denotes the point where $\mathcal{A}_H(t)$ is no longer decreasing ($\mathcal{A}_h(T_L) = 0$), denoting local trend change. However, the function continues to be superadditive up to point $T_G$. On the subinterval of time between $T_L$ and $T$, the curve is concave, indicating reliability growth over $[T_L, T]$.

Situation D is the converse of that of C. Up to point $T_G$, the cumulative number of failures is subadditive, denoting reliability growth over $[0, T_G]$; from $T_G$, it is no longer subadditive. On $[T_L, T]$ the function is superadditive, corresponding to reliability decrease over this time interval. Combining C and D leads to situations where the trend may change more than once.

### 10.2.5 Some particular situations

What precedes shows that the notion of reliability growth is related to the interval of time considered and, thus, strongly associated with the origin of the time interval. Two types of particular situations can therefore be found in practice: (1) when the subadditivity factor is constant (or null) over a given interval and (2) when $\mathcal{A}_h(t)$ varies but remains positive over a given interval while the concavity of $H(t)$ may change. These two specific cases will be reviewed.

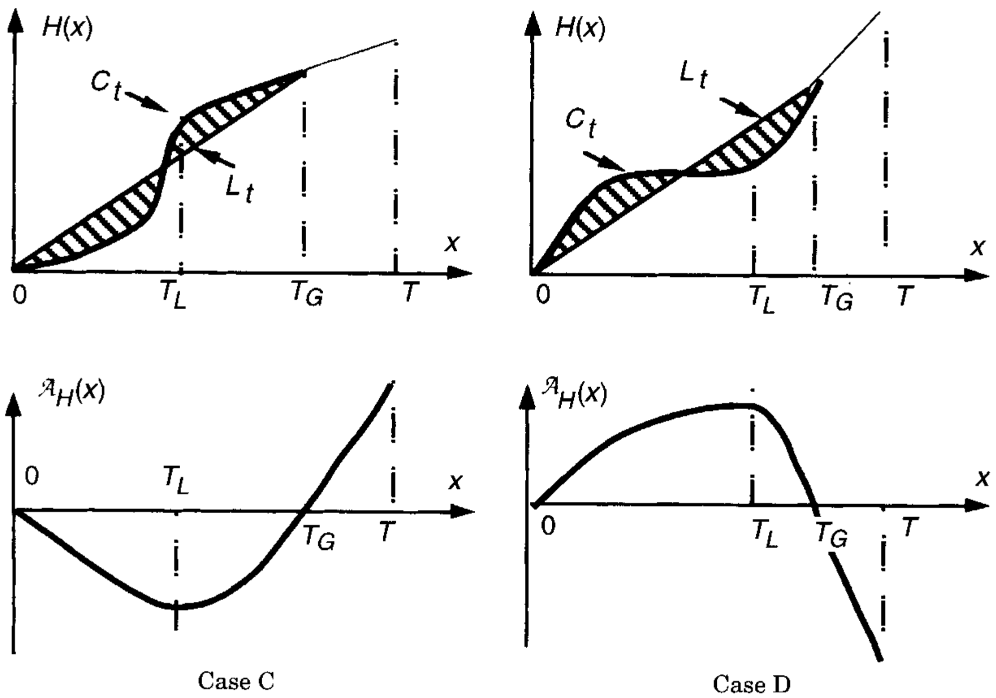


**Figure 10.3** Subadditivity/superadditivity and local trend variation.

The first case (where $\mathcal{A}_H(t)$ is constant or null over a given interval, say, $[t_1, t_2]$) is characterized by the fact that the derivative of $\mathcal{A}_H(t)$, $\mathcal{A}_h(t)$, is null over $[t_1, t_2]$. Integration of $H(t) - th(t) = 0$ leads to a linear cumulative number of failures function, i.e., a constant failure intensity over $[t_1, t_2]$. The constancy of $\mathcal{A}_H(t)$ thus indicates *stable reliability* over this time interval.

Finally, the case of $\mathcal{A}_h(t)$ being positive over a given interval while the concavity of $H(t)$ may change leads to the notion of transient or temporary behavior. Positive $\mathcal{A}_h(t)$ means that $\mathcal{A}_H(t)$ is not decreasing over the considered interval. This is shown in the example in Fig. 10.4. $H(t)$ is subadditive; $h(t)$ is fluctuating, leading to $H(t)$ concavity change; whereas $\mathcal{A}_H(t)$ is not decreasing ($\mathcal{A}_H(t)$ is not shown). The transient variations of $h(t)$ cannot be detected by the sign of $\mathcal{A}_H(t)$ and do not correspond to a trend variation as defined by the subadditive property. They are due to the random nature of the process and are identified as a *transient* or *temporary* behavior of the software.

Defining reliability growth through the subadditive property is thus very attractive since it is not sensitive to the transient and temporary behavior. The subadditive property constitutes a form of smoothing of the software behavior, as shown in Fig. 10.4.

### 10.2.6 Summary

Reliability growth/decrease is entirely characterized by the subadditivity factor $\mathcal{A}_H(t)$ and its derivative $\mathcal{A}_h(t)$. $\mathcal{A}_H(t)$ gives information about the trend on average over a given interval, whereas $\mathcal{A}_h(t)$ informs about local trend variation as follows:

- $\mathcal{A}_H(t) \geq 0$ over $[0, T]$ implies reliability growth on average over $[0, T]$.
- $\mathcal{A}_H(t) \leq 0$ over $[0, T]$ implies reliability decrease on average over $[0, T]$.
- $\mathcal{A}_H(t)$ constant over $[0, T]$ implies stable reliability on average over $[0, T]$.
- Changes in the sign of $\mathcal{A}_H(t)$ indicate reliability trend changes.

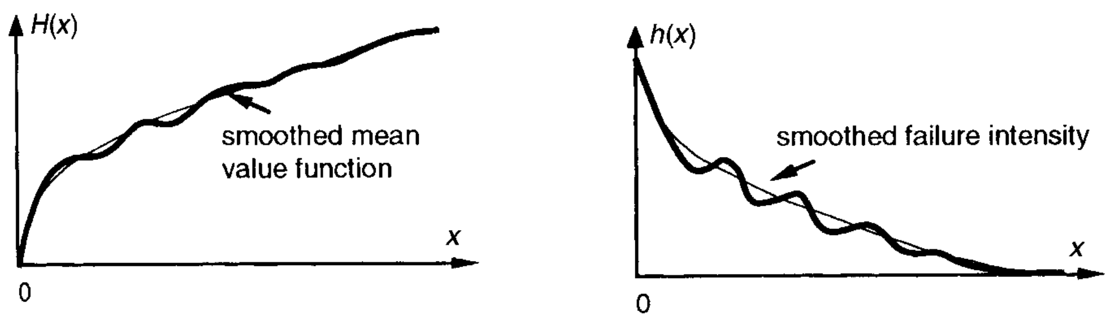


**Figure 10.4** Subadditivity and transient/temporary behavior.

- $\mathcal{A}_h(t) \geq 0$ over a subinterval $[t_1, t_2]$ implies local reliability growth over $[t_1, t_2]$.
- $\mathcal{A}_h(t) \leq 0$ over a subinterval $[t_1, t_2]$ implies local reliability decrease over $[t_1, t_2]$.
- Changes in the sign of $\mathcal{A}_h(t)$ indicate local reliability trend changes.
- Transient variations of the failure intensity are not detected by the subadditivity property.

## 10.3 Trend Analysis

Reliability growth can be analyzed by trend tests. In this section we will present only the most often used and most significant trend tests and place the emphasis on the relationship between the Laplace test (the most common one) and the subadditive property. The presentation of the tests is followed by a discussion of how they can be used to follow up software reliability.

Failure data can be collected in one of two forms: *interfailure times* or *number of failures per unit of time.* These two forms are related. Knowing the interfailure times enables us to obtain the number of failures per unit of time (the second form needs less precise data collection).

The use of data in the form of number of failures per unit of time reduces the impact of transient variations on software reliability analysis and evaluation. The unit of time is a function of the type of system usage as well as the number of failures occurring during the considered units of time, and it may be different for different phases. For instance, since more failures are likely to occur during development, the selected unit of time may be smaller than the one selected for operational life.

### 10.3.1 Trend tests

There are a number of trend tests which can be used to help determine whether the system undergoes reliability growth or decrease. These tests can be grouped into graphical and analytical tests [Asch84]. Graphical tests consist of plotting some observed failure data such as the interfailure times or the number of failures per unit of time versus time in order to visually obtain the trend displayed by the data. As such they are informal. Analytical tests correspond to more rigorous tests since they are based on statistical considerations. The raw data are processed to derive trend factors. The principle of analytical tests is to test a null hypothesis $H_0$ versus an alternative $H_1$. Usually, $H_0$ corresponds to one of the following assumptions for the underlying process: it is assumed to be either (1) a homogeneous Poisson process

(HPP) or (2) a stationary renewal process. Very often, $H_1$ corresponds to "the process undergoes monotonic trend," i.e., increasing (decreasing) interfailure times or decreasing (increasing) failure intensity.

Theoretical definition and comparison of analytical trend tests have given rise to several publications [Cox66, Asch84, Gaud90]. In the latter reference, detailed presentation, analysis, and comparison of some analytical tests (e.g., Laplace, MIL-HDBK 189, Gnedenko, Spearman, and Kendall tests) are presented. In particular, it is shown that

- From a practical point of view, all these tests yield similar results for the detection of reliability trend variations.
- The Spearman and Kendall tests have the advantage of being based on less restrictive assumptions (that is, $H_0$: the underlying process is a renewal process).
- The Gnedenko test is interesting since it uses exact distributions.
- From the optimality point of view, the Laplace test is superior and recommended for use when the NHPP assumption is made (even though its significance level is not exact and its power cannot be estimated).

These results confirm our experience in the processing of real failure data. We have observed the agreement between the results of these various tests and the superior efficiency of the Laplace test.

All the aforementioned tests are performed relative to a monotonic trend. Linked with the subadditive property, a test for subadditivity (referred to subsequently as the Hollander test) was derived by Hollander and Proschan in [Holl74] and Hollander in [Holl78]. The Hollander test deals with $H_0$ and $H_1$ defined by:

$H_0$: the failure process is an HPP

$H_1$: the mean value function is superadditive

It is thus more general than the previous ones and also more in line with our definition of reliability growth/decrease. Further details on the Laplace and Hollander tests will be provided in the following subsections.

The trend can be analyzed using interfailure times data or failure intensity data, both of which we will now examine.

**10.3.1.1 Interfailure times.** Two trend tests are commonly carried: the arithmetical mean and the Laplace tests. The *arithmetical mean* of the interfailure times is a popular test. It consists of calculating the arithmetical mean $\tau(i)$ of the observed interfailure times $\theta_j$, $j = 1, 2, \ldots, i$ ($\theta_j$ are realizations of $T_j$):

$$\tau(i) = \frac{1}{i} \sum_{j=1}^{i} \theta_j \tag{10.7}$$

An increasing series of $\tau(i)$ indicates reliability growth and, conversely, a decreasing series suggests reliability decrease. This straightforward test is directly related to the observed data. A variant of this test consists of evaluating the mean of interfailure times over periods of time of the same length in order to put emphasis on the local trend variation.

Let $N(T)$ denote the cumulative number of failures over the observation period $[0, T]$. The *Laplace test* [Cox66] consists of calculating Laplace factor, $u(T)$ which is derived as follows. The occurrence of events is assumed to follow an NHPP whose failure intensity is decreasing and given by

$$h(t) = e^{a+bt} \qquad b \leq 0 \tag{10.8}$$

If $b = 0$, the Poisson process becomes homogeneous and the occurrence rate is time-independent. Under this hypothesis ($b = 0$, that is, $H_0$: the failure process is an HPP), the test procedure is to compute:

$$u(T) = \frac{\dfrac{1}{N(T)} \displaystyle\sum_{n=1}^{N(T)} \sum_{j=1}^{n} \theta_j - \dfrac{T}{2}}{T \sqrt{\dfrac{1}{12N(T)}}} \tag{10.9}$$

This factor may be evaluated step by step, after each failure occurrence, for instance. In this case $T$ is equal to the time of a failure occurrence, say, failure $i$, and failure at time $T$ is to be excluded. Equation (10.9) thus becomes

$$u(i) = \frac{\dfrac{1}{i-1} \displaystyle\sum_{n=1}^{i-1} \sum_{j=1}^{n} \theta_j - \dfrac{\sum_{j=1}^{i} \theta_j}{2}}{\displaystyle\sum_{j=1}^{i} \theta_j \sqrt{\dfrac{1}{12\,(i-1)}}} \tag{10.10}$$

Practical use of Laplace test in the context of reliability growth can be summarized as follows:

- Negative values of the Laplace factor indicate a decreasing failure intensity ($b < 0$).
- Positive values suggest an increasing failure intensity ($b > 0$).
- Values varying between −2 and +2 indicate stable reliability.

These practical considerations are derived from the significance levels associated with the normal distribution; e.g., for a significance level of 5 percent,

- The null hypothesis "$H_0$ : HPP" versus "$H_1$ : the failure intensity is decreasing" is rejected for $u(T) < -1.645$.
- The null hypothesis "$H_0$ : HPP" versus "$H_1$ : the failure intensity is increasing" is rejected for $u(T) > 1.645$.
- The null hypothesis "$H_0$ : HPP" versus "$H_1$ : there is a trend" is rejected for $|u(T)| > 1.96$.

The Laplace test can be simply interpreted as follows:

- $T/2$ is the midpoint of the observation interval.
- $1/[N(T) \sum_{n=1}^{N(T)} \sum_{j=1}^{n} \theta_j]$ corresponds to the statistical center of the interfailure times.

Under the assumption of failure intensity decrease (increase), the interfailure times $\theta_j$ will tend to occur before (after) the midpoint of the observation interval; hence the statistical center tends to be smaller (larger) than the mid-interval.

**10.3.1.2 Failure intensity and cumulative number of failures.** Two very simple graphical tests can be used: the plots giving the evolution of the observed cumulative number of failures and the failure intensity (i.e., the number of failures per unit of time) versus time, respectively. The inevitable local fluctuations exhibited by experimental data make smoothing necessary before the reliability trend can be determined, and favor the cumulative number of failures rather than failure intensity. Reliability trend is then related to the subadditive property of the smoothed plot, as seen in Sec. 10.2.

The formulation of the *Laplace test* for failure intensity (or cumulative number of failures) is as follows. Let the time interval $[0, T]$ be divided into $k$ units of time of equal length, and let $n(i)$ be the number of failures observed during time unit $i$. Following the method outlined in [Cox66], the expression of the *Laplace factor* is given by (for a detailed derivation consult [Kano91a])

$$u(k) = \frac{\sum_{i=1}^{k} (i-1)n(i) - \frac{(k-1)}{2} \sum_{i=1}^{k} n(i)}{\sqrt{\frac{k^2-1}{12} \sum_{i=1}^{k} n(i)}} \qquad (10.11)$$

The same results as previously stated apply: negative values of $u(k)$ indicate a decreasing failure intensity (reliability growth) whereas positive values point out an increasing failure intensity (reliability decrease).

The *Hollander test* for subadditivity [Holl74] consists of evaluating the statistic $Q_n$ based on the times of failures

$$s_i = \sum_{j=1}^{i} \theta_j$$

$$Q_n = 2K_n/n(n-1)(n-2)$$

$$K_n = \sum{}^{*} \; [\phi(s_{\alpha 3} + s_{\alpha 2}, T) - \phi(s_{\alpha 3}, s_{\alpha 1} + s_{\alpha 2}) \cdot \phi(s_{\alpha 1} + s_{\alpha 2}, T)] \qquad (10.12)$$

where $n = N(T)$, $\phi(a, b) = 1$ if $a \le b$, else $\phi(a, b) = 0$, and $\sum^*$ is over all $\frac{1}{6}n(n-1)(n-2)$ choices of subscripts such that $1 \le \alpha 1 < \alpha 2 < \alpha 3 \le n$. Critical values of the $K_n$ statistic are given in the same reference for various levels of significance.

**10.3.1.3 Relationship between the Laplace test and the subadditive property.** The Laplace test may be used in the same way as any statistical test with significance levels as indicated above. However, we derive a relationship between the Laplace factor and the subadditivity factor allowing extension of the properties of the latter.

Let $n(i)$ denote the number of failures during the $i$th unit of time (i.e., $N(k) = \sum_{i=1}^{k} n(i)$). The numerator of Eq. (10.11) can be written as

$$\sum_{i=1}^{k} (i-1)[N(i) - N(i-1)] - \frac{(k-1)}{2} N(k)$$

which is equal to

$$\left[kN(k) - \sum_{i=1}^{k} N(i)\right] - \frac{(k-1)}{2} N(k) = \frac{k+1}{2} N(k) - \sum_{i=1}^{k} N(i)$$

Equation (10.11) thus becomes

$$u(k) = -\frac{\displaystyle\sum_{i=1}^{k} N(i) - \frac{k+1}{2} N(k)}{\sqrt{\dfrac{k^2-1}{12} N(k)}} \qquad (10.13)$$

The $u(k)$ numerator is nothing other than the subadditivity factor, $\mathcal{A}_H(k)$. Therefore, testing the sign of $u(k)$ leads to testing the sign of the difference of areas between the curve plotting the cumulative number of failures and the chord joining the origin and the current cumulative number of failures. This shows that the Laplace factor (fortunately) integrates the unavoidable local fluctuations which are typical of experimental data, because the numerator of this factor is directly related to the subadditive property.

In the previous section certain features related to the subadditive property were introduced (i.e., reliability growth over a given interval and local trend change). We use a simple hypothetical example to illustrate the relationship between these features and the Laplace test. Figure 10.5 shows the failure intensity considered, the corresponding cumulative number of failures, $N(k)$, the derived subadditivity factor, and the evaluated Laplace factor.

Considering the whole data set (Figure 10.5*a*) leads to the following comments:

- $\mathcal{A}_H(k)$ is negative until point 9, thus indicating superadditivity and hence reliability decrease up to this point.
- The trend of $\mathcal{A}_H(k)$ changes around point 6, indicating local trend change. (This is also noticed when looking directly at the failure intensity, which is decreasing from point 6.)
- The sign and variations of the Laplace factor follow the sign and variations of $(-\mathcal{A}_H(k))$.

If we consider the data set from point 6 and plot the same measures (Fig. 10.5*b*), the results of this time origin change are as follows:

- $\mathcal{A}_H(k)$ is positive for each point showing reliability growth over [6, 21].
- The Laplace factor becomes negative.

To conclude, the denominator in the expression of the Laplace factor (Eq. (10.13)) usually does the following:

- Amplifies the Laplace factor variations when compared to those of $\mathcal{A}_H(k)$, at the beginning of the time interval considered
- Reduces the scale variation of the Laplace factor when compared to the variations of $\mathcal{A}_H(k)$ on the whole time interval, as it acts as a norming factor

It is also worth noting that changes in the time origin similarly impacts the subadditivity factor and the Laplace factor. However, the change in the time origin does not result in a simple translation in all situations.

The preceding statements are now illustrated by a more complex case corresponding to the experimental data of the TROPICO-R

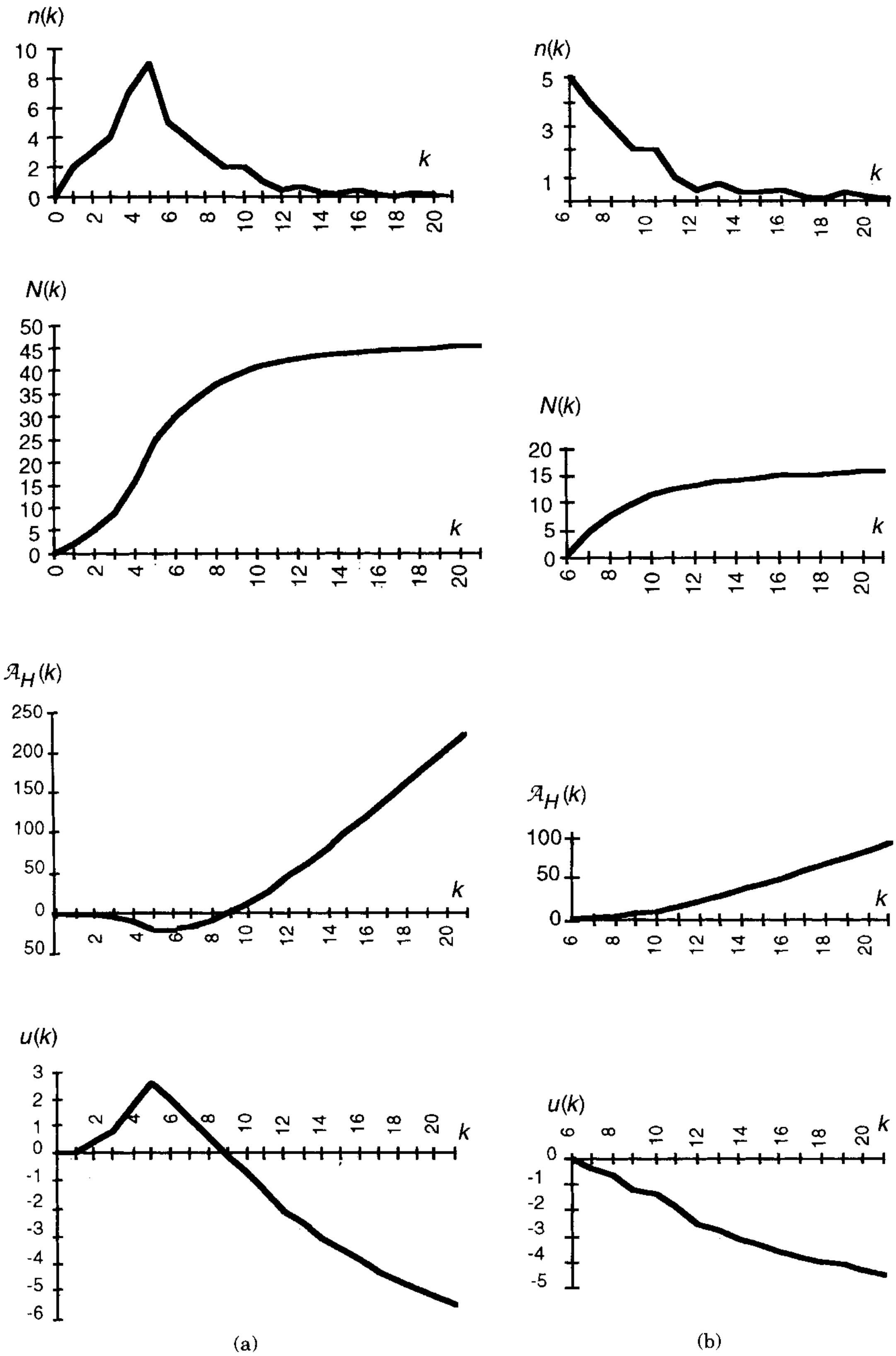


**Figure 10.5** Example of relationship between failure intensity, cumulative number of failures, subadditivity factor, and Laplace factor.

switching system studied in [Kano91a]. Figure 10.6 gives the Laplace factor for the whole data set from validation to operation. At the beginning of validation, a reliability decrease took place, due to the occurrence of 28 failures during the third unit of time, whereas only 8 failures had occurred during the first two time units, and 24 failures occurred during the next two time units. This is a common situation at the start of validation: reliability decrease is due to the activation of a large number of faults. Applying the Laplace test without the data belonging to the two first units of time leads on average to reliability growth from unit time 3 despite local trend changes (Fig. 10.7). The evolution of the subadditivity factor $\mathcal{A}_H(k)$ is shown in Fig. 10.8, which also depicts the influence of the first two data items. Even though only two data items were removed, the comparison of Figs. 10.6 and 10.7 (and of the curves of Fig. 10.8) confirms the previous remarks, that is,

- The curve's shape is preserved when suppressing the first data items, which cause the reliability decrease at the beginning and preservation of the local trend change points.
- $\mathcal{A}_H(k)$ seems less sensitive to the local trend variations when compared to the Laplace factor. This difficulty may be overcome by considering, for instance, a reduced data set as indicated in Fig. 10.9, which depicts $\mathcal{A}_H(k)$ over the validation phase only.

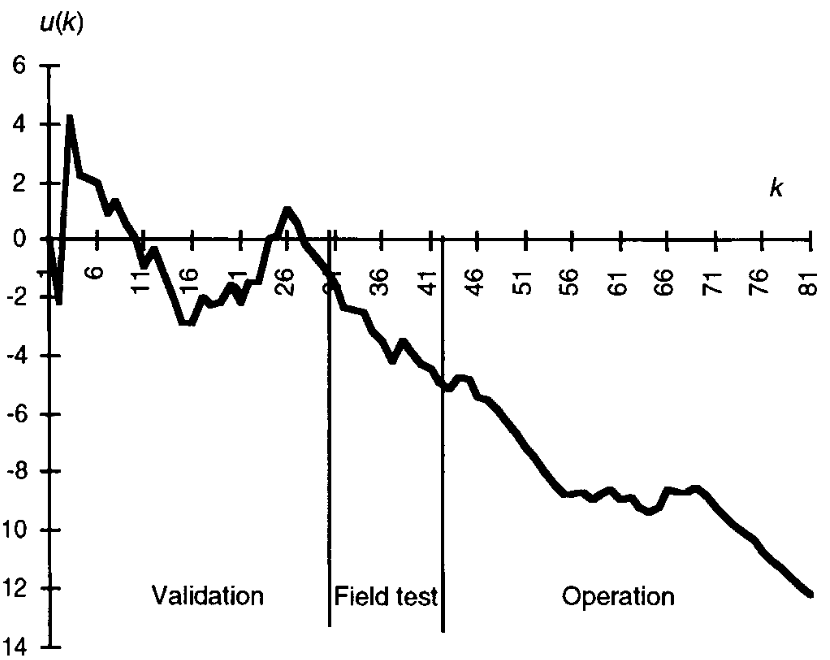


**Figure 10.6** Laplace factor for the TROPICO-R considering the whole data set.

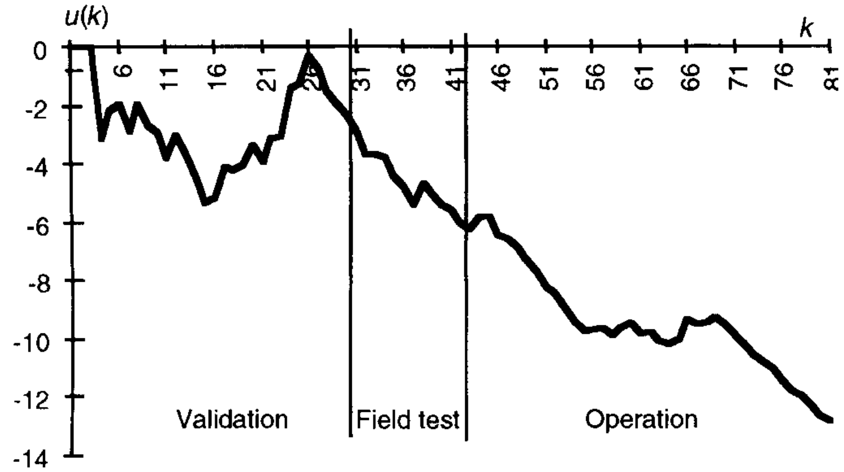


**Figure 10.7** Laplace factor for TROPICO-R without considering the first two units of time.

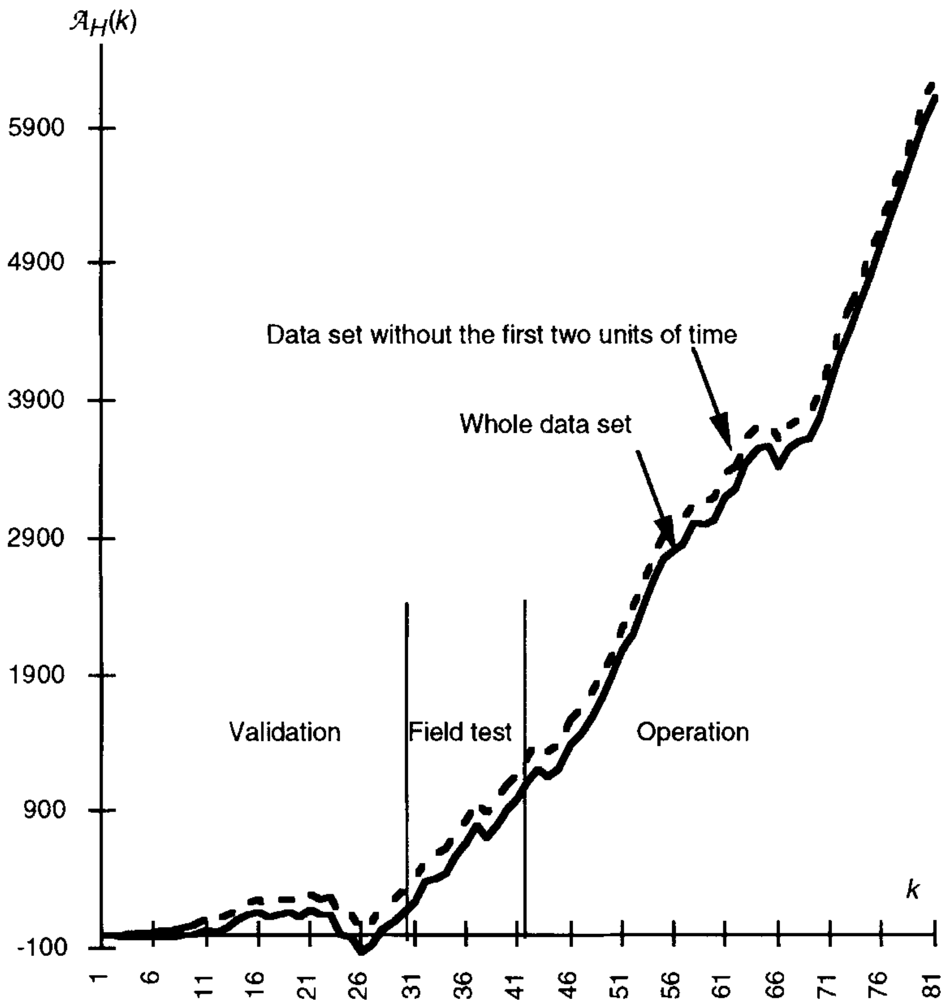


**Figure 10.8** Subadditivity factor for TROPICO-R with and without considering the first two units of time.

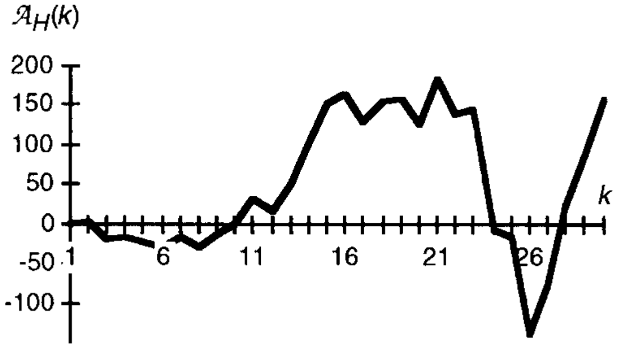


**Figure 10.9** Subadditivity factor for TROPICO-R for the validation phase.

### 10.3.2 Example

By way of example, we illustrate the features of the Laplace factor introduced in the previous paragraph. The data are those collected on system 2 presented in [Musa79], denoted S2 in this chapter. Fifty-four failures occurred during the observation period covering system testing.

The left column of Table 10.1 gives the number of failures, $i$; the second column lists the execution time (in seconds) from system restart after failure $i - 1$ to failure $i$. We use Eq. (10.10) to evaluate the Laplace factor, $u(i)$, from $i = 2$ up to $i = 54$. Numerical values of $u(i)$ are given in the third column of Table 10.1. Certain values are worth commenting upon. Time to failure 2 is larger than time to failure 1. As a result $u(2)$ is negative, indicating reliability growth. The times to failures are still increasing up to failure 8, and the Laplace factor is also negative and decreasing, indicating reliability growth over [1, 8]. Time to failure 9 is smaller, thus evidencing local variation which leads to an increasing but still negative Laplace factor, indicating reliability growth over [1, 9], and so on. The Laplace factor is indeed negative over all the period considered. It is illustrated in Fig. 10.10*a*. This figure shows that from failures 31 to 41, $u(i)$ is increasing, indicating a local reliability decrease over this time interval despite a global reliability growth. If we consider data items from failure 31 only and evaluate again $u(i)$ for $i = 32$ to 54, the Laplace factor becomes positive up to $i = 42$, confirming reliability decrease over this period of time. The numerical results are given in the right column of Table 10.1, and the corresponding curve in Fig. 10.10*b*.

Consider the failure intensity for the same system. It is obtained by computing the number of failures over periods of time of equal length (called *units of time*). Cumulative execution time obtained by summing all the times to failures is 108,708 s. The unit of time considered is 5000 seconds of execution time, which leads to 22 units of time. This choice is a trade-off. Indeed, if a small unit of time (2000–3000 s) were considered it would lead to several units of times during which zero failure would be observed mainly during the last period of observation, and, on

**TABLE 10.1 Times to Failures and Laplace Factor for System S2**

**(In seconds of execution time)**

| Failure no. ($i$) | Time to failure (seconds) | Laplace factor $u(i)$ | Laplace factor (from 31) |
|---|---|---|---|
| 1 | 191 | 0.00 | |
| 2 | 222 | –0.13 | |
| 3 | 280 | –0.31 | |
| 4 | 290 | –0.36 | |
| 5 | 290 | –0.36 | |
| 6 | 385 | –0.55 | |
| 7 | 570 | –0.93 | |
| 8 | 610 | –1.15 | |
| 9 | 365 | –0.97 | |
| 10 | 390 | –0.86 | |
| 11 | 275 | –0.64 | |
| 12 | 360 | –0.57 | |
| 13 | 800 | –0.99 | |
| 14 | 1,210 | –1.59 | |
| 15 | 407 | –1.4 | |
| 16 | 50 | –0.95 | |
| 17 | 660 | –1.07 | |
| 18 | 1,507 | –1.73 | |
| 19 | 625 | –1.67 | |
| 20 | 912 | –1.78 | |
| 21 | 638 | –1.72 | |
| 22 | 293 | –1.47 | |
| 23 | 1,212 | –1.75 | |
| 24 | 612 | –1.67 | |
| 25 | 675 | –1.64 | |
| 26 | 1,215 | –1.86 | |
| 27 | 2,715 | –2.6 | |
| 28 | 3,551 | –3.35 | |
| 29 | 800 | –3.19 | |
| 30 | 3,910 | –3.79 | |
| 31 | 6,900 | –4.68 | 0.00 |
| 32 | 3,300 | –4.78 | 1.88 |
| 33 | 1,510 | –4.62 | 2.07 |
| 34 | 195 | –4.28 | 2.62 |
| 35 | 1,956 | –4.22 | 2.41 |
| 36 | 135 | –3.89 | 2.88 |
| 37 | 661 | –3.65 | 3.07 |
| 38 | 50 | –3.33 | 3.47 |
| 39 | 729 | –3.14 | 3.53 |
| 40 | 900 | –2.99 | 3.50 |
| 41 | 180 | –2.71 | 3.79 |
| 42 | 4,225 | –3.19 | 2.44 |
| 43 | 15,600 | –4.97 | –0.59 |
| 44 | 0 | –4.65 | –0.09 |
| 45 | 0 | –4.34 | 0.38 |
| 46 | 300 | –4.07 | 0.76 |
| 47 | 9,021 | –4.77 | –0.31 |
| 48 | 2,519 | –4.72 | –0.23 |
| 49 | 6,890 | –5.06 | –0.69 |
| 50 | 3,348 | –5.06 | –0.67 |
| 51 | 2,750 | –5.00 | –0.58 |
| 52 | 6,675 | –5.25 | –0.91 |
| 53 | 6,945 | –5.48 | –1.19 |
| 54 | 7,899 | –5.73 | –1.49 |

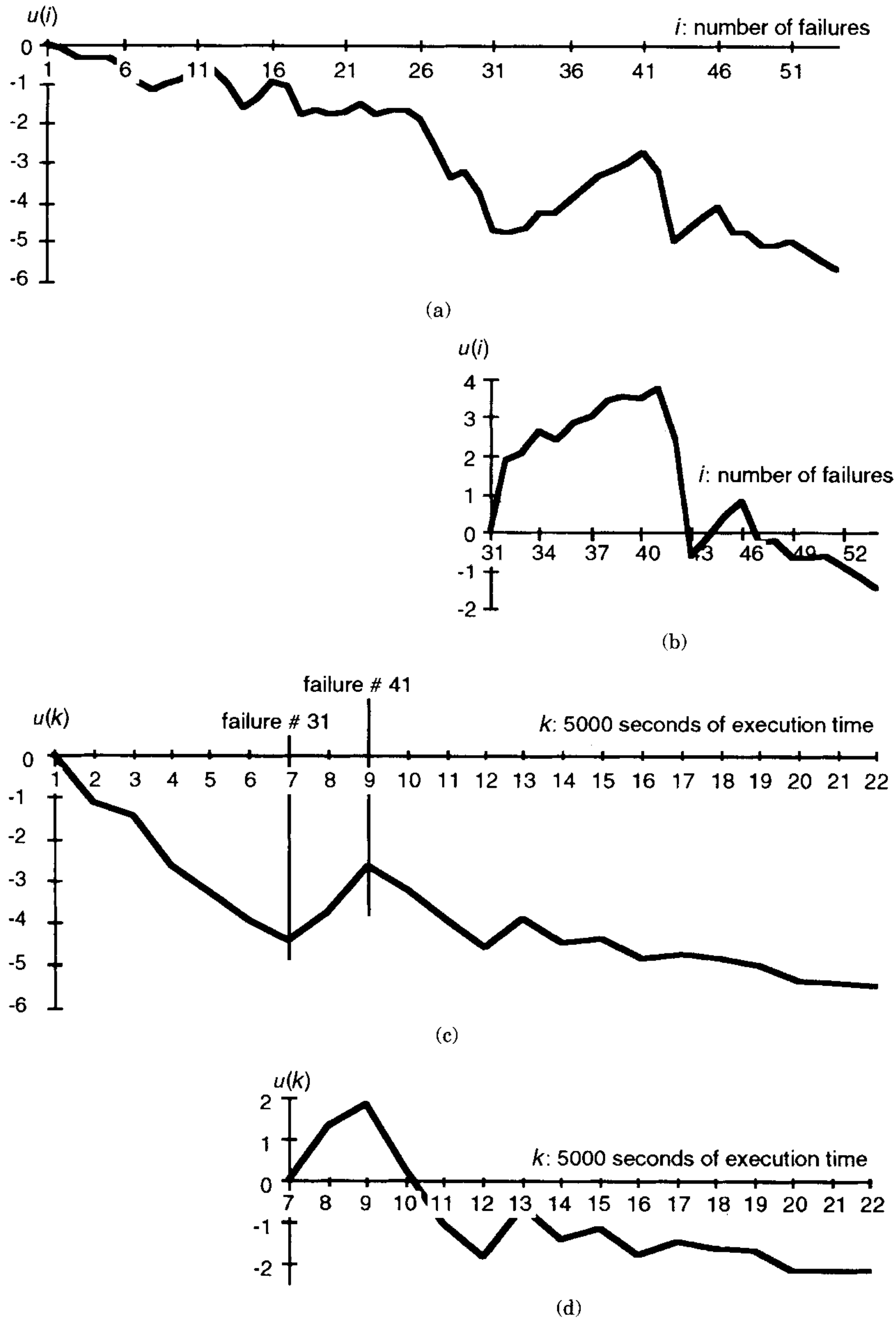


**Figure 10.10** Laplace factor for system S2. (*a*) Laplace factor for the times to failure considering the whole data set. (*b*) Laplace factor for the times to failure from failure 31. (*c*) Laplace factor for the failure intensity considering the whole data set. (*d*) Laplace factor for the failure intensity from unit time 7.

the other hand, a larger unit of time would lead to a smaller number of units of time and hence to less detailed information.

The first three columns of Table 10.2 list, respectively, the number of time unit, the corresponding number of failures during this time unit, and the Laplace factor evaluated using Eq. (10.11) or Eq. (10.13), which are equivalent. The Laplace factor is also displayed in Fig. 10.10*c* and shows a reliability decrease between unit times 7 and 9. Failure 31 occurred during unit of time 7, and failure 42 during unit of time 9. The comparison of Fig. 10.10*a* and *c* shows that the Laplace factor gives similar results when considering failure intensity or times to failure. Application of the Laplace test to failure data from the 7th unit of time (last column of Table 10.2 and Fig. 10.10*d*) leads to results similar to those obtained when considering data items from failure 31.

### 10.3.3 Typical results that can be drawn from trend analyses

Trend analyses are a great help when it comes to appreciating the efficiency of test activities and controlling their progress. They are particularly helpful for following up the software development. Indeed, graphical tests are often used in the industry [Grad87, Ross87,

**TABLE 10.2 Number of Failures Per Periods of 5000 s and Laplace Factor for System S2**

| Unit of time ($k$) | Failure intensity | Laplace factor $u(k)$ | Laplace factor (from 7) |
|---|---|---|---|
| 1 | 12 | 0.00 | |
| 2 | 7 | −1.15 | |
| 3 | 6 | −1.47 | |
| 4 | 2 | −2.67 | |
| 5 | 2 | −3.28 | |
| 6 | 1 | −3.96 | |
| 7 | 1 | −4.40 | 0.00 |
| 8 | 4 | −3.73 | 1.34 |
| 9 | 6 | −2.66 | 1.85 |
| 10 | 1 | −3.22 | 0.26 |
| 11 | 0 | −3.95 | −1.02 |
| 12 | 0 | −4.56 | −1.86 |
| 13 | 4 | −3.90 | −0.62 |
| 14 | 0 | −4.46 | −1.42 |
| 15 | 2 | −4.38 | −1.19 |
| 16 | 0 | −4.85 | −1.81 |
| 17 | 2 | −4.71 | −1.48 |
| 18 | 1 | −4.84 | −1.61 |
| 19 | 1 | −4.96 | −1.71 |
| 20 | 0 | −5.34 | −2.17 |
| 21 | 1 | −5.40 | −2.17 |
| 22 | 1 | −5.45 | −2.17 |

Vale88]—even though they are referred to differently (e.g., descriptive statistics or control charts).

Also worth pointing out here is that the role of the trend analysis is only to *draw attention* to problems that might otherwise pass unnoticed until it is too late, thus providing an early warning likely to speed up the search for a solution. Trend analysis can be used to enrich the interpretation of someone who knows the software from which the data are derived, as well as the development process and the user environment.

In the following, three typical situations are outlined: reliability decrease, reliability growth, and stable reliability.

*Reliability decrease* at the beginning of a new activity is generally expected and considered normal. Examples of such activities are

- New life cycle phase
- Change in the test sets within the same phase
- Addition of new users
- Activation of the system in a different profile of use

Also, reliability decrease may result from regression faults. Trend analysis allows this kind of behavior to be noticed. If the duration of the decrease period seems long, there may be cause for alarm. In some situations, if it continues to decrease there may be some problems within the software. Analyzing the reasons for this decrease as well as the nature of the activated faults is of prime importance in these kinds of situations. Such analysis may result in the decision to reexamine the corresponding software part.

*Reliability growth* following reliability decrease is usually welcomed, since it indicates that, after removal of the first faults, the corresponding activity reveals fewer and fewer faults. When calendar time is used, mainly in operational life, sudden reliability growth may result from a period of time during which the system is less used or not used at all; it may also be due to some failures that are not recorded. When this is noticed, particular care must be taken and, more important, the reasons for this sudden increase have to be analyzed.

*Stable reliability* indicates that either (1) the software does not undergo corrective maintenance or (2) the corrective actions performed have no visible effect on reliability. When the software is under validation, stable reliability with almost no failures means that the corresponding activity has reached a saturation: the application of the corresponding test sets does not reveal new faults. One has to either stop testing, introduce new test sets, or proceed to the next phase. More generally, it is recommended that the application of a test set continue as long as it exhibits reliability growth and end when stable reliability with almost no failures is reached. Thus, in practice, if stable reliability has not been reached the validation team (and the manager) may decide to continue testing before software delivery (since it will be more efficient and cost-effective to do so) and to remove faults during validation rather than during operation.

Finally, trend analyses may greatly help reliability growth models to give better estimations, since they can be applied to data displaying trends in accordance with their assumptions rather than blindly. Applying reliability growth models blindly may lead to nonrealistic results when the trend displayed by the data differs from that assumed by the model. Failure data can be partitioned according to the trend:

- In case of reliability growth, most of the existing reliability growth models can be applied.
- In case of reliability decrease, only models allowing an increasing failure intensity can be applied.
- When the failure data exhibit reliability decrease followed by reliability growth, an S-shaped model [Ohba84] can be applied.
- When stable reliability is noticed, a constant failure intensity model can be applied (HPP model); reliability growth models are in fact not needed.

### 10.3.4 Summary

In this section, we have presented a few trend tests and tried to keep the presentation as simple as possible by skipping mathematical developments and giving graphical and practical interpretations. The Laplace and the Hollander tests can be used as conventional statistical tests with significance levels. However, the graphical interpretation of the subadditive property and the link between the Laplace factor and the subadditivity factor enable both local trend change and trend on average to be identified at a glance. In practice, we will mainly plot the Laplace factor and possibly the subadditivity factor in order to follow up the software reliability. Processing of the failure data from system S2 showed the benefit of using the Laplace factor to identify local trend changes as well.

## 10.4 Application to Real Systems

This section is intended to illustrate the type of results that may be expected from trend analysis during development and operational phases, as well as from the application of reliability growth models.

Since the previous section showed the link between the Laplace factor and the subadditivity factor, we will use both. The aim of this section is simply to illustrate some of the points introduced in the previous section, and not to make detailed analyses of the data sets considered. For further details about the systems considered, you may consult the publications referenced. We will analyze the trend of five data sets, some of them being in the form of times to failures and the others being in the form of failure intensity. In order to show the consistency of the results derived from the various trend tests, more than one test will be applied for some data sets.

The considered software systems are as follows:

1. System *SS4* of [Musa79]
2. The system also considered in [Musa79] referred to as *S27*
3. The system corresponding to the switching system of section 2 in [Kano91a], called *SS1*
4. The so-called *SS2* system, corresponding to the switching system observed during validation and part of operational life [Kano93b]
5. The system corresponding to an avionic application, referred to as *SAV*

For each one, we give the results of the trend analysis and comment on the type of reliability growth models that could be (or has been) used. The analyses are performed using SoRel, a tool for reliability analysis and evaluation presented in App. A.

### 10.4.1 Software of system SS4

Failure data gathered on SS4 correspond to operational life. Application of the arithmetical mean test in Fig. 10.11*a* shows that the mean time to failure is almost constant: about 230,000 units of time. The corresponding Laplace factor given in Fig. 10.11*b* oscillates between −2 and +2, also indicating stable reliability for a significance level of 5 percent.

As for system S2 considered in Sec. 10.3, we evaluate the failure intensity considering a unit of time of $10^6$ seconds of execution time. The application of Laplace test to the failure intensity (displayed in Fig. 10.11*c*) also indicates stable reliability at the same significance level. For this system, a constant failure rate (i.e., HPP model) is well adapted to model the software behavior and is simpler to apply than a reliability growth model. This is not surprising since the software was in operational life without corrective maintenance. The behavior of the software is thus similar to that of the hardware:

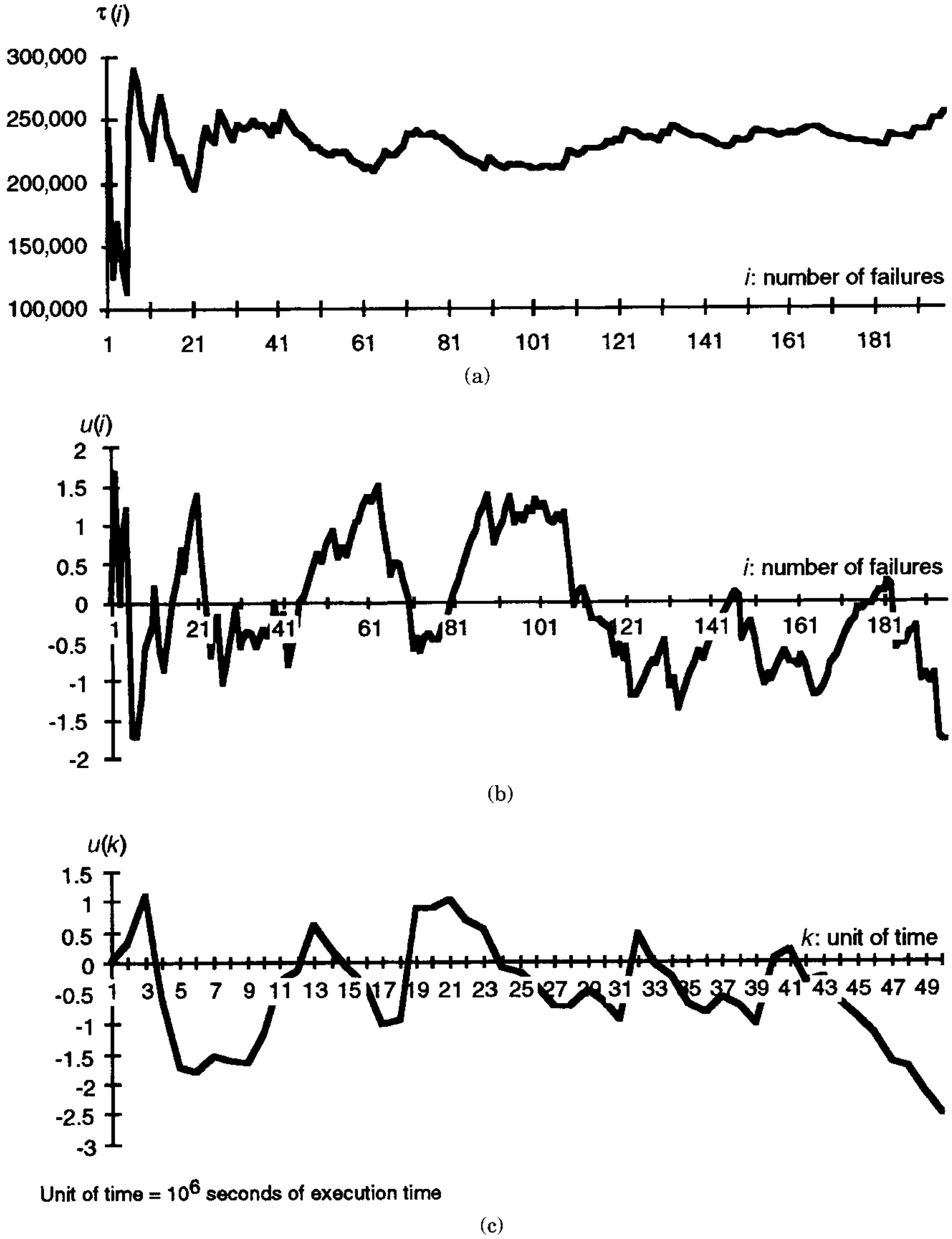


**Figure 10.11** Trend tests for SS4. (*a*) Arithmetical mean of the times to failure. (*b*) Laplace factor of the times to failure. (*c*) Laplace factor of the failure intensity.

- For the hardware, the repair actions are intended to replace the failed part by another one which is identical.
- For nonmaintained software, the system is restarted with an input pattern different from the one having led to failure.

In both cases the system's ability to deliver a proper service is preserved (i.e., stochastic identity of the successive times to failure).

### 10.4.2 Software of system S27

S27 is an example of systems that exhibit two phases of stable reliability. The transition between them took place around the 24th failure, as indicated in Fig. 10.12. This system was under test and the reasons of this sudden change (which may be due to a singular behavior of the software) must be investigated. Unfortunately, the published data did not allow us to identify the reasons of this behavior. In this case, data may be partitioned into two subsets, each one being modeled by a constant failure rate: the failure rate of the second subset (from 24 to 42) being lower than the failure rate of the first. If we remove failure data up to failure 23 and again apply the Laplace test, the corresponding factor shown in Fig. 10.12*c* confirms the stable reliability over [24–41], except for two points.

Figure 10.12*a* and *b* illustrates the link between a graphical test (the mean of the interfailure times) and the results of the Laplace factor. Both of them point out the discontinuity in software behavior.

### 10.4.3 Software of system SS1

Trend tests accounting for the whole data set collected on this system are displayed in Figs. 10.6 and 10.7. Applying the Laplace test separately to each phase (ignoring data collected during the previous phases) is illustrated in Fig. 10.13.

The following comments apply to both Figs. 10.6 and 10.13:

- Reliability decrease from $k = 14$ to $k = 25$ was induced by the changes in the nature of the tests within the validation phase. This period corresponds to the application of *quality* and *performance* tests after *functional* tests in the previous period. This decrease is due to their dynamic nature (traffic simulation) which has activated new parts of the program.
- Transitions from validation to field test and from field test to operation did not give rise to a reliability discontinuity, which means that the tests applied during the end of validation are representative of operational conditions.
- Figure 10.6 indicates that from $k = 55$ up to $k = 70$ reliability tends to be stabilized: $u(k)$ is almost constant, suggesting stable reliability. However, when considering the trend results obtained for operational data only in Fig. 10.13 we notice in fact a reliability decrease over this time interval. The difference in perception of the reliability variation (from reliability growth to stable reliability or reliability decrease) is related to the interval of time considered. When considering the whole data set, a relative stable reliability is observed, and when considering only operational failure data, a relative reliability

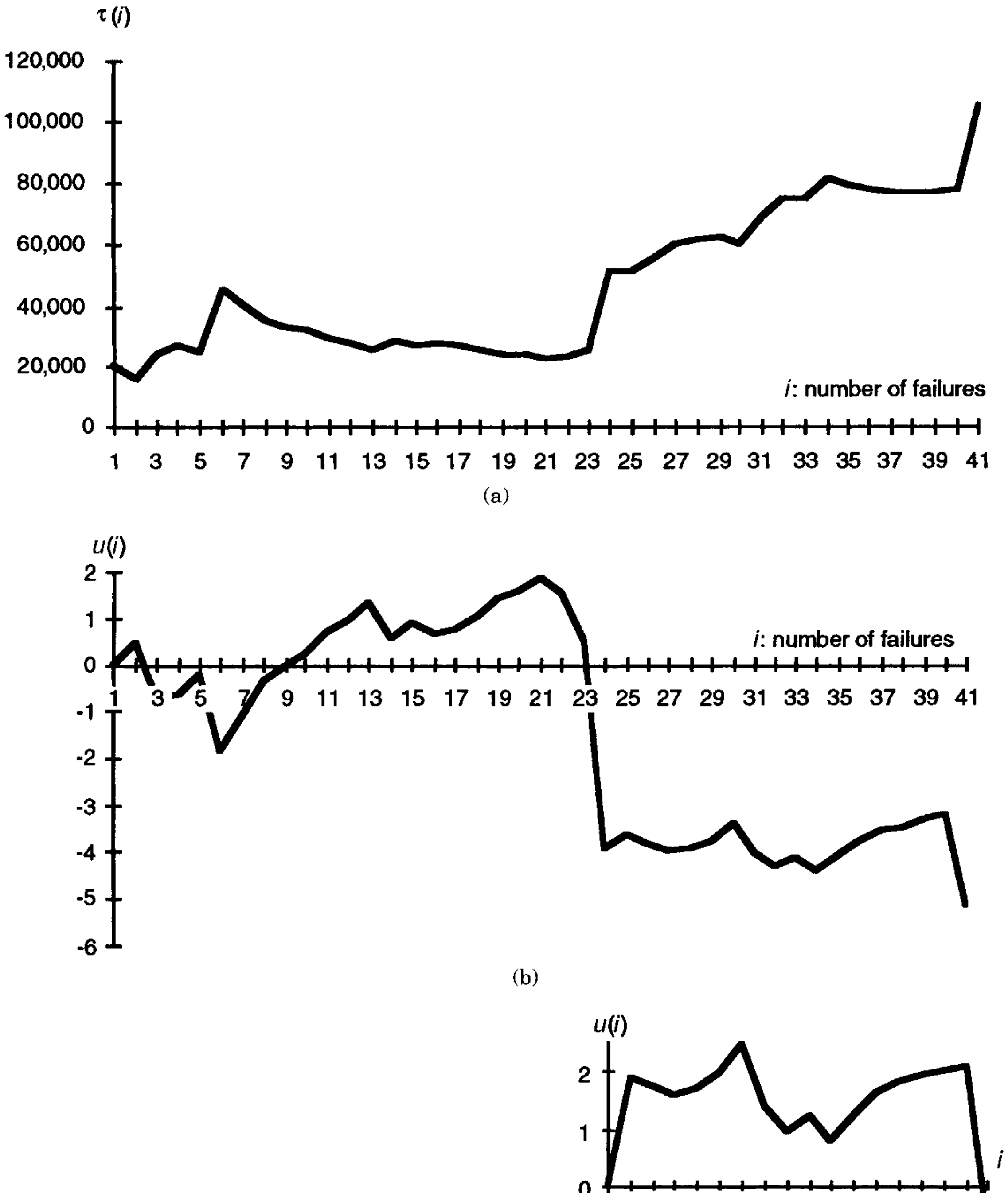


**Figure 10.12** Trend tests for S27 considering the times to failure. (a) Arithmetical mean of the times to failure. (b) Laplace factor of the times to failure. (c) Laplace factor from the 24th failure.

decrease is observed. It can also be noted that the trend change points do not vary from Fig. 10.6 to Fig. 10.13.

- From $k$ = 70, the trend is reversed. This failure behavior is directly related to the number of installed systems over the periods considered, during which about 12 systems were installed and the number

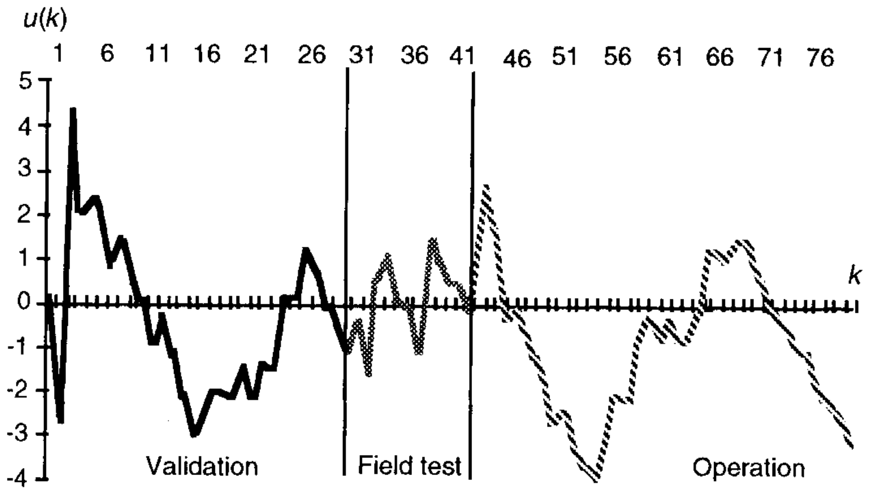


**Figure 10.13** Laplace factor of SS1 for each phase.

of failures reported by the users increased. By time unit 70, a new generation of systems had been released and no additional former system had been installed, which corresponds to the period of reliability growth from time unit 70.

Applying the reliability growth models blindly to this data set would have produced no significant results. However, taking into account the increasing number of installed systems and using the trend analysis results led to trustworthy predictions from reliability growth models. These results [Kano91a] are in agreement with those observed later.

### 10.4.4 Software of system SS2

The subadditivity factor, $\mathcal{A}_H(k)$, for this system is given in Fig. 10.14. SS2 displayed a reliability decrease during validation; reliability growth took place during operational life only. This is confirmed by Fig. 10.15, where the subadditivity factor for operational life is applied to the data collected during operation only. It can also be seen that some reliability fluctuations took place starting from unit time 15; this fluctuation is due to the introduction of new users. Clearly, no reliability growth model can be applied during validation. Nonetheless, an S-shaped model can be applied to the whole data set. Also, any reliability growth model with a decreasing failure rate can be applied to the data collected during operation (from unit of time 9) [Kano93b].

### 10.4.5 SAV

We consider the data collected during 70 units of time including the end of validation and operational life. Only a few failures were discovered during operation (which started at unit time 28), and even these

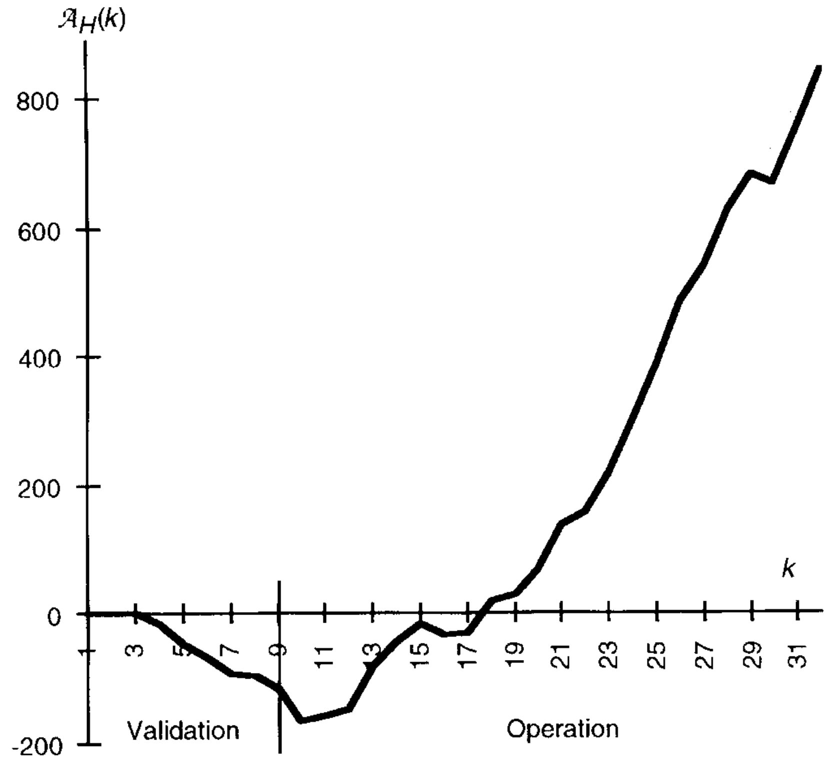


**Figure 10.14** Subadditivity factor for SS2 considering the whole data set.

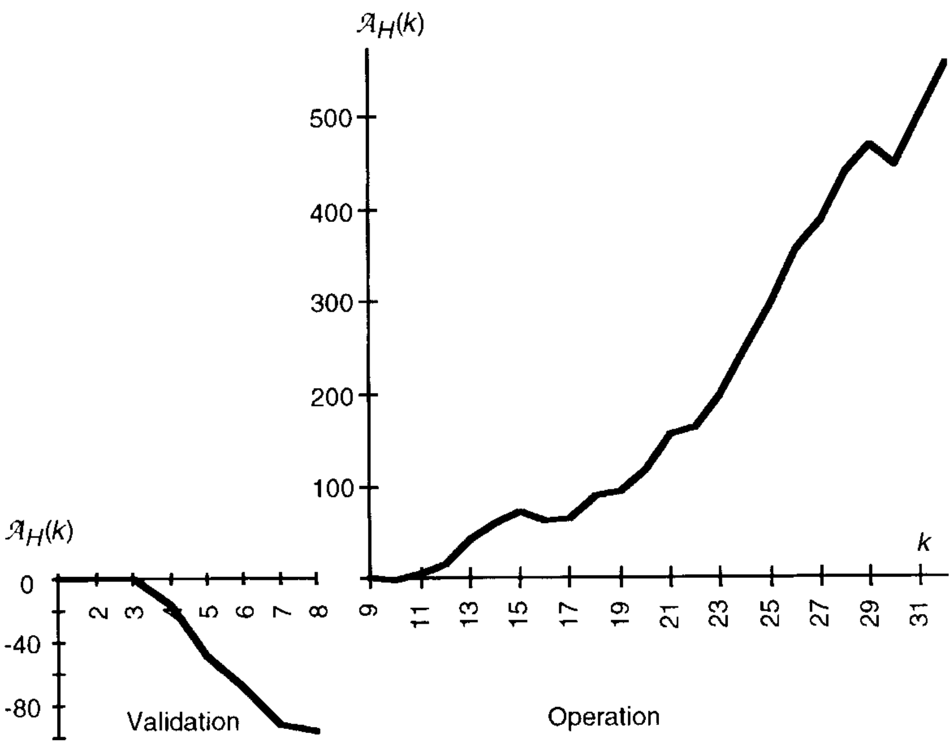


**Figure 10.15** Subadditivity factor for SS2 considering each phase separately.

failures were detected by the software manufacturer during introduction of new functionalities (specification changes). The Laplace factor for this system is given in Fig. 10.16. The reliability decrease around the 24th unit of time is due mainly to the introduction of new versions corresponding to changes in the specifications. It can be seen that a significant reliability improvement took place during operation (when considering the whole data set).

## 10.5 Extension to Static Analysis

### 10.5.1 Static analysis conduct

It is well known that software fault-fixing in the earlier development phases is much less expensive than fault-fixing at a later stage of development and during operation [Boeh81]. It is also well known that static analysis (carried out either by means of walk-through or code review or inspections) substantially reduces the corrective maintenance. Thus it is important to provide the software developer with some statistical criteria to guide the decision to drop from one phase to the following phase of the inspection process or testing.

The figures published in [Faga76, Ross87, Bush90, and Saye90] show that, for systems undergoing thorough walk-throughs or inspections, the majority of faults are detected before software execution (75 to 95 percent of the faults being found before software testing). For these systems, the analysis of data collected during testing can be advantageously preceded by analysis performed on data related to the trouble reports recorded during walk-throughs or inspections.

Generally, walk-throughs and code review vary greatly in terms of regularity and thoroughness, whereas inspections have well-established

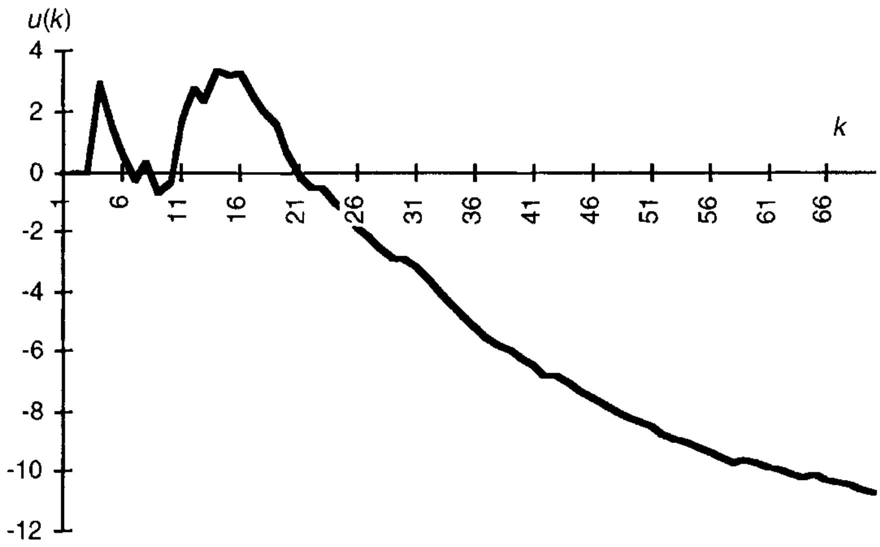


**Figure 10.16** Laplace factor for the observed failures of SAV.

and rigorous procedures [Faga76]. During inspections, exit from one operation to the following one is based on criteria that are checkpoints in the development process through which every programming project must pass. The sets of exit criteria are defined for each project and should be as objective as possible so as to be repeatable [Faga86]. For walk-throughs, such criteria do not exist and exit from the different operations is left to responsible judgment.

Faults detected during static analysis of the code and during testing in the absence of static analysis are faults located in the code; using the same approach in both phases seems natural. For design-level inspections, even though the nature of the detected faults may differ from those detected later, data analysis may be conducted as the analysis of failure reports recorded during testing and operation. Inspection progress can thus be monitored using trend analysis results (in addition to the exit criteria already used for exit between operations)* as for testing. However, due to the differing nature of the faults detected during the various inspection levels (or phases), each level has to be monitored separately to handle data of the same nature (note that this is common practice when processing data from different test phases).

As far as we know, statistical processing of information that can be derived from troubles reported during inspection or walk-through phases is seldom used to guide the management of these phases; the work published in [Grad86] constitutes an exception. Generally, these data are either processed together with data from software testing by application of reliability growth models, as carried out in [Ross89], or processed by a tracking model based on the results related to previous similar projects as in [Kan91] in order to approximate the final quality index of the software and not to guide its management as proposed here.

Theoretical aspects and practical results presented in Secs. 10.2 and 10.3 can thus be adapted to faults detected by static analysis. Data in the form "number of faults detected per unit of time" is better suited to these phases, since the supplier is more interested in the process of removing faults than in evaluating intervals between two fault detections. The choice of unit of time duration is determined by the duration of the inspection phases and the number of faults detected. The unit of time may range from a few hours to one day when several faults are detected during such periods, or a few days when less faults are detected; and the unit of time can be changed from one phase to the

* It is worth noting that exit criteria for software inspection and monitoring criteria using trend tests are not contradictory. The first one has to be applied to internal operations within an inspection level, whereas the second constitutes a criterion based on the number of troubles detected by this activity and helps make the decision to exit the inspection level considered.

next. However, since the software is not executed during inspection, talking about reliability growth or decrease is meaningless. We are interested in the evolution of the number of troubles reported versus time and not in the evolution of the interfailure times or number of failures. Nevertheless we will abusively use "reliability growth" (respectively, decrease) to characterize situations where the number of trouble reports is decreasing in time (respectively, increasing).

### 10.5.2 Application

Consider again SAV, the software of the avionic application studied in the previous section. For this system, troubles detected during specification and code review have been recorded in trouble reports analogous to failure reports. Since the system specifications kept changing during the life cycle, trouble reports were drawn up even when the system was in operation (while adding new parts to the software). For this system, more than 50 percent of faults were detected by specification and code review.

The Laplace factor corresponding to the trouble reports is given in Fig. 10.17 which, as in the case of software execution, reveals an almost steady reliability growth from unit of time 16. Reliability fluctuation at the beginning is due to the review of new specifications and the corresponding software code. From unit of time 25, it is interesting to note that the very local fluctuations of the Laplace factor corresponding to trouble reports (indicated by the arrows in Fig. 10.18) are also followed by local fluctuations of the Laplace factor corresponding to software failures. The time lag corresponds to the time interval between the review of the new part of the software (either specification or code) and the execution (testing) of this part of the software.

## 10.6 Summary

In this chapter, we have characterized reliability growth using the notion of the subadditive property. Then a graphical interpretation of this property was derived and we have shown the equivalence between this property and the Laplace factor, thus allowing for the Laplace test to be extended to local trend-change identification.

We have shown (1) that trend analyses constitute a major tool during the software development process, from static analysis to system integration and (2) how the results can guide the project manager to control the progress of the development activities and even to make the decision to reexamine the software for specific situations. Extension of trend analyses to trouble reports from static analysis is all the more useful as the majority of faults are detected by design and code inspec-

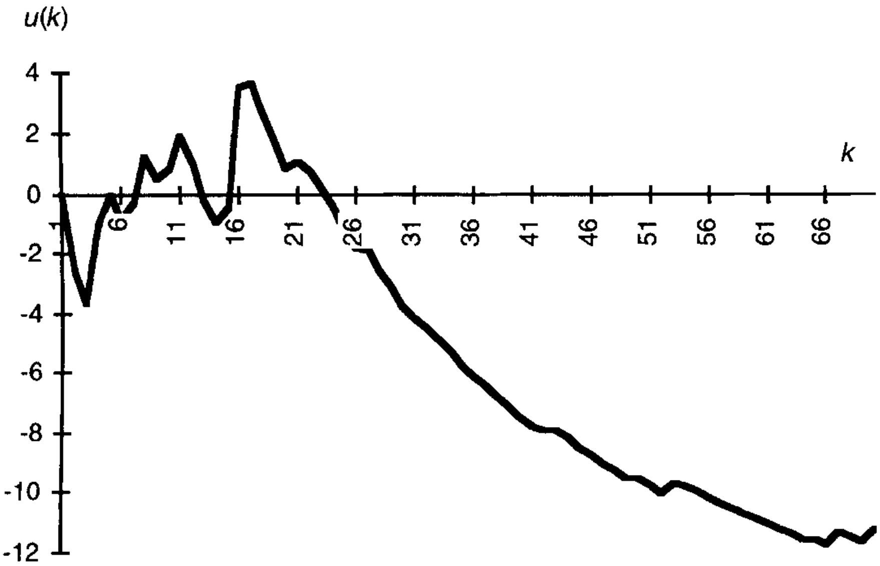


**Figure 10.17** Laplace factor for the trouble reports related to SAV review.

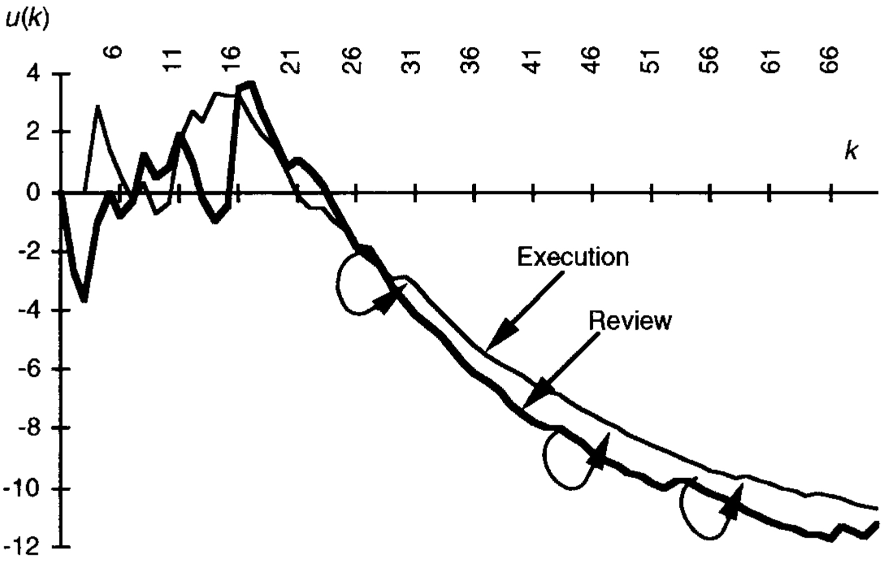


**Figure 10.18** Laplace factor for the trouble reports and for the observed failures related to SAV.

tion; monitoring the inspection activities is thus of prime importance in these situations.

Trend analyses are also helpful when reliability evaluation is needed. They allow periods of reliability growth and reliability decrease to be identified in order to apply reliability growth models to data exhibiting trends in accordance with their modeling assumptions. Trend analysis and reliability growth models are part of a global method for software reliability analysis and evaluation which is presented in [Kano88, Kano93b] and which has been successfully applied to data collected on real systems [Kano87, Kano91a].

## Problems

**10.1** The table in Fig. 10.19 gives the successive interfailure times observed during the validation test of an application software (read from left to right). Which trend tests can you apply to this data set? Apply at least one of them. Does this curve or the results of the trend test application reveal a possible abnormal behavior?

**10.2** Assuming that Fig. 10.19 represents the observed failure intensity (number of failures per week), answer the same questions as in Prob. 10.1.

**10.3** The Laplace factor calculated from the failure intensity data collected during testing is plotted in Fig. 10.20. Identify the various periods of reliability growth/decrease. Can you think of some reasons for this reliability decrease? Comment on this and recommend one or more reliability growth models which could be applied according to the trend.

**10.4** After three months of software testing without specification changes, the observed failure intensity is given in Fig. 10.21. Identify the period(s) of reliability decrease. What could be the reasons for this decrease? (Give some reasons that seem acceptable from a tester's point of view, and others that could help improve the validation procedures.)

**10.5** Repeat Prob. 10.4 assuming two new versions of the software have been introduced due to specification changes. Locate approximately the times of the introduction of the new versions. Comment on this.

**10.6** In the text (Sec. 10.2.2), we have shown how Eq. (10.3) implies Eq. (10.5) in the case where $K$ is even. Show this implication when $K$ is odd.

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 12 | 10 | 16 | 13 | 6 | 7 | 6 | 5 | 7 | 9 | 8 | 9 | 7 | 5 | 6 |
| 3 | 5 | 3 | 7 | 9 | 8 | 10 | 9 | 12 | 10 | 14 | 12 | 15 | 13 | |

**Figure 10.19**

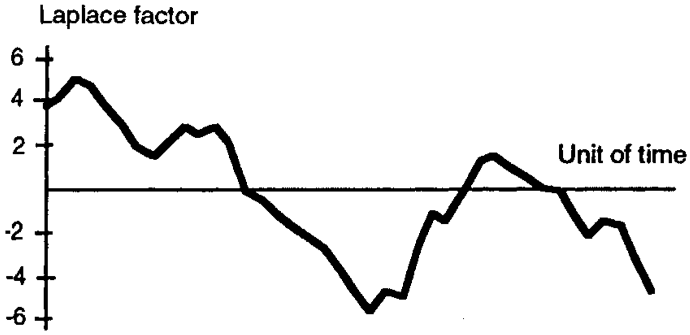


**Figure 10.20**

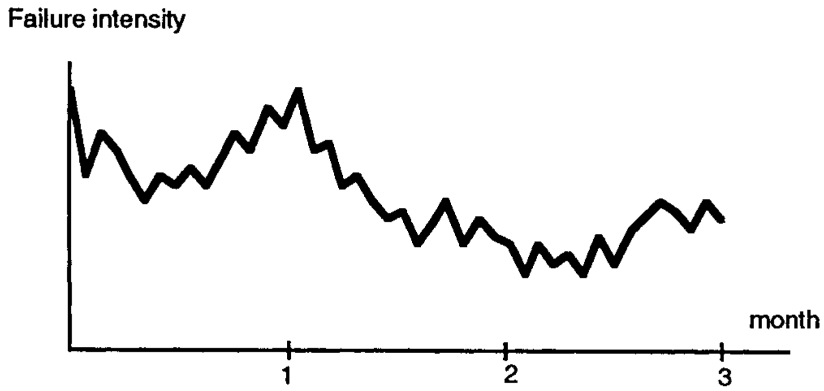


**Figure 10.21**

**10.7** The observed failure intensity during the last six months of software testing is given in Fig. 10.22*a* (number of failures per week). It is assumed that there are no quantified reliability objectives. Our aim in this exercise is to use qualitative and informal criteria from trend test results and the observed failure intensity to guide the development process. The informal criteria for software delivery is the following: "the software has reached a stable reliability behavior with a few failures per week—more precisely, less than 10 failures per month."

*a.* Plot the failure intensity and the corresponding Laplace factor. Delivery of the software is planned for the end of the year. Does this aim seem reachable?

*b.* The failure intensity observed during the following three months (months 7 to 9) is given in Fig. 10.22*b* (number of failures per week). Comment about the trend. Do you think that it is still reasonable to plan delivery for the end of the year?

*c.* The failure intensity observed during the following three months (months 10 to 12) is given in Fig. 10.22*c* (number of failures per

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 30 | 31 | 25 | 30 | 28 | 27 | 32 | 25 | 30 | 28 |
| 29 | 25 | 27 | 28 | 26 | 25 | 23 | 24 | 19 | 20 |
| 17 | 15 | 16 | 12 | 15 | 17 | | | | |

**Figure 10.22(*a*)**

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 12 | 15 | 13 | 12 | 9 | 12 | 10 | 11 | 9 |
| 10 | 14 | 15 | | | | | | |

**Figure 10.22(*b*)**

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 17 | 16 | 14 | 17 | 15 | 10 | 8 | 9 | 6 | 7 |
| 5 | 3 | 4 | | | | | | | |

**Figure 10.22(*c*)**

week). At the end of the 10th month do you think that it is still reasonable to plan delivery for the end of the year? Plot the overall failure intensity and the Laplace factor. What could be the reasons for the failure intensity increase at the beginning of the last quarter?

**10.8** The average failure intensity observed during four weeks of testing is (assuming seven working days):

- 5 failures per day the first week
- 3 failures per day the second week
- 2 failures per day the third week
- 0.2 failures per day the fourth week

Plot the failure intensity and the cumulative number of failures. Show that the latter is subadditive over the considered period. Plot the Laplace factor.

**10.9** After 12 weeks of testing, the observed failure intensity can be approximated as follows:

- For $0 \le t < 3$, $h(t) = 2 + 3t$
- For $3 \le t < 7$, $h(t) = 17 - 2t$
- For $7 \le t < 9$, $h(t) = 10 - t$
- For $9 \le t \le 12$, $h(t) = 1$

Plot the failure intensity and the cumulative number of failures. What about the subadditive property? Locate approximately the region of trend change. Plot the Laplace factor.

**10.10** The failure intensity (i.e., the number of failures per week) observed during the validation of a software system is given in Fig. 10.23 (read from left to right). Plot the failure intensity and the Laplace factor. What conclusions can be drawn for the trend? Partition the data according to the trend and plot the Laplace factor for the subset displaying reliability growth.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 15 | 17 | 16 | 19 | 20 | 17 | 20 | 19 | 21 | 19 |
| 25 | 28 | 26 | 24 | 26 | 25 | 23 | 24 | 19 | 20 |
| 17 | 15 | 16 | 12 | 15 | 17 | 12 | 15 | 13 | 12 |
| 9 | 12 | 10 | 11 | 9 | 13 | 10 | 9 | 10 | 8 |
| 7 | 9 | 5 | 6 | | | | | | |

**Figure 10.23**