Proc. Natl. Sci. Counc. ROC(A)
Vol. 23, No. 5, 1999. pp. 555-566

(Invited Review Paper)

Degradation Analysis and Related Topics: Some Thoughts and a Review

MIN-TE CHAO

Institute of Statistical Science
Academia Sinica
Taipei, Taiwan, R.O.C.

(Received December 1, 1998; Accepted April 27, 1999)

ABSTRACT

Degradation is a phenomenon where certain measurements of quality characteristics deteriorate over time. When an item subject to life testing is too well made to fail, we often turn our attention to degradation measurements and hope to obtain life time information from this type of data. The first part of this paper presents a general discussion of the degradation phenomenon. The second part reviews some scattered works related to the analysis. Topics involve shelf life studies, growth curve, S-shaped curves, experimental design, stochastic process modeling, accelerated life testing and step-stress models.

Key Words: accelerated life testing, degradation, design of experiments, growth curve, S-shaped curves. step-stress models

I. Introduction

I started to do a relatively intensive literature search on topics concerning degradation analysis when we were awarded a 3-year project by the National Science Council in 1996. At that time, there was no intention to perform a thorough review except for fear of missing something important. What is degradation? As time goes on, things become worse. Degradation is just a technical term used to describe this phenomenon. Since the degradation phenomenon takes place naturally, there is a vast amount of work in the literature and our task is necessarily limited by this author's inability and his selection bias. A major problem we encountered is that most of these works are closely related to certain physical processes. Subject matter knowledge, at least at the superficial level, is needed when in depth understanding is called for. Nevertheless, the statistical behavior of degradation data and its related applications have been well studied.

Degradation analysis involves the degradation data. Four sets of such data are reported in Section II just to provide a flavor of the subject matter. We examine and explain the reasons why such data are collected. The true objective is to study the life time distribution of certain subjects under inspection, but we are driven by the need to extrapolate into the unknown because the subjects are too well made and no or little failure is observed during the testing period. The idea is simple indeed: if it takes too long to observe a true failure, we simply try to observe how bad the item under inspection can become. In doing so, we hope that perhaps some conclusion can be made which is still useful in real world terms.

One of the subjects most carefully studied using degradation data is determining the shelf life of drugs. There are good reasons: on the one hand, it is an official requirement. Stating on the labels the expiration dates for all marketable drugs is required by health authorities in most developed countries. On the other hand, there is a powerful profit-driven pharmaceutical industry behind the scenes. In spite of the amount of work actually done, the methodology used to determine the shelf lives is basically mature, however. Some recent works are reported in Section III.

In Section IV, we examine the development of another related topic, the study of growth curves. As in the shelf life problem, the basic underlying assumption of growth curve studies is that observations are jointly normal. This implies that, among other things, the basic methodology is solid. Strictly speaking, nothing is surprising.

The classical S-shaped curves, also called sigmoids, are discussed in Section V. This type of curve fitting is traditionally used to model different growth rates of a subject or a group of subjects under study when there is a natural limit of growth.

We then move to the problem of how degradation data are collected. For traditional data collection, the most effective way is through a well designed experiment where the criterion is to select observations at points maximizing certain information measurement. For degradation problems, since the objective is to make inference on the extreme tail part of the life data, a class of problems, called extrapolation design, come into the picture naturally (Section VI).

In Section VII, we examine the methods used to model the degradation process. There are basically two types of probability models that have been used: the compound Poisson process and the Brownian motion. The former is the basis of the cumulative damage model, and the latter allows the use of stochastic differential equations.

The original version of this work was done in 1996 when the project started. As time went on more recent papers were included. Efforts have not been made to search back for the purpose of giving credits for priorities; rather, the objective of this work is to provide readers with enough background material to start work. Most practical problems are solved using known methods. Since doing good data analysis is only technical in nature, true scientific contributions for this type of problems are rare. I have, however, tried to cite works by local researchers as much as possible.

II. The Degradation Phenomenon and Why We Study It

1. Reliability Analysis

Most things have a life span, defined in one way or another. These lives, when measured, present us with data sets we can use to explore further for scientific or other purposes. It is natural to study the life time distribution of a subject through a set of measured data. One of the earlier topics in this respect, which is still very much alive in research activities, is the theory of reliability. A generic definition of reliability is:

> Reliability is the probability of a product or a system performing its intended function without failure for a specified period of time under specific conditions.

For reliability analysis, the basic subject matter to study is the probability distribution of the life times of the subject under study. For this purpose, the standard method is to take a set of observed life times T_1, T_2, ..., T_n, censored sometimes, though basically we assume $T_i \sim F(\cdot;\theta)$. From the likelihood function constructed from this sample, we can make an inference with respect to the unknown parameter θ.

When the form of $F(\cdot;\theta)$ is known and the complete distribution of F is determined by a finite dimensional parameter θ, then we have a classical parametric model; if F is completely unknown except for some qualitative descriptions such as continuity or smoothness, then we have a non-parametric model; finally, if F is unknown but the parameter θ has some structure to explore, then we have a semi-parametric model.

When the parameter θ exhibits some structure, we will naturally embed our inference problem into traditional, and time-tested as well, models for statistical analysis. These include techniques such as experimental design, regression, logistic regression, accelerated life testing, etc. These methods incorporate various situations that one may encounter in practice. There is no need, however, to restrict the inference to the classical frequentists' parametric setup. We can, if the situation requires, use the Bayesian method or even the empirical Bayes techniques.

What we have described is the general setup for the reliability inference. The real problem, particularly when applied on the shop floor or in a reliability testing center, is of course much more complicated. We have at best provided a very rough description for this important topic, but it cannot be too far away from the general picture. There is extensive literature on reliability analysis, and research is still very active. On the other hand, several excellent books exist on this topic, and industrial standards have been established, e.g., by the Department of Defense, U.S.A. (1991): MIL-HDBK-217F.

There is a common requirement for reliability analysis: we have to make observations on the true life times T_1, T_2, ..., T_n. A certain percentage of the observed T's may be be censored, but all methods fail when this percentage becomes too high.

2. Data Collection Problems

In the last section, it was mentioned that one has to take true lifetime observations to make a sensible inference on the properties of the life distribution of the subject matter under study. In reality, however, it is difficult to actually observe the true failure of a well-made item even under accelerated environment. With rapid advances in technology, products nowadays are just too well-made to fail within a reasonable amount of time. For example, the average life time of a light-emitting diode (LED) is around 100,000 hours in theory. But this is about 11 years, and if a firm has to wait for that long, or even half that long, to obtain information about its mean life before failure (MLBF, one of the most important pieces of reliability information), then it will lose purchase orders for these items to other competitors. It is entirely possible that some other devices will be invented in the mean time which will replace the functions of LED, and in 11 years the LED may no longer be needed in the market. On the other hand, reliability information, and the ability to supply it, is an important means of persuading a buyer. We must develop some other methods to cope with this problem when the items are slow to fail.

One way is to observe, and calculate, to what extent the items under study have deteriorated. We may take measurements on one or several characteristics over a period of time and observe how bad they become. From the data set thus collected, the life time distribution, or some characteristics of the life distribution, under study can be estimated. For example, by measuring luminosity of fluorescent lamps, Tseng *et al*. (1995) estimated the time when the amount of degradation in the luminous flux fell below 60% of its original value.

The degradation phenomenon takes place naturally: iron can rust, copper may oxidize, bacteria make food rotten. In general, the impact of the environment, be it natural or artificial, takes its toll gradually. The item under inspection may still be useable and the amount of deterioration, properly measured, can provide us with useful information. In some cases, it can be translated into regulations: drugs and food and other biomasses have proper shelf lives; we expect the life expectancy of an under sea cable to be 25 years, etc.

Some degradation phenomenon can in theory be calculated. One example is the half-life of a certain radioactive material. Another example is the amount of time that a satellite stays in orbit. For many others degradation phenomenon the underlying theory is limited, and the only sensible way to make an inference is to actually take samples and establish a model on which an inference can be made. Still many others degradation phenomenon lie between these two extremes: there is some theoretical background, but further verifications are needed. Degradation data, however, basically have the following characteristics: (1) there is a subscript t, indicating the time sequence in which the data were obtained; (2) an important part of the data changes monotonically with time. This is, among other things, the basic essence of degradation is that it changes with time along a definite direction. Milk spoiles with time, iron cannot become less rusty, the cracks in a reinforced concrete structure can only grow, the luminosity of fluorescent lamps may fluctuate due to unstable voltages but basically goes down with time.

Table 1. Dielectric Breakdown Strength

Temperature\Weeks	1	2	3	4	5	6	7	8
180	15	14	13.5	15	18.5	12.5	13	13
	17	16	17.5	15	17	13	13.5	12.5
	15.5	13	17.5	15.5	15.3	16	16.5	16.5
	16.5	13.5	13.5	16	16	12	13.5	16
225	15.5	13	12.5	13	13	11	11.5	11
	15	13.5	12.5	10.5	14	9.5	10.5	11.5
	16	12.5	15	13.5	12.5	11	13.5	10.5
	14.5	12.5	13	14	11	11	12	10
250	15	12.5	12	12.5	12	11	7	7.2667
	14.5	12	13	12	12	10	6.9	7.5
	12.5	11.5	12	11.5	11.5	10.5	8.8	6.7
	11	12	13.5	11.5	12	10.5	7.9	7.6
275	14	13	10	6.5	6	2.7	1.2	1.5
	13	11.5	11.5	5.5	6	2.7	1.5	1
	14	13	11	6	5	2.5	1	1.2
	11.5	12.5	9.5	6	5.5	2.4	1.5	1.2

Source: Nelson (1981)

3. Basic Data Form

A. Example 1

We want to estimate the dielectric breakdown strength (DS) of certain insulation specimens. For this purpose, a sample of 128 such specimens is obtained under 4 different temperature settings: 180 °C, 225 °C, 250 °C, and 275 °C. Eight aging times are used: 1, 2, 4, 8, 16, 32, 45, and 64 weeks. The data in Table 1 represent the measured dielectric strengths. We are interested in knowing the amount of time needed for a typical specimen to degrade to 2 kVolts when the temperature is set at 150 °C (Nelson, 1981).

The amount of degradation in this data set is measured on the different specimen. Therefore the observations are independent and there is no need to do time series analysis. However, when we plot the measurements against time, it is clear that for any fixed temperature, the trend with time is obvious in scatter plots.

Let

$$X_{itk} \ (i=1,\ 2,\ 3,\ 4;\ t=1,\ 2,\ \ldots,\ 64;\ k=1,\ 2,\ 3,\ 4)$$

be the DS values measured at temperature T_i at the t-th week. Figure 1 shows the respective scatter plots and the related regression lines.

B. Example 2

Table 2 shows a typical set of data obtained from a stability study of certain drugs for a new drug application. There are 24 batches, and measurements of the contents of their major components are taken at 0, 12, 24, and 36 months. The main purpose of analyzing this type of data is to determine the shelf lives of these drugs. In general, the potency of various components decreases with time. This data set comes from Chow and Shao (1991).

Talbe 2. Data from a Stability Study

batch	age				batch	age			
	0	12	24	36		0	12	24	36
1	105	104	101	98	13	105	104	99	95
2	106	102	99	96	14	104	103	97	94
3	103	101	98	95	15	105	103	98	96
4	105	101	99	95	16	103	101	99	96
5	104	102	100	96	17	104	102	101	98
6	102	100	100	97	18	106	104	102	97
7	104	103	101	97	19	105	103	100	99
8	105	104	101	100	20	103	101	99	95
9	103	101	99	99	21	101	101	97	90
10	103	102	97	96	22	102	100	99	96
11	101	98	93	91	23	103	101	99	94
12	105	102	100	98	24	105	104	100	97

Source: Chow and Shao (1991)

C. Example 3

A copper bearing intra-uterine contraceptive device (IUCD) consists essentially of a length of copper wire coiled around a plastic former. When placed *in uterus*, the copper is released, which increases the efficacy of the device. The contraceptive action is dependent upon the surface area of the copper, which may be regarded as the main factor controlling the rate of copper release. The data set consisting of 811 IUCD's was collected through intra-uterine device research networks and family planning centers in the Manchester area, U.K. The copper content was determined by using atomic absorption spectroscopy.

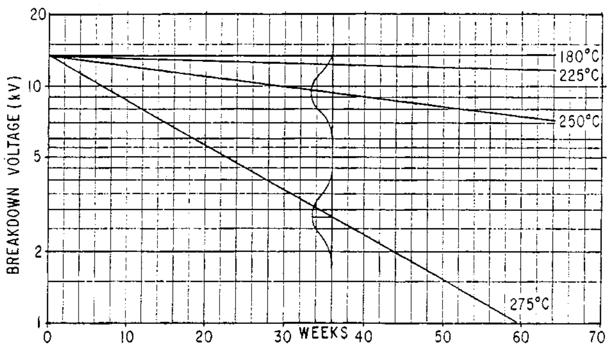

Fig. 1. Aging of dieletric strength. [Adapted from Nelson (1981)]

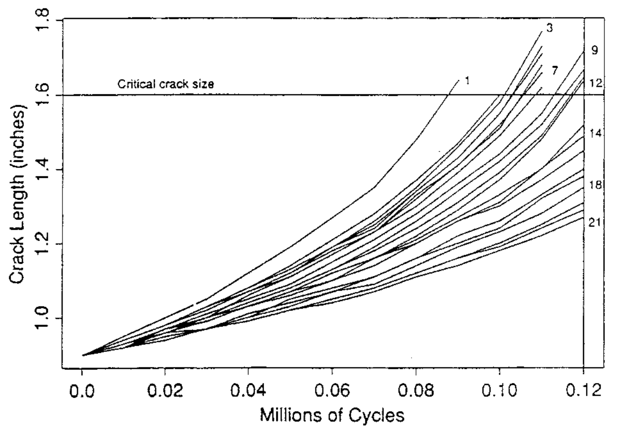

Fig. 3. Growth of cracks. [Adapted from Bogdanoff and Kozin (1985)]

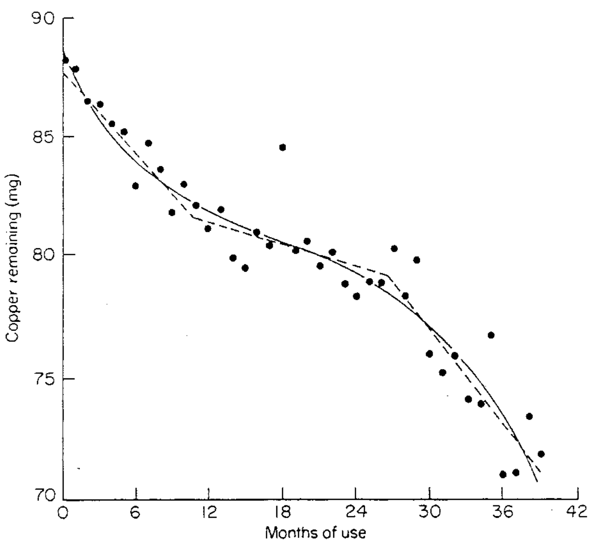

Fig. 2. Degradation of copper content. [Adapted from Faragher *et al.* (1985)]

The data and two types of fit are plotted in Fig. 2 (Faragher *et al.*, 1985), and it is clear that the copper content decreases with time.

D. Example 4

Figure 3, taken from a book by Bogdanoff and Kozin (1985), indicates the growth of crack length caused by stress. Here, we use "million cycles" in the abscissa, instead of time, to indicate the stress levels applied.

This data set has been thoroughly analyzed by Lu and Meeker (1993).

4. Some Conditions on the Form of the Data

Summarizing the examples given in the previous section, we may use the following form to describe the type of degradation data:

$$\{X_{at},\ a\in A,\ t\in T\},$$

where A denotes a set of labels and T is either $[0,\infty)$ or $\{t_0, t_1, t_2, ...\}$, implying time. For real observations made, the set T is discrete; but theoretically we may consider T as being continuous.

The label set A may have some structure which indicates the conditions under which these observations were made. In general, if $a_i\in A$ are all distinct, we may assume that $X_{a_i t}$, i=1, 2, ... are mutually independent. But for the same a and different $t_1<t_2<...,<t_k$, there may be some dependence among the observations modeled in probability terms.

The theme "degradation" implies that when $t_1<t_2$, we should expect one of the following:

Condition D. $E[X_{at_1}]\geq E[X_{at_2}]$
We believe that Condition D can reasonably be imposed on all degradation data. Unfortunately, this condition has not been sufficiently utilized in the literature. A more stringent condition is the following.

Condition SD. For all $t_1<t_2\in T$, $X_{at_1}\geq X_{at_2}$
Because of measurement errors and perhaps other noise factors (such as an unstable voltage), strict degradation conditions are hard to satisfy. In practice, when degradation data are clearly defined and the analysis is basically sound, Condition D usually holds.

5. Data Analysis Method in General

If we can write down the form of the likelihood function (or some variation of it) of the observed data, be it parametric, non-parametric, or semi-parametric, basically for this set of data the method of analysis is more or less determined. What is left is either technical in nature (such as the Newton-Raphson iteration method) or doctrine of schools, for example, Bayesian or frequentist. Therefore, the basic problem is to choose a proper probability model to describe the behavior of the observed data.

Works along this direction, except for the use of Condition D or SD, are abundant in the literature. Their basic strategy is to look at the data set first. For example, when the label set A is a singleton set, then all we have is the data $\{X_t,\ t\in T\}$, a time series. If T contains 25 or more points, we will naturally try to use traditional time series models, such as the autoregressive integrated moving average (ARIMA) model. If A has some linear structure and T has less than 10 points, then we may employ growth curve analysis, which is also well developed. If we can find some background knowledge, say that X_t satisfies a certain set of differential equations, then we may try to see if the use of stochastic differential equations can be a better approach.

6. Danger of Extrapolation

We mentioned in previous sections that the reason for doing degradation analysis is that we need to make use of the degradation phenomenon to model an unobservable time of failure. This means that we need to do extrapolation. All methods of extrapolation are dangerous, however.

The simplest method of extrapolation is to fit the data with a straight line and extend the line further into the future. Straight lines can indeed be extended, but any small error is magnified as the abscissa increases. The further we extrapolate, the more mistakes we may expect. This is a basic problem with all extrapolation (and forecast) methods. One may replace the straight line with some other curves, but the problem remains. It is fair to say that academic rigor is inversely related to the degree of extrapolation.

We may need to understand, or make a judgement on, whether an under sea cable has a useful life of 25 years, or whether an LED can last for 11 years. But the accuracy of these judgements depends on the appropriateness of the fitted statistical model. When the model is wrong, the forecast may still be valid. A simple model may out-perform the more sophisticated ones in forecasting. Nevertheless, more sophisticated models models continue to appear and disappear. Long range forecasting and extrapolation cannot be accurate. Depending on their final use, the danger due to inaccuracies of extrapolation can be serious, not so serious or sometimes (fortunately) irrelevant. In many cases, these may be the only educated guesses. It is difficult to make a distinction between true scientific work and pseudo scientific work; nevertheless, the distinction is important. An example is provided at the end of Section V which shows that even a well planned experiment can lead to an incorrect result if extrapolation is not carried out with care.

III. Shelf Lives of Drugs

One of the formally documented procedures in degradation analysis concerns the determination of the shelf lives of drugs. All drugs are specifically labeled "to be used before (a certain date)". How are these expiration dates determined?

Studies of this kind are called stability studies in pharmaceutical literature. A more common term is "shelf life studies". For formal documentation, see

FDA, U.S.A. (1987), MHW, Japan (1991), and ICH, Switzerland (1993). A monograph by Chow and Liu (1995) gives the technical details of shelf life studies and illustrates them using well-prepared examples. It is perhaps the best text on this topic so far. In what follows, we will only discuss shelf life studies related to statistical analysis.

1. Official Definition of Shelf Life

Suppose there are k batches of drugs to be tested. Let Y_{ij} be the result of the j-th assay of the i-th batch. The following model is used:

$$Y_{ij}=x'_{ij}\beta_i+\epsilon_i,\ i=1,\ \dots,\ k;\ j=1,\ \dots,\ n$$

where x_{ij} is of dimension $p\times 1$ and denotes the appropriate regressor, and ϵ_i denotes the error term.

The simplest form of x is of the form $x_{ij}=(1,t_j)$, where $t_1<t_2<\dots<t_n$ denote the time points when these assays are performed.

The first step after data collection and cleaning is to test whether there is between batch variation. That is, the null hypothesis

$$H_0{:}\beta_1=\dots=\beta_k$$

is tested. The size of the type I error used is $\alpha=0.25$, which is employed based on a suggestion made by Bancroft (1964). If H_0 is accepted then we may pool the k sets of data together to obtain a shelf life estimate. If H_0 is rejected, then we will find a shelf life estimate for each batch, and the smallest among k shelf lives will be used.

The method used to estimate the shelf life of any batch (or a combined batch) is as follows. First, we take $E\beta=b$, so $EY=x'b$. In general x is a function of t, so we write $x=x(t)$. The main concern of the FDA is whether the contents of the drug components fall below the amount as claimed on the label. If η is the content of a component described on the label, then the "true shelf life" is then the time needed for the value of η to reach a certain lower level. Typically, we set the threshold at $\eta'=0.9\eta$; in doing so, it is implicitly assumed that $x(t)'b$ is a decreasing function of t. Therefore,

$$SL_{True}=\inf\{t{:}x(t)'b=\eta'\}.$$

From the frequentists' point of view, SL_{True} is an unknown parameter, so it is possible to define its confidence interval. The one-sided lower confidence level is often used, for conservative reasons. The level is usually set at the standard 95% level.

2. Fixed Effect Model

The estimation of shelf life, as described in the previous section, can be modified. When the statistical methodology becomes maturer, it is sensible to believe that for the same data, better estimates can be obtained than before.

The fundamental question is: what is SL_{True}? When we follow the procedure under which the shelf life is defined, it seems proper to say that a clear definition is possible only when no batch to batch variation is discovered. When the problem is examined from the consumers' point of view, any single purchased drug should have a 95% chance that the amount of its components is more than $\eta'=0.9\eta$, as stated on the label. Here, "any drug" means that a drug chosen at random, so from the users' point of view, the essence of the shelf life problem is the prediction of a future observation rather than estimation of some unknown parameters. There has been basically no research done on shelf life (of drugs) with prediction in mind. The only related work, in terms of prediction and life length estimation, is a paper by Carey and Koenig (1991). Their topic is estimation of the life of some under sea cables, which we will describe later.

When we assume the viewpoint of a consumer, the shelf life problem is then predictive in nature. Since it is the value of a typical future observation that is our concern, we must include variation due to the observation itself in our prediction interval. A natural consequence is that the modified estimate of SL will be shorter. Here, the emphasis is on random selection of a single drug. This is different from selection of a batch of drugs at random from many batches at hand, however. The latter, termed "the random effect" model, is well represented in the literature. We will discuss it in the next section.

On the other hand, if H_0 is rejected, then using the shortest shelf life only assures that SL_{True} is one of the true average shelf lives. Since this method is conservative, the estimated SL must at least be a safe number to use. When this approach is employed, it is implicitly assumed that these few batches must be representative batches. Strictly speaking, in this situation we are not very sure what the true definition of SL_{True} is and yet efforts have been made to find its respective confidence intervals. The work is still under the framework of frequentists, but its explanation is not entirely clear.

When the number of batches k is large, the shortest length method may be too conservative. Efforts have been made to solve this problem (Shao and Chow, 1994). Let SL_i, be the shelf life of the i-th batch. Then when $k\to\infty$,

$$\min_{1 \le i \le k} SL_i \to 0.$$

Hence, when k is large, it is unreasonable to simply select the shortest shelf life among all the batches. For most new drugs the value of k is about 3 or 5. But for a so-called marketing stability study, a larger value of k is used and the sampling intervals will be closer (Chow and Shao, 1991).

The simplest model for the i-th batch is perhaps of the form

$$Y_{ij}=\alpha_i+\beta_i t_{ij}+\epsilon_{ij}.$$

So far, we have assumed that α_i, β_i are fixed but unknown parameters. All tests for

$$H_0:\beta_1=\ldots=\beta_k,$$

treat β's as unknown parameters as well. To test H_0, the method of analysis of covariance is employed. For some real case examples, see Ruberg and Stegeman (1991). This paper deals with the problem of how to pool data from various batches when the main concern is fixing the power of the test.

3. Random Effect Models

We may treat the parameter as being of 2 or more dimensions. For random effect models, the coefficients α, β may be modeled as random variables. Thus, we will consider parameters of the form (α,β). Here, **(i)** α_i and β_i can both be unknown parameters; **(ii)** α_i can be an unknown parameter and β_i can be treated as random; or **(iii)** α_i and β_i can both be random.

Statisticall analysis in these cases, since they all assumpe normality, can be carried out without too much difficulty, at least for large sample cases. For example, Chen *et al*. (1995) treated β_i as missing data and used a unified approach to treat cases **(i)**-**(iii)** mentioned earlier using the same EM algorithm. In principle, this falls into the framework of Dempster *et al*. (1981) for the Bayesian/EM setup. The EM algorithm can be slow sometimes. In reality computing efficiency is not a main concern, however, so the general technique developed in this work is sufficient to solve almost all shelf life problems under the assumption of normality.

There are other scattered works. Murphy and Weisman (1990) considered case **(ii)**; as did by Chow and Shao (1991). Most works on shelf life have used the ordinary least squares method to estimate the parameters (α,β), but weighted least squares may also be employed. Kirkwood (1977) used the maximum likelihood directly, and in that paper, some earlier works on accelerated degradation tests are cited. We should also mention the work of Shao and Chow (1994). It summarizes some known exact results and suggests a method for actually estimating the shelf life in case **(ii)**. Wei (1998) suggested simple method for determining release limits for drug products. The method is still somewhat artificial, and this author feels that even under the assumption of a normal model, the problem of shelf life estimation is relatively narrow and mostly consists of piecemeal research topics. This is not a glamour topic to jump into, however, and young statisticians should look at it carefully before putting efforts into it.

IV. Growth Curves

Compared with research works related to the determination of shelf life, a more general framework is the growth curve. Growth curves stand alone as a self-sufficient topic, and there is a vast amount of literature dealing with this subject. For a succinct introduction, yet with considerable depth, see Ware (1983).

The basic data form of a growth curve is of the form $\{X_{at}, a\in A, t\in T\}$. The earliest example of a growth curve study examined the ramus heights of 20 boys measured at age 8, 8.5, 9 and 9.5, respectively (Elston and Grizzde, 1962). Independence among the boys was assumed, of course, but the measurements taken at different times for each boy were dependent. Traditionally, a "growth curve" does not have to deal with real growth (i.e., Condition D is not needed), but it must consider the dependence structure at different points of time. Furthermore, normality is still assumed.

Because of the assumption of normality, the study of growth curves is so well developed that rather complicated situations may be considered. See, for example, Lundbye-Christensen (1991). The general picture is that we are not too far away from the traditional likelihood setup, and it is often necessary to apply the EM algorithm. Dempster *et al*. (1981) is the standard reference, where the mixed effect model (some of the β are fixed effects while other β are random effect) is carefully sorted out. Using this idea, Chen *et al*. (1995) discussed the shelf life problem mentioned earlier. However, since the tests are destructive, it is reasonable to assume that measurements taken at different time points are independent. There is no such a convenience in traditional growth curve studies.

Our main concern is still prediction. For fixed a, $\{X_{at}, t\in T\}$ are dependent. Therefore, even with the normality assumption, for prediction at time $t_1>\sup T$, it is necessary to assume that the covariance matrix $\{X_{at}, t\in T\}$ has a natural definition for $t_1>\sup T$. When the covariance matrix is of either form listed below:

(a) $\Sigma=\delta_{ij}\sigma^2+(1-\delta_{ij})\rho\sigma^2$, or

(b) $\Sigma=\sigma^2(\rho^{|i-j|})$

Lee (1988) has a clear solution. Of course, case **(b)** corresponds to an AR(1) model. We believe it is possible to treat the case of AR(p). This generalization, if it can be done, is not too surprising.

Special credit must be given to Rao (1987). This article is rich in content and essence and is a key reference for this topic. For a more recent review, see Lee and Geisser (1996).

V. Sigmoid and the Like

The problems discussed in this section are all related to the fit of data to an S-shaped curve. A classical term for these curves is "sigmoid" (see Stone (1980)).

The basic idea of this type of curve fitting is that certain growths are by nature bounded above, so an S-shaped curve may be used to model the changing rate of growth from the beginning to the end. At first growth is slow and gradually speeds up. After it reaches a peak it then begins to slow down and eventually levels off.

Let $F(x)$ be any cumulative distribution function, then we may use $y=\theta F(x)+\epsilon$ to model this type of growth (see Boulanger and Escobar (1994)). Not all F's are S-shaped, of course, and except of traditional reasons, there is no theoretical reason to stick with the S-shape either. One of the most popular S-shaped growth curves is the Gompertz distribution.

The model

$$y_i=\frac{\alpha}{1+\exp\{A-Bt_i\}}+\epsilon_i,\ i=1,\ 2,\ \ldots,\ n \qquad \textbf{(1)}$$

is often called a logistic growth model. When $\epsilon_i\sim N(0,\sigma^2)$, it is necessary to use non-linear regression to estimate the parameters α, A, and B. In practice, α, $B>0$, and as $t\to\infty$, $y_i(t)\uparrow\alpha$ and α may be over-estimated. This model, and a slightly modified version, is used to estimate the yield of crops 15 days before harvest in India (see Jain *et al*. (1992)).

The logistic growth model was perhaps one of the earliest fitted growth models. Under certain birth-death processes, it can be shown that the expected size of the population at time t can be roughly represented by a logistic function (Tan and Pientadosi, 1991). For a more systematic treatment, see Tan (1992).

The classical approach starts from Eq. (1) with $\epsilon=0$. A differential equation is used to describe this curve. This suggests the possibility that a growth curve may be obtained from a differential equation.

Using this idea, various growth models can be obtained. Let $X(t)$ be a birth-death process with $X(0)=N_0$ which satisfies

$$Pr[X(t+\Delta t)=j+1|X(t)=j]=jb(t)\Delta t+o(\Delta t),$$

$$Pr[X(t+\Delta t)=j-1|X(t)=j]=jd(t)\Delta t+o(\Delta t).$$

Tan (1986) proved that

$$E[X(t)|x(0)=N_0]=N_0\exp\{\int_0^T\gamma(x)dx\},$$

where $\gamma(x)=b(x)-d(x)$. Here, we have some freedom in choosing the form of $\gamma(x)$. If we choose $\gamma(x)=\beta\exp\{b(x)-d(x)\}$, then Gompertz's form is obtained.

The logistic growth curve is also used to model rational selection by customers between two products to study the market shares of these products (see Oren and Schwartz (1988)).

Another example, which was used to study the propagation delay of an under sea cable, is given by

$$y=\theta(1-\exp\{-\sqrt{\lambda t}\})+\epsilon. \qquad \textbf{(2)}$$

Using this model, together with temperature as the accelerating factor (i.e., the Arrenius Law), Carey and Koenig (1991) obtained measurements at 100 °C, 150 °C and 175 °C and obtained 3 data sets. The purpose was to determine whether the propagation delay could be more than 2 nano seconds ($=2\times10^{-9}$ seconds) in 25 years at a use temperature of 40 °C. This is a well-documented report with many technical details except for raw data. Regressions, linear and non-linear, are used. This is an engineering paper; although it is not 100% rigorous, it is mostly satistactory. Their basic theory follows that of Jennrich (1969), which can be extended to mixed effect models (see Wu (1996)).

The real shortcoming of Carey and Koenig (1991) is that the condition "25 years" has never been used. In fact, its conclusion is based on the estimated value of θ, which is the value of y at $t\to\infty$. Therefore, if their conclusion is valid for 25 years, then it also applies to 10000, and the same logical conclusion is also obtained.

When used in extrapolation, the final inference from this type of model is sensitive to the functional forms employed. Its intermediate behavior is usually explained by the differential equation and its implied movement. Unless there are good reasons to believe that the same physical law applies to the far end of time, the conclusion, when far stretched, is relatively thin.

In the old days when computer use was not very popular, various techniques were used to facilitate estimation of an S-curve. But these usually caused confusion in the meaning of residuals. The comment by Oliver (1964) that "there is no substitute for full least squares in estimating the logistic function" is perhaps correct, and the same comment may be applied to other S-curves. For related software, see Ross (1980).

VI. Design of Experiment

Only a limited number of works have utilized the concept of experimental design in the analysis of degradation data. One report is that of Cullis and McGilchrist (1990). Although the model considered is a little bit complicated, basically it still is one of the growth models with the assumption of normality. Another work is that of Tseng *et al*. (1995), where a 2^{4-1} fractional factorial design was used. For some similar works, see Yu and Tseng (1999). These works start with the boundary crossing of the degradation path, and from this it is concluded that the respective life times follow a lognormal distribution. Then analysis of design data from this design is carried out under (transformed) normality. In this type of degradation study, however, little attention is paid to prediction. Rao (1987), when dealing with the prediction problem in growth curve models, found that when making a prediction, the weights of the "last few data points" play a more important role than do the other ones. Therefore, for a reasonable design where prediction (or extrapolation) is the main concern, one should consider the unbalanced version. A basic concept here is so-called *c*-optimality, and the proper paper to start with is Hoel and Levine (1964) (see also Chao (1995)).

Optimal designs are statistical methods seeking the proper locations of various settings that can be used to take observations in scientific experiments. Most traditional optimal designs are based on the invariance property of the respective design matrix X, typically characterized by its eigenvalues. The *c*-optimal design, on the other hand, looks for settings that maximize the variance of $\Sigma_i c_i \hat{\theta}_i$, where c_i's are constants and $\hat{\theta}_i$'s denote generic estimates. For degradation analysis, our main interest lies in the upper quantiles of the underlying distribution, which is usually of form $\Sigma_i c_i \hat{\theta}_i$. A typical example is the normal quantile which is estimated by $\hat{\mu}+z_\alpha\hat{\sigma}$, where $z_\alpha=\Phi^{-1}(1-\alpha)$. It is natural to expect that positive effects will result when the design concept is incorporated into growth curve analysis. But I haven't seen any work that uses the concept of experimental design in the study of growth curves with prediction.

VII. Stochastic Process

There are two basic approaches used to describe the degradation process: the direct approach and the indirect approach. In the direct approach, equations are set up first, and data fitting is the basic means of analysis. No, or little, attention has been paid to the question of why the model works. The indirect approach, on the other hand, starts from theoretical justification and then tries to derive more theory. Once in a while, we see something in between; the logistic growth curve, for example, is supported by both data and theory (Tan, 1986, 1992; Tan and Piantadosi, 1991).

Degradation is caused by enduring stress. The model based on this logic is the cumulative damage (CD) model. The earliest CD model was perhaps Miner's Law, which is deterministic, and we shall not discuss it further. Most of the CD models nowadays are related to the compound Poisson process (see Sanders (1982)).

The system fails when a certain amount of cumulative damage is inflicted. Let $X(t)$ denote a CD process, and let T be the life time of this system; then,

$$T=\inf\{t:\ X(t)\geq d\}$$

In principle, when the distribution of $X(t)$ is known, we can find the distribution of T.

Use of this technique is not limited to the CD models. Let

$$X(t)=a+bt+W(t)$$

(where $W(t)$ is a Brownian motion); then, T has a distribution which is inverse Gaussian. This is one of the few cases where we can find the exact solution (see Bhattacharyya and Fries (1982)). Using this relation, Doksum (1991) and Doksum and Normand (1995) performed analysis on degradation data (also see Whitmore (1995)). Work in this direction seems to have concentrated on finding the distribution or asymptotic distribution of T (see for example, Sethuraman and Young (1986), Desmond (1985), Berman (1970) and Whitmore *et al*. (1998)). All these works were essentially motivated by the classical work of Cramér and Leadbetter (1965).

It is natural to extend this work into the framework of Itô integral and related stochastic differential equations. This approach has been well studied in other fields, such as economics and finance. A good review paper is that of Bollerslev *et al*. (1994). I know of very little work in this direction using degradation data except for the initial report by Ueng (1988). Another approach is to study the fatigue of metal under stress and the growth of the crack length therein. A related physical law is the Paris Law (see Cinlar (1996) and Palettas and Goel (1996)). More recent works are those of Lu and Meeker (1993) and Meeker *et al*. (1998).

VIII. Step Stress Models

When the subject under inspection is too good to fail, one way to study its life time distribution is to employ accelerated life testing (ALT); when failure is even more difficult, we employ degradation analysis. In doing so, acceleration factors with increasing stress levels may also be applied. But there is no reason to restrict study to one constant level of stress; we may even increase the stress level on the same testing item as time goes on. This idea is emphasized in the step stress models.

Let $\{I_i, i=1, 2, ...\}$ be disjoint time intervals that sum to $[0,T]$. For $t \in I_i$, stress level V_i is applied to the testing item, and the amount of degradation $L(t;\theta)$ is observed. Here we assume that the I's have increasing end points, and that V_i increases with i. The parameter θ is of finite dimension, say $\theta=(\theta_1, \theta_2, ..., \theta_p)$.

The process $L(t,\theta)$ is observed and modeled piecewisely in each interval, with continuity assumed at the junctions of intervals. The key part of this model is that the model for $V=V_0$, the level of stress at the use condition, depends on only part of θ, say θ_1, θ_2. For $t \in I_i$, all θ_j's ($j\geq 3$) are treated as nuisance parameters, and the main problem is to estimate the important part of θ, i.e., (θ_1, θ_2).

Most of the step stress studies have been done by Tseng and his group (see Tseng and Wen (1996), Tseng and Yu (1997), Yu and Tseng (1998) and Tseng and Chiao (1998)).

IX. Conclusion

In this paper, we first described the general scenario of degradation analysis, and then examined the problem from various viewpoints. Almost all related works are motivated by the need for extrapolation. These methods are, by nature, inaccurate.

There has been no intention for this paper to do a thorough search. Rather, topics have been clustered to show the type of work that applied researchers have done based on their perspectives. There are various ways to describe the degradation phenomenon, and for solution there is no panacea. Among other things, the view we take is that we start from a probabilistic/ statistical model to describe the basic degradation process and hope to capture some essence of the underlying life distribution. Very little attention has been paid to methods based on physical laws in order to build a differential equation-based model. This does not imply, of course, that this approach is unimportant (see, for example, Wu (1993)). A good starting point in this direction is a book by Crandall and Mark (1963). Another related topic we have ignored is that of statistical methods related to nondestructive evaluation (see Olin and Meeker (1996)).

Acknowledgment

This research was partially supported by the National Science Council, R.O.C., under contracts NSC 85-2121-M-001-027 and NSC 87-2129-M-001-007.

References

Bancroft, T. A. (1964) Analysis and inference for incompletely specified models involving the use of preliminary test(s) of significance. *Biometrics*, **20**, 427-442.

Berman, S. M. (1970) Occupation times of stationary Gaussian processes. *J. App. Prob*., **7**, 721-733.

Bhattacharyya, G. K. and A. Fries (1982) Fatigue-failure models – Birnbaum-Saunders vs. inverse Gaussian. *IEEE Reli*., **R-31**, 439-440.

Bogdanoff, J. L. and F. Kozin (1985) *Probability Models of Cumulative Damage*. Wiley, New York, NY, U.S.A., London, U. K., and Sydney, Australia.

Bollerslev, T., R. F. Engle, and D. B. Nelson (1994) ARCH models. In: *Handbook of Ecomometrics*, IV. R.F. Engle and D.L. McFadden Eds. Elsevier Science, Amsterdam, Netherlands.

Boulanger, M. and L. A. Escobar (1994) Experimental design for a class of accelerated degradation tests. *Technometrics*, **36**, 260-272.

Carey, M. B. and R. H. Koenig (1991) Reliability assessment based on accelerated degradation: a case study. *IEEE Reli*., **40**(5) 499-506.

Chao, M. T. (1995) Some robust type of extrapolation design. *Proceedings of the International Conference on Statistical Methods and Statistical Computing for Quality and Productivity improvement*, **1**, pp. 99-108. Seoul, Korea.

Chen, J. C., J. S. Hwang, and Y. Tsong (1995) Estimation of the shelf-life of drugs with mixed effects models. *J. Biopharmaceutical Statist*., **5**(1), 131-140.

Chow, S. C. and J. P. Liu (1995) *Statistical Design and Analysis in Pharmaceutical Science*. Marcer Dekker, New York, NY, U.S.A.

Chow, S.C., and J. Shao (1991) Estimating drug shelf-life with random effects. *Biometrics*, **47**, 1071-1079.

Cinlar, E. (1996) Fatigue crack growth. In: *Reliability and Maintenance of Complex Systems*. pp. 37-52. S. Ozekici Ed. Springer, New York, NY, U.S.A. and Heidelberg and Berlin, Germany.

Cramér, H. and M. R. Leadbetter (1965) The moments of the number of crossings of a level by a stationary normal process. *Ann. Math. Statist.*, **36**, 1656-1663.

Crandall, S. H. and W. D. Mark (1963) *Random Vibration in Mechnical Systems*. Academic Press, New York, NY, U.S.A.

Cullis, B. R. and C. A. McGilchrist (1990) A model for the analysis of growth data from designed experiments. *Biometrics*, **46**, 131-142.

Dempster, A. P., D. B. Rubin, and R. K. Tsutakawa (1981) Estimation in covariance components models. *JASA*, **76**, 341-353.

Department of Defense, U.S.A. (1991) *MIL-HDBK-217F: Reliability Prediction of Electronic Equipment*. Department of Defense, U.S.A., Washington, D.C., U.S.A.

Desmond, A. (1985) Stochastic models of failure in random environments. *Can. J. of Statist.*, **13**(2), 171-183.

Doksum, K. A. (1991) Degradation rate models for failure time and survival data. *W. I. Quarterly*, **4**, 195-203.

Doksum, K. A. and S. T. Normand (1995) Gaussian models for degradation processes – Part I: Methods for the analysis of biomarker data. *Life Data Analysis*, **1**, 131-144.

Elston, R. C. and J. E. Grizzle, (1962) Estimation of time response curves and their confidence band. *Biometrics*, **18**, 148-159.

Faragher, E. B., E. N. Chantler, and C. T. Roberts (1985) A statistical evaluation of the rate of degradation of copper releasing IUCDs. *The Statistician*, **34**, 243-248.

FDA, U.S.A. (1987) *Guideline for Submitting Documentation for the Stability Study of Human Drugs and Biologies*. Center for Drugs and Biologies, Office of Drug Research and Review, Food and Drug Administration, Rockville, MD, U.S.A.

Hoel, P. G. and A. Levine (1964) Optimal spacing and weighting in polynomial prediction. *Ann. Math. Statist.*, **36**, 1483-1493.

ICH, Switzerland (1993) *Harmonised Tripartite Guideline on Stability Testing of New Drug Substance and Products*. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, The ICH Secretariat, Geneva, Switzerland.

Jain, R. C., R. Agrawal, and K. N. Singh (1992) A within year growth model for crop yield forecasting. *Biometrical J.*, **34**(7), 789-799.

Jennrich, R. I. (1969) Asymptotic properties of nonlinear least squares estimators. *Ann. Math. Statist.*, **40**, 633-643.

Kirkwood, T. B. L. (1977) Predicting the stability of biological standards and products. *Biometrics*, **33**, 736-742.

Lee, J. C. (1988) Prediction and estimation of growth curves with special covariance structures. *JASA*, **83**(402), 432-440.

Lee, J. C. and S. Geisser (1996) *On prediction growth curves. In: Modeling and Prediction Honoring Semour Geisser*, pp. 77-103. J.C. Lee, W. Johnson, and A. Zellner Eds. Springer, New York, NY, U.S.A. and Heidelberg and Berlin, Germany.

Lu, C. J. and W. Q. Meeker (1993) Using degradation measures to estimate a time-to-failure distribution. *Technometrics*, **35**(2), 161-174.

Lundbye-Christensen, S. (1991) A multivariate growth model for pregnancy. *Biomertics*, **47**, 637-657.

Meeker, W. Q., L. A. Escobar, and C. J. Lu (1998) Accelerated degradation tests: modeling and analysis. *Technometrics*, **40**, 89-99.

MHW, Japan (1991) *Stability Guidelines*. Ministry of Health and Welfare, Tokyo, Japan.

Murphy, J. R. and D. Weisman (1990) Using random slopes for estimating shelf-life. In: *Biopharmaceutical Proceedings*, pp. 196-200. Amer. Statist. Asso., Alexandria, VA, U.S.A.

Nelson, W. (1981) Analysis of performance-degradation data form accelerated tests. *IEEE Reli.*, **R-30**(2), 149-155.

Olin, B. D. and W. Q. Meeker (1996) Applications of statistical methods to nondestructive evaluation (with discussion). *Technometrics*, **38**, 95-130.

Oliver, F. R. (1964) Methods of estimating to logistic growth functions. *Applied Statist.*, **xiii**(2), 57-66.

Oren, S. S. and R. G. Schwartz (1988) Diffusion of new products in risk-sensitive markets. *J. of Forecasting*, **7**, 273-287.

Palettas, P. N. and P. K. Goel (1996) Predictive modeling for fatigue crack propagation via linearizing time transformations. In: *Reliability and Maintenance of Complex Systems*, pp. 53-69. S. Ozekici Ed. Springer, New York, NY, U.S.A. and Heidelberg and Berlin, Germany.

Rao, C. R. (1987) Prediction of future observations in growth curve models (with discussion). *Statistical Science*, **2**(4), 434-471.

Ross, G. J. S. (1980) MLP, *Maximum Likelihood Program (a Manual)*. The Statistics Department, Rothamsted Experimental Station, Harpenden, Hertfordshire, U.K.

Ruberg, S. J. and Stegeman, J. W. (1991) Pooling data for stability studies: Testing the equality of batch degradation slopes. *Biometrics*, **47**, 1059-1069.

Sanders, S. C. (1982) Cumulative damage models. In: *Encyclopedia of Statistical Science*, **2**, pp. 230-233. S. Kotz and N.L. Johnson Eds. Wiley, New York, NY, U.S.A., London, U.K., and Sydney, Australia.

Sethuraman, J. and T. R. Young (1986) Cumulative damage threshold crossing models. In: *Reliability and Quality Control*, pp. 309-319. Elsevier, Amsterdam, Netherlands.

Shao, J. and Chow, S. C. (1994) Statistical inference in stability analysis. *Biometrics*, **50,** 753-763.

Stone, R. (1980) Sigmoids. *BIAS*, **7**(1), 59-119.

Tan, W. Y. (1986) A stochastic Gompertz birth-death process. *Stat. Prob. Letters*, **4**, 25-48.

Tan, W. Y. (1992) Logistic stochastic growth models and applications. In: *Hand-books of the Logistic Distribution*, pp. 397-426. N. Balakrishnan Ed. Dekker, New York, NY, U.S.A.

Tan, W. Y. and S. Piantadosi (1991) On stochastic growth procedures with applications to stochastic logistic growth. *Statistica Sinica*, **1**, 527-540.

Tseng, S. T. and C. H. Chiao (1998) NSC Report for Project NSC 86-2115-N-007-037, National Science Council, R.O.C., Taipei, Taiwan, R.O.C.

Tseng, S. T., M. Hamada, and C. H. Chiao (1995) Using degradation data to improve fluorescent lamp reliability. *J. Qual. Tech.*, **27**(4), 363-369.

Tseng, S. T. and Z. C. Wen (1996) *Step-stress Accelerated Degradation Analysis of Highly Reliable Products*. NSC Report for Project NSC 85-2121-M-007-041, National Science Council, R.O.C., Taipei, Taiwan, R.O.C.

Tseng, S. T. and H. F. Yu (1997) A termination rule for degradation experiments. *IEEE Reli.*, **46**, 130-133.

Ueng, C. Y. (1988) NSC Report for Project NSC 87-2199-M-001-007, National Science Council, R.O.C., Taipei, Taiwan, R.O.C.

Ware, J. H. (1983) *Growth curves. In: Encyclopedia of Statistical Science*, **3**, pp. 539-542. S. Kotz and N.L. Johnson Eds. Wiley, New York, NY, U.S.A., London, U.K., and Sydney, Austrlia.

Wei, G. C. G. (1998) Simple methods for determination of the release limits for drug products. *J. of Biopharmaceutical Statist.*, **8**, 103-112.

Whitmore, G. A. (1995) Estimating degradation by a Wiener diffusion process subject to measurement error. *Life Data Analysis*, **1**, 307-319.

Whitmore, G. A., Crowder, M. J., and Lawless, J. F. (1998) Failure inference from a marker process based on a bivariate Wiener model. *Life Data Analysis*, **4**, 229-251.

Wu, S. J. (1996) *Reliability Analysis Using the Least Squares Method in Non-linear Mixed Effect Degradation Models*. Ph.D. Dissertation. Dept. of Statis., University of Wisconsin, Madison, WI, U.S.A.

Wu, W. F. (1993) Computer simulation and reliability analysis of fatigue crack propagation under random loading. *Engineering Fracture Mechanics*, **45**, 697-712.

Yu, H. F. and S. T. Tseng (1998) Online procedure for terminating an accelerated degradation test. *Statistica Sinica*, **8**(1), 207-220.

Yu, H. F. and S. T. Tseng (1999) Designing a degradation experiment. *Naval Research Logistics*, (to appear).

衰變分析及相關問題：一些想法和回顧

趙民德

中央研究院統計科學研究所

摘　要

衰變是物體逐漸變壞的現象，經由測量其某些品質特性而表現出來。當受測物品因品質太好而不易失效時，衰變分析就成為一種可由此對該物品的壽命分布作統計推論的工具。本文分為兩部分，第一部分作對衰變現象及相關問題作一般性的探討；第二部分則回顧了一些相關的特殊議題，這包括了藥品的上架壽命、成長曲線、S-型曲線、外插實驗設計、隨機過程模型及逐步應力模型等。