A critical look at some point process models for repairable systems

Citation for published version (APA):

DOI:
10.1093/imaman/4.4.375

Document status and date:
Published: 01/01/1992

Document Version:
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:
• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher’s website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:
www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:
openaccess@tue.nl
providing details and we will investigate your claim.
A critical look at some point-process models for repairable systems

MARTIN NEWBY
Eindhoven University of Technology, and Frits Philips Institute for Quality Management, The Netherlands

The use of a model-driven approach to the analysis of repairable systems is considered and shown to be useful as a way of understanding the characteristics of such a system. However, considerable statistical problems arise from the use of a set of standard model-building elements. In particular, identification problems arise in many of the models. The argument is illustrated by examples from software reliability and mechanical reliability. The conclusion is that, in many cases, the exploratory data-analysis approach is as effective as the use of more sophisticated models.

1. Introduction
The successful use of stochastic models and the application of statistical techniques in the study of complex repairable systems seems to remain rather limited. There seems to be a gap between a model-driven approach and the statistical (data-analysis) approach. Models can be used to give insight into system behaviour, but are difficult to analyse statistically in terms of finding estimators of their parameters. Simpler approaches based on what is intuitively attractive in statistical terms also seem to lose the essential features of the systems studied. In other words, given knowledge of the structure of the system and the behaviour of the components, we can model it successfully; but, when we must estimate model parameters in order to extrapolate the behaviour of a system, we meet problems.

In this paper, I first outline some of the models and indicate theoretically why the estimation through likelihood methods is not always successful. Secondly, on the basis of a few examples, the difficulties with likelihood are also illustrated. The conclusions to be drawn are that, in many cases, the exploratory data-analysis approach outlined by Bendell & Walls (1985) and Ansell & Phillips (1989) yields as much information as the data are capable of giving. In particular, the lack of identifiability in some of the model-based approaches means that observation of explanatory variables adds nothing to the knowledge obtained from the event data.

2. Dependency in system modelling
Some of the difficulties seem to lie with our feelings about what a model should describe, and with our lack of clarity about what we want to know. The problems mainly seem to arise from the desire for simplification so that statistical analysis is possible. The difficulty forms a sub-theme in Ascher & Feingold (1984), where they discuss extensively the problems of aggregation. In a complex system, subsystems
and components follow different processes, but the data are often recorded without reference to these subprocesses. Thus we try to analyse the data with one of the simpler or better-known point processes. The history of the subject also plays a part. Engineers have always been more concerned with models that deal with structural dependence—mostly in a deterministic sense until relatively recently, but also through the use of Markov models (e.g. Billinton & Allan 1983, 1984). These models capture the essence of system behaviour, and allow different scenarios to be explored, but are usually difficult from the point of view of statistical analysis. On the other hand, many of the methods of dealing with dependence on explanatory variables have been borrowed from other areas of statistics—medical statistics in particular (Gore et al. 1984; Gail et al. 1980; Prentice et al. 1981). The questions asked in medical statistics are frequently about the effect of a particular treatment on survival; they deal rarely with the times between reoccurrences of an illness, and pay little regard to the structure of the system. Recently developments in econometrics have led to techniques for dealing with the statistical analysis of stochastic processes that more closely resemble the problems faced in reliability modelling (Heckman & Singer 1984; Lancaster 1990).

In this paper, some natural approaches to dealing with dependency on covariates for repairable systems are outlined. The approach maintains one common feature in almost all the models for repairable systems: the assumption that the intervals between failures are statistically independent. The development through time is reflected in the parameters of the distribution of the interval, and these parameters are often in turn modelled as functions of explanatory covariates. The covariates may be as simple as the index number of the failure interval or a measure of maintenance effort. Most of the models fall in the class of semi-Markov processes, but in only one is the state transition matrix explicitly required (Brown & Proschan 1983).

3. Nonrepairable systems

A nonrepairable system in this context is simply one that experiences only one type of event, called failure, and is not restored to working order when it fails. Nonrepairable systems are the subject of survival analysis and are therefore important in medical statistics and the study of component reliability. The knowledge of component behaviour obtained from survival analysis can be used as a building block in the study of complex systems. A renewal process describes a system that, after a failure, is simply replaced by a new system with the same characteristics, so that the life distribution of the system is enough to deduce all the properties of the system. In more complex cases, if a system has regeneration points, and if the distribution of time between regeneration points is known, then a great deal can be learnt about the system’s behaviour. Many models for nonrepairable systems can reappear in what Cox (1973) calls modulated renewal processes. Proportional-hazards extensions of the idea of a modulated renewal process are discussed in detail by Prentice et al. (1981).

As well as studying the effects of the measurable covariates on the life of the component, we may also have to deal with heterogeneity in the observed systems. The heterogeneity may be intrinsic—for example, different suppliers may deliver systems of differing qualities—or there may be errors of measurement in the duration data and the covariates (Lancaster 1990; Vaupel et al. 1979). In the following paragraphs, the basic models are outlined and illustrated; they can be used in combination as building blocks to model system behaviour.

4. Accelerated failure time

Perhaps the most intuitively attractive model is that of accelerated failure time. In this model, the effect of the covariates is assumed to be seen as changes in the time scale for the system. Suppose that the duration is a random variable $T_i$ and that the covariates are summarized in a vector $z = (z_1, ..., z_p)$; then the assumption is that, if the duration of a system under standard conditions is $T_0$, then its duration under conditions described by $z$ is

$$T = \psi(z, \beta) T_0,$$

where $\beta$ is a vector of parameters. The life is increased or decreased according to whether $\psi < 1$ or $\psi > 1$. The behaviour of the model is most easily understood by considering the effect of the transformation (4.1) on a family of distributions. Suppose that $F_0(\cdot)$ is a scalar distribution with density function $f_0(\cdot)$ and hazard rate $h_0(\cdot)$. The system operating with covariate $z_j = (z_{j1}, ..., z_{jm})$ has a scale-transformed distribution given by

$$F_j(x) = F_0\left(\frac{x}{\psi(z_j, \beta)}\right) (j = 1, ..., r),$$

Estimation and inference for this model are particularly straightforward if the acceleration factor is taken to be a quasi-linear, in the sense that we can express $\psi(z, \beta)$ in the form $\psi(\beta^T z)$. A detailed treatment of estimation and interpretation for such models can be found in Newby & Winterton (1983) and Newby (1985, 1988).

Introducing a location parameter offers the possibility of reflecting more of the features of a process as it unfolds through time. These models can be written in terms of the baseline distribution as

$$F_j(x) = F_0\left(\frac{x + \theta(z_j)}{\psi(z_j, \beta)}\right) (j = 1, ..., r).$$

The presence of the location parameter allows the effects of accumulated aging or imperfect repair to be more effectively represented. The likelihoods for these models are not much more complicated than those for the two-parameter models, although care has to be taken with the estimation of location parameters (Newby 1988; Smith & Naylor 1987). Generally a grouped likelihood performs better than the likelihood itself (Cheng & Amin 1983; Cheng & Iles 1987).

5. Proportional hazards

One of the most widely used methods for the study of the effects of covariates is the proportional-hazards model. The basis of the model is the simple assumption that
the hazard rate is affected in a multiplicative way by a relative risk factor. Since the hazard rate is a measure of aging, an increase in relative risk is an indication of more rapid aging. The model is simply expressed in terms of a baseline hazard rate \( h_0(x) \) and a relative risk factor \( \psi(z, \beta) \) dependent as above on a covariate vector \( z \). The system operating with covariate \( z_j = (z_{j1}, \ldots, z_{jm}) \) has a hazard rate

\[
h_t(x) = \psi(z_j, \beta) h_0(x) \quad (j = 1, \ldots, r)
\]

with the assumption that, for standard conditions, \( \psi(z) = 1 \). Naturally \( \psi \) must be non-negative, and the commonest choice is the Cox model (Cox 1972) in which \( \psi(z) \) is independent as above on a covariate vector \( z \). The model is simply expressed in terms of a baseline hazard rate

\[\psi(z, \beta) = \exp \beta^T z.\]

The basic properties of the model are easy to deduce. However, the most valuable property is that, if we are prepared to leave the baseline hazard rate unspecified, the analysis of the effects of covariates can proceed on the basis of a partial likelihood alone. With such a nonparametric approach, the importance of covariates can be explored, and predictions can be made about the effect of covariates on the hazard rate is a measure of aging, an increase in relative risk is an indication of more rapid aging. The model is simply expressed in terms of a baseline hazard rate \( h_0(x) \) and a relative risk factor \( \psi(z, \beta) \) dependent as above on a covariate vector \( z \). The system operating with covariate \( z_j = (z_{j1}, \ldots, z_{jm}) \) has a hazard rate

\[
h_t(x) = \psi(z_j, \beta) h_0(x) \quad (j = 1, \ldots, r)
\]

with the assumption that, for standard conditions, \( \psi(z) = 1 \). Naturally \( \psi \) must be non-negative, and the commonest choice is the Cox model (Cox 1972) in which \( \psi(z) \) is independent as above on a covariate vector \( z \). The model is simply expressed in terms of a baseline hazard rate

\[\psi(z, \beta) = \exp \beta^T z.\]

The basic properties of the model are easy to deduce. However, the most valuable property is that, if we are prepared to leave the baseline hazard rate unspecified, the analysis of the effects of covariates can proceed on the basis of a partial likelihood alone. With such a nonparametric approach, the importance of covariates can be explored, and predictions can be made about the effect of covariates on the relative risk; however, to make predictions about the future behaviour of systems, a parametric model generally has to be employed. The details of these models are readily accessible in the literature (Ansell & Phillips 1989, 1990; Bendell et al. 1986; Cox & Oakes 1984; Prentice et al. 1981) and are also available in statistical packages such as BMDP.

6. Additive hazards

In this case, the baseline hazard \( h_0 \) is assumed to be modified in an additive way by the effect of covariates (Pijnenburg 1991):

\[h_j(x) = \psi^j + h_0(x) \quad (j = 1, \ldots, r),\]

where we define

\[\psi_j = \psi(z_j, \beta).\]

For \( j = 1, \ldots, r \), the cumulative hazard functions are

\[H_j(x) = \psi_jx + H_0(x),\]

the survivor functions become

\[R_j(x) = \exp[-H_j(x)] = \exp[-(\psi_jx + H_0(x))] = R_0(x) \exp[-\psi_jx],\]

and the densities are then

\[f_j(x) = [\psi_j + h_0(x)] \exp[-(\psi_jx + H_0(x))].\]

On considering the competing-risks model, it is clear that the additive-hazards model is a form of competing-risks model with two (possibly fictional) components in series. However, the failing component cannot be determined. This remark is useful for simulating data from an additive model, since the failure time \( X \) is simply the minimum of (i) an exponential random variable \( X_\psi \) sampled from the distribution with parameter \( \psi_j \) and (ii) a random variable \( X_0 \) sampled from the baseline distribution. Mercer (1961) showed how an additive form arose naturally in a wear-dependent renewal process in which a stochastic wear process contributed additively to the failure rate of a component.

The moments are easily obtained by Laplace transformation. It is easy to see that the transforms of the distributions and densities are

\[
\tilde{R}_j(s) = \int_0^\infty R_0(x) e^{-sx} \exp[-sx] dx = \tilde{R}_0(s + \psi_j),
\]

\[
\frac{1}{s + \psi_j} \left(1 - \int_0^\infty f_0(x) e^{-(s + \psi_j)x} dx\right) = \frac{1}{s + \psi_j} \left[1 - \tilde{f}_0(s + \psi_j)\right],
\]

\[
\tilde{F}_0(s + \psi_j) = \frac{1}{s + \psi_j} \tilde{f}_0(s + \psi_j), \quad \tilde{f}_0(s + \psi_j) = \psi_j \tilde{R}_0(s + \psi_j) + \tilde{f}_0(s + \psi_j),
\]

for \( j = 1, \ldots, m \). The transform of \( f_0 \) can be written in terms of the noncentral moments \( \nu_n \) \((n = 0, 1, \ldots)\) as

\[
\tilde{f}_0(s + \psi_j) = \sum_{n=0}^\infty (-1)^n (s + \psi_j)^n \nu_n.\]

Setting \( s = 0 \) in the above gives the first moments; higher moments can then be derived by differentiating under the integral.

\[
m_1(\psi_j) = \int_0^\infty R_0(x) dx = \int_0^\infty R_0(x) e^{-vx} dx = \int_0^\infty x \psi_j + h_0(x) dx = \frac{1}{\psi_j} \left(1 - \int_0^\infty f_0(x) e^{-vx} dx\right) = \frac{1}{\psi_j} \left[1 \tilde{f}_0(0)\right].
\]

Write \( c_n = \nu_{n+1}/(n + 1) \). Then the first moment is

\[
m_1(\psi_j) = \sum_{n=0}^\infty (-1)^n \psi_j^n c_n.
\]

By repeatedly differentiating (6.1), the higher moments are obtained as

\[
m_p(\psi_j) = (-1)^{p-1} \frac{d^{p-1}}{d\psi_j^{p-1}} m_1(\psi_j).
\]

The results yield the following expression for the noncentral moments:

\[
m_p = \sum_{n=0}^\infty (-1)^n \psi_j^n \frac{d^{p-1}}{d\psi_j^{p-1}} c_{n+p-1}.
\]

To derive the likelihood, assume that observations are made at \( n \) values \( z_1, \ldots, z_n \) of the covariate, \( x_{ij} \) is the \( i \)th observation (censored or actual) at level \( z_j \) \((i = 1, \ldots, l_j; j = 1, \ldots, n)\), and \( L^0(\beta, \theta) \) denotes the contribution to the log-likelihood for the observations (actual and censored) at \( z_j \), where the parameter vector \( \beta \) is \( p \)-dimensional and \( \theta \in \mathbb{R}^k \) is the parameter vector for \( f_0 \). The required
likelihood equations are

\[
L^{(j)}(\beta, \theta) = \sum_{i \in C_j} \ln f_j(x_{ji}, \psi_j) + \sum_{i \in C_j \cup D_j} \ln R_j(x_{ji}, \psi_j) = \sum_{i \in C_j} \ln [\psi_j + h_0(x_{ji})] - \psi_j \sum_{i \in C_j \cup D_j} x_{ji} - \sum_{i \in C_j \cup D_j} H_0(x_{ji}),
\]

where \( C_j \) is the \( j \)th index set of censored observations and \( D_j = \{1, \ldots, l_j\} \cap C_j \) \((j = 1, \ldots, n)\). The overall likelihood is

\[
L = \sum_{j=1}^{n} L^{(j)}(\beta, \theta).
\]

Now write \( \psi_j(\beta, \theta) = (\partial/\partial \beta_i) \psi(z_{ji}, \beta) \) \((j = 1, \ldots, n; k = 1, \ldots, p)\). The necessary derivatives are

\[
\frac{\partial L^{(j)}}{\partial \beta_k} = \psi_j \left( \sum_{i \in D_j} \frac{1}{\psi_j + h_0(x_{ji})} - \sum_{i \in C_j \cup D_j} x_{ji} \right) = \psi_j \mu_j(\beta, \theta),
\]

which can be written as

\[
\frac{\partial L^{(j)}}{\partial \theta_i} = \sum_{i \in C_j \cup D_j} \frac{h_0(x_{ji})}{\psi_j + h_0(x_{ji})} - \sum_{i \in C_j \cup D_j} H_0(x_{ji}),
\]

where \( h_0 \) and \( H_0 \) denote the derivatives of \( h_0 \) and \( H_0 \) with respect to \( \theta_i \). The likelihood equations are

\[
\frac{\partial L}{\partial \theta_i} = \sum_{j=1}^{n} \frac{\partial L^{(j)}}{\partial \theta_i} = \sum_{j=1}^{n} \left( \sum_{i \in C_j \cup D_j} \frac{h_0(x_{ji})}{\psi_j + h_0(x_{ji})} - \sum_{i \in C_j \cup D_j} H_0(x_{ji}) \right) = 0,
\]

\[
\frac{\partial L}{\partial \beta_k} = \sum_{j=1}^{n} \frac{\partial L^{(j)}}{\partial \beta_k} = \sum_{j=1}^{n} \psi_j \mu_j(\beta, \theta) = 0.
\]

This last can be written as a matrix equation on setting

\[
\Psi(z, \beta) u(\theta, \beta) = 0,
\]

namely

\[
\Psi(z, \beta) u(\theta, \beta) = 0,
\]

where \( u = (u_1, \ldots, u_n) \). Equation (6.2) shows that, for a fixed \( \theta \), there is always the solution

\[
u(\beta, \theta) = 0,
\]

i.e.,

\[
\sum_{i \in D_j} \frac{1}{\psi_j + h_0(x_{ji})} - \sum_{i \in C_j \cup D_j} x_{ji} = 0 \quad (j = 1, \ldots, n),
\]

and that the existence of other solutions depends on the rank of \( \Psi \). When \( p < n \), the rank of \( \Psi \) is at most \( p \) and there more solutions than those of \( u(\beta, \theta) = 0 \). Further

\[
E\left( \frac{1}{h(X)} \right) = E(X).
\]

Thus, if \( h(t) = \psi_j + h_0(t) \) is a hazard rate, then it should satisfy (6.4), and the empirical version of (6.4) is just (6.3). This means that it is unlikely that a set of solutions can be found which, at the same time, satisfy the conditions for a hazard rate and are also other than the trivial solutions of (6.3).

7. Competing risks

The competing-risks model (David & Moeschberger 1978) has two interpretations: the first, as its name suggests, describes the lifetime of a system subject to several potential causes of failure; the second describes the lifetime of a series of components which fails as soon as one of the components fails. The occurrences of the potential failures can be regarded as a vector of random variables \( (T_1, \ldots, T_n) \) so that the failure time is the minimum of the \( T_i \). If the survival function of \( T_i \) is \( R_i(\bullet) \) and the random variables \( T_i \) are independent, the system survival function is easily found to be

\[
R_S(x) = \prod_{i=1}^{n} R_i(x),
\]

with hazard rate

\[
h_S(x) = \sum_{i=1}^{n} h_i(x).
\]

Competing risks arise naturally in reliability problems, in particular in the analysis of series systems and from 'weakest-link' arguments for systems consisting of statistically independent components. The '\( \beta \)-factor' method for dealing with dependency is also a version of a competing risks model since, in effect, it analyses the system as a series of (i) a subsystem of independent components and (ii) a common-cause component (Lewis 1987). The (non-)identifiability results reported in Crowder (1991) show how to construct a competing-risks model that correctly reflects the behaviour of the system.

8. Frailty or mixtures

The idea of frailty or mixture models can be used in two ways: first simply as a way of introducing an idea of heterogeneity into the construction of a model, and secondly as an object of interest in itself. In demography and econometrics, the identification of a mixing distribution is of some importance; in reliability, the use of mixing distributions is more commonly a step in model building.

Frailty (Vaupel et al. 1979) is defined rather like proportional hazards, but differs in that the relative risk factor is a random variable in this case. Vaupel et al. define frailty \( i \) in terms of the hazard rates of individuals in a population. They write

\[
h(x | i) = \lambda h_0(x)
\]
for the hazard rate of an individual with frailty \( \lambda \); \( h_0 \) is again a standardized, or baseline, hazard rate. If the frailty at time \( x \) has a density \( \omega_x(\cdot) \), the average hazard rate at time \( x \) is

\[
\hat{h}(x) = \int_0^\infty h(x | \lambda) \omega_x(\lambda) d\lambda = h_0(x) \int_0^\infty \lambda \omega_x(\lambda) d\lambda = \hat{\lambda} h_0(x).
\]

Now, if the frailty \( \lambda_x \) decreases with time, so will \( \hat{\lambda}_x \) (since the weakest die young), and we see the apparent effect of the average hazard rate for the population declining more rapidly than the hazard rate for individuals.

Mixture models also arise naturally in reliability (Lancaster 1990; Littlewood & Verrall 1973) when we write the hazard rate as a conditional hazard rate \( h(x | z, \lambda) \); here, as usual, \( z \) is a covariate and \( \lambda \) is a random variable with density \( w \), and the unconditional density and survivor function for \( x \) are

\[
\begin{align*}
  f(x | z) &= \int f(x | z, \lambda) w(\lambda) d\lambda = \int h(x | z, \lambda) \exp[-H(x | z, \lambda)] d\lambda, \\
  R(x | z) &= \int R(x | z, \lambda) w(\lambda) d\lambda = \int \exp[-H(x | z, \lambda)] d\lambda.
\end{align*}
\]

But note that the hazard rate defined in terms of frailty is not the hazard rate of the unconditional distribution, that is,

\[
h(x | z) = \frac{f(x | z)}{R(x | z)} = \int h(x | z, \lambda) \omega(\lambda) d\lambda = \hat{h}(x | z).
\]

This observation also applies when we interpret the mixing model as a Bayesian model with prior \( w(\lambda) \) for a parameter \( \lambda \); here again, care has to be taken over the choice of meaning for a hazard function, particularly when we wish to estimate a hazard function from data. For example, the Littlewood–Verrall model (Littlewood & Verrall 1973) mixes the density functions of the inter-failure times between the appearance of faults in a piece of software. However, the mixing process also seems to carry with it identification problems (Lancaster 1990; Ridder 1990).

### 9. Repairable systems

The choice of an appropriate model depends on what is expected of the data. In a repairable system, a renewal process typifies the behaviour of system in which a failed component is replaced by an identical new component. A lamp with the replacement of failed bulbs is a perfect example of a renewal process. The two points about the renewal process are that the system returns instantly to an ‘as new’ state and ages in exactly the same way as before. If the ideas of a renewal process are extended, then the system can be returned to the working state, but in a condition between the new state and the failed state: after a repair, the system may age more rapidly than before. All of the above models can incorporate each of these features, although not always in one model. The questions that are asked about such a system are ‘can the development of the system through time be predicted?’ and ‘can the effects of different operating conditions be incorporated in the model (for example, can the effects of modification or a different maintenance regime be modelled)?’ The easiest assumption is that the interval lengths remain statistically independent; the use of covariates models changes in the distribution of lengths. If the process is described by an intensity function \( \lambda(\cdot) \) which gives the probability that the current interval ends in \([t, t + dt] \) as \( \lambda(t) dt \), then we have a nonhomogeneous Poisson process (with \( \lambda \) continuous over the whole time period). If the interval after the event at \( t_i \) has hazard rate \( \hat{\lambda}_i(t - t_i) \), then the process can be called a modulated renewal process (Thompson 1981; Cox 1973; Prentice et al. 1981). In this second case, \( \lambda \) is a piecewise continuous function with a discontinuity at each failure/renewal time. It is also possible to have an intermediate situation in which the rate \( \lambda \) has discontinuities only at certain events and not at others (Brown & Proschan 1981).

The techniques outlined for a single sample can be extended, following Cox (1973), to a modulated renewal process in place of the ordinary renewal process: the assumption of independent and identically distributed intervals is dropped and replaced by that of the interval length depending only on the state at the beginning of the interval, which may in turn depend on the number of preceding intervals but not on their length. The assumption is that the \( j \)th interval length \( X_j \) is distributed as \( F_j(\cdot) \). The aim of the models is to reflect in the \( F_j \) the changes in the system as its history unfolds. Through such models it is hoped to discover whether the system improves or deteriorates through time, and for example, to determine optimal maintenance or replacement policies.

The simplifying assumptions that the \( j \)th interval has distribution \( F_j \) means that the durations of the \( j \)th interval can be treated simply as a sample from \( F_j \) and used to construct a log-likelihood \( L^{(b)} \) for that interval alone, the overall log-likelihood is the sum

\[
L = \sum_{j=1}^n L^{(b)}.
\]

This likelihood also provides likelihoods for the parameters of \( F_j \) as functions of explanatory variables. The estimators are, in principle, obtained as solutions of the likelihood equations obtained by setting the appropriate derivatives of \( L \) equal to zero. Systematic use of the chain rule frequently produces simple relationships between the ordinary log-likelihoods and those where there is assumed to be an underlying model.

Techniques for the analysis of renewal processes can also be carried over, in principle, to the modulated renewal process, although explicit closed formulæ for measures of interest are mostly not available. If the explanatory variables are deterministic, interval number for example, we can proceed as for a renewal process: the time \( t_n \) of event \( n \) is simply the sum of the \( n \) independently distributed interval lengths \( X_i \):

\[
t_n = \sum_{i=1}^n X_i,
\]

so that the Laplace transform of \( g_\lambda(\cdot) \), the density of \( t_n \), can be written as

\[
\hat{g}_\lambda(s) = \prod_{i=1}^n \hat{f}(s).
\]
where \( f_i \) is the density function for \( F_i \), and the renewal function \( V(\cdot) \) has Laplace transform
\[
\hat{V}(s) = \frac{1}{s} \sum_{i=1}^{\infty} \hat{G}_i(s).
\]

When it comes to using these building bricks, a number of points are clear. Proportional-hazards and accelerated-failure-time approaches can only model the rate of aging between events. However, if these models are based on a baseline hazard rate \( h_0(\cdot) \) with the property \( h_0(0) = 0 \), then every interval hazard rate \( h_i \) also satisfies \( h_i(0) = 0 \). If the hazard rate is zero at time zero, then the system is instantaneously as good as new after a repair, even though thereafter it may age faster. An accumulation of aging will almost always require the interval hazard rate to satisfy this condition.

Proportional-hazards and accelerated-failure-time approaches can only model the accelerated-failure-time model with a location shift satisfy this condition in general.

These two models also enable some more complicated models, imperfect repair, and the two simple models are a nonhomogeneous Poisson process and a renewal process, which have as the unique common model the ordinary Poisson process. These two models also enable some more complicated models, imperfect repair, and alternating renewal processes to be analysed with the same techniques.

The history of the simplest processes is described by a sequence of times \( t_1, t_2, \ldots \) at which events, usually called failures, occur, in effect, repair times are ignored. The process can also be described in terms of the inter-failure times. Define \( t_0 = 0 \) for consistency; then the inter-failure times are
\[
x_j = t_j - t_{j-1}.
\]

The types of repair considered here fall into three classes: a perfect repair which returns a system to its new state, a minimal repair which returns the system to working order but in exactly the same state as just before the failure, and an imperfect repair which produces a working state that is better than immediately before the failure but not as good as the original state.

10. The nonhomogeneous Poisson process

This process has two evocative names: bad-as-old and minimal-repair (Ascher & Feingold 1984). The assumptions are that, after a failure, the system is instantaneously restored to the state in which it was immediately before the failure. That is, only sufficient work is done (minimal repair) to restore the system to working order; after repair, the system is as exactly as it was before the breakdown. The meaning of this assumption is that, if we look at a system immediately after a breakdown at time \( s \), then we cannot distinguish it from a system of the same type which has had no breakdown up to time \( s \); all breakdowns can be regarded as the first breakdown after a period \( s \) with no breakdown. The survival function for the time to the first breakdown is \( R(\cdot) \). Thus the conditional survival function for the waiting time \( w \) to the next breakdown can immediately be written as
\[
R(\cdot|w) = R(s+w)/R(s),
\]
and it is easy to show that the intensity for the process is just the current value of the hazard rate for the time to first failure regarded as a function and evaluated at the current time. A standard argument (Cox & Lewis 1978) shows that the process so defined is a nonhomogeneous Poisson process (NHPP).

The NHPP is widely used in reliability growth modelling (the Duane model, Ascher & Feingold 1984) and in software reliability growth modelling (Miller & Keilier 1991). The strength of the model is that it is simple and relatively flexible, and the weakness is that it tends to be used as a black-box approach to data from repairable systems. It can provide a useful way to detect trends and dependence in data. The case of a constant intensity \( \lambda \) is a standard Poisson process, and many results from the standard Poisson process can be used by noting that, in the time scale \( t = \Lambda(t) \), where \( \Lambda(\cdot) \) is the mean-value function, the process becomes a standard Poisson process with unit rate. Time dependency is revealed through a nonconstant \( \lambda \).

Conversely, an NHPP with intensity \( \lambda \) defines a bad-as-old or minimal-repair model. The important distinguishing feature is that the intensity \( \lambda \) is a continuous function of time (Thompson 1981).

The scope for the use of covariates is somewhat limited. The problem is that \( \lambda \) is a continuous function; for non-time-varying covariates, the effect applies once and for all at the start of the process. When restricted to time-independent covariates, it is impossible to model the effect of changes in a system (for example, maintenance policy or improved components) during its lifetime. Continuously time-varying covariates can, however, appear simply in the rate function. A once-and-for-all effect can be determined by comparing the rate of occurrence of failures (ROCOF) under different policies. What the ROCOF can tell us is whether the system performance changes over time, and this aspect is widely used in modelling reliability growth and software reliability. Moreover, for this system, the expected number of failures in a fixed interval is easy to calculate.

Both accelerated-failure-time and proportional-hazards versions are easy to construct using a baseline intensity \( \lambda_0 \); the intensity functions are respectively
\[
\dot{\lambda}(t|z) = \frac{1}{\psi(z)} \lambda_0 \left( \frac{t}{\psi(z)} \right), \quad \dot{\lambda}(t|z) = \psi(z) \lambda_0(t).
\]

In view of the above remarks, it is clear that estimation of \( \psi \) requires data from a number of systems operating with different values of \( z \), and a single process history will not allow \( \psi \) to be estimated unless \( \lambda_0 \) is known a priori.

11. Imperfect repair

The imperfect-repair model is easy to describe. It is assumed that two sorts of repair can take place after a failure: the first kind is a perfect repair that returns the system to its original state; the second type is a minimal repair that restores the system to the working state, but only to the state just before the failure. We shall not examine this model in detail, but remark that the sequences of time of a perfect repair are regeneration points for the process, and the process of intervals between imperfect repairs is a renewal process (Brown & Proschan 1983). The techniques applied to
renewal and modulated renewal processes can be borrowed to determine many of
the properties of this model. The treatment is analogous to the treatment of the
alternating renewal process (Whittaker & Samaniego 1989).

12. Additive hazards

This approach arises naturally from the desire to model a system that, after a repair,
is better than it was just before the repair, but not as good as new. Thus, suppose
that a new system has interval hazard rate \( h_0(\cdot) \); then, after the \((j - 1)\)th failure, the
interval hazard rate is \( \psi_j + h_0(\cdot) \). The attraction of this model is that it appears as a
simple additive analogue of the proportional-hazards approach, and that each
failure contributes something to the age of the system. Moreover, this model offers
hazard rates that are not zero at time zero. However, in view of the problems of
estimation described above, it can only be used in a phenomenological way to
measure the magnitude of the jumps \( \psi \) in the hazard rate, and models for the \( \psi_j \)
which make use of explanatory variables are unlikely to produce satisfactory estimators of the parameters (Pijnenburg 1991).

13. Virtual age

Here, after a repair, the system returns to working order, and the interval hazard
rate is that of a system which is not new but is better than a system which has
undergone a minimal repair. This can be modeled by assuming that, after \( j \)th repair,
the system begins to operate as a system with an age \( \tau_j \). The parameter \( \tau_j \) is the
virtual age after a repair. This is also an approximation to Downton’s model for the
analysis of data from a nonhomogeneous Poisson process (Downton 1969) in which
several copies of the process are observed, but where the failure times and the
identities of the processes are not recorded—only the inter-failure times are known.
The reliability function for the interval is

\[
R_j(x) = R_0(x + \tau_j)/R_0(\tau_j).
\]

The hazard rate is

\[
h_j(x) = h_0(x + \tau_j).
\]

The virtual-age model also yields a measure of the relative survival chance at the
beginning of each interval through \( p_j = R_0^{-1}(\tau_j) \). Thus it can be seen that, if \( \tau_j = 0 \)
for all \( j \), then the system is a renewal process; if \( \tau_j = \tau_j \) for all \( j \), then the system
becomes an NHPP. This makes the model attractive in that it lies naturally between
the NHPP and the renewal process. The model can be extended in two ways using
proportional-hazards ideas or accelerated-failure-time ideas. The proportional-
hazards version is

\[
h_j(x) = \psi_j h_0(x + \psi_j).
\]

With this proportional-hazards model, the presence of the \( \tau_j \) means that a parametric
likelihood function must be used to estimate the parameters \( \psi \) and \( \tau \) of the functions;

the accelerated-failure-time version has a hazard rate of the form

\[
h_j(x) = \frac{1}{\psi_j} h_0\left(\frac{x + \tau_j}{\psi_j}\right).
\]

Later the accelerated-failure-time version will be illustrated. The parameter \( \tau \) is a
measure of accumulated aging, and the parameter \( \psi \) is a measure of the rate at which
aging occurs. In its most general form, the parameters \( \psi \) and \( \tau \) are assumed to be
functions of explanatory variables, for instance a simple trend model would have \( \psi \)
constant and \( \tau = \tau(\psi + \beta) \) or vice versa. However, just as with the additive hazards
model, there are—as yet unresolved—identification problems.

This model also solves a trivial-seeming difficulty with hazard rates. For the most
commonly occurring distributions, apart from the exponential, the hazard rate and
density at time zero are either zero or infinite. On examining data from repairable
systems, the interval density almost always appears to be finite but nonzero at time
zero. Noting that the \( j \)th interval density is

\[
f_j(x) = f_0(x + \tau_j)/R_0(\tau_j)
\]

shows that

\[
f_j(0) = f_0(\tau_j)/R_0(\tau_j) = h_j(0) = h_0(\tau_j).
\]

14. Examples

Software reliability

The two most common forms of model used in software reliability are versions of the
NHPP (Goel & Okumoto 1979) and mixture models based on order statistics
(Littlewood 1991). In this example, a family of NHPP models based on a mean-value
function is used to analyse a data set from Musa. The distinguishing feature of many
analyses of software failure is that it is almost always possible to obtain a good fit
to the data, but the forecasting power of the models is poor. The class of models
used is described by Al-Ayyoubi et al. (1990) and Miller et al. (1991). The mean value
for a software failure process is bounded because there is a finite initial number of
faults; indeed the process generated is indistinguishable from an order statistics
process. If the original number of faults follows a Poisson distribution with mean \( \alpha \),
and the time taken to detect a fault is distributed as \( F(t) \), then the observed failure
times come from a thinned Poisson process with mean-value function

\[
E[N] = \alpha F(t).
\]

This model also delivers a process with a bounded mean-value function. In this case,
the accelerated-failure-time approach is natural, and the scale parameter of the
distribution \( F \) is then a measure of the relative performance of the debugging process
for a particular piece of software.

Using the scale-parameter model, the mean-value function is

\[
\Lambda(t) = \mu(t) = \alpha F(t/b; \beta),
\]
and the intensity function is

$$\lambda(t) = \frac{d\mu}{dt} = \frac{x}{b} f(t; \beta).$$

Since the intensity is a density function, it must eventually decline to zero, suggesting that these models represent a situation in which the intervals between faults are increasing but that the expected time to the discovery of the last fault is infinite.

The following familiar models are recovered by particular choices of distribution:

For the Littlewood model, based on the Pareto distribution,

$$\lambda(t) = x \left(1 - \frac{1}{(1 + t/b)^\alpha}\right);$$

for the Goel and Goel–Okumuto models, using the Weibull distribution,

$$\lambda(t) = x(1 - e^{-\alpha t^\beta});$$

and the extreme-value distribution gives

$$\lambda(t) = x e^{-\alpha t^\beta}.$$

The likelihood in this case, for grouped data \((m_i, t_i)\) with \(m_i\) failures in the interval \([t_{i-1}, t_i)\), is

$$L(x, \beta, b) = x^n \exp[-xF(u^*)] \prod_{i=1}^{n} \left[F(u_i) - F(u_{i-1})\right]^{m_i},$$

where \(n = \sum_{i=1}^{n} m_i\), with \(u_i = t_i/b\) and \(u^* = t^*/b\) the end of the observation period. In all cases the estimator for \(x\) is

$$\hat{x} = n/F(u^*),$$

which can be used to give a profile likelihood

$$L^*(\beta, b) = \prod_{i=1}^{n} \left(\frac{F(u_i) - F(u_{i-1})}{F(u^*)}\right)^{m_i}.$$

The profile likelihood was maximized using NAG routine E04KBF, and the results were rather unstable. The model is illustrated using a data set from Musa. What is striking about the likelihood is that there are multiple maxima and that the estimators are highly correlated. Although the model can yield a good fit to the data, the interpretation requires some care. The results are summarized in Table 1. The contour map of the profile likelihood for the Pareto-based model is remarkable in that it exhibits multiple maxima along a ridge at about 45° (Al-Ayyoubi et al. 1990). This ridge also shows the high correlation between the shape and scale parameters in agreement with the estimated correlation in Table 1. The extreme-value model also gives a likelihood which indicates a second maximum for the log-likelihood function. The predictions from the three models also vary widely, but all the models fit the data satisfactorily as measured by the Kolmogorov–Smirnov statistic, which does not reject the hypothesis that the models fit at the 5% level. Miller et al. (1991) included these models in a 'super model' and showed that, after different lengths of time, different simple models were chosen on the basis of fit. What can be seen from Table 1 is that there is again an identification problem—in the sense that the likelihood values are all more or less the same and that, in this particular case, the fits are also more or less the same. The 'super model' fails to distinguish between the simple models on the basis of likelihood. Similar results are reported by Etzadz- Amoli & Ciampi (1987) in a study of a hazard regression model based on spline approximations; their model included proportional-hazards and accelerated-failure-time models as special cases. They also report 'ridges' in the likelihood and a nearly singular Hessian matrix as indicative of an identification problem in which some parameters must be arbitrarily assigned to allow estimation of all the parameters of interest.

McCollin et al. (1989) discuss more fully the use of explanatory variables so analyse the times between failure for a number of pieces of software. They indicate that the most widely used approach is the proportional-hazards one, in combination with one of the other basic models described in this paper. Their conclusions were somewhat tentative but indicated that only two rather simple covariates could be used in modelling: the age of the software, and the sequence number of the fault. Although they fitted proportional-hazards models in some cases, the fits were always marginal. With these two covariates, the model reduces directly to one of the kind discussed above.

The investigation of van Pul (1982a,b) into the likelihoods for the Musa and Littlewood models demonstrates that the number of observations needed to attain asymptotic normality in the maximum-likelihood estimators is extremely large, of the order of thousands of observations. Moreover, not all samples yielded acceptable estimators of the parameters of interest. The problem appears to be twofold: firstly there is an identification problem in that the models cannot distinguish between (i) a system having a large number of faults with a small rate of occurrence per fault and (ii) a system with few faults but with a high rate of occurrence per fault (Wright & Hazelhurst 1988); and secondly the likelihoods themselves are ill-conditioned.

**Table 1**

<table>
<thead>
<tr>
<th></th>
<th>Pareto</th>
<th>Weibull</th>
<th>Extreme-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log-likelihood</td>
<td>-143.7</td>
<td>-143.9</td>
<td>-143.8</td>
</tr>
<tr>
<td>(\hat{x})</td>
<td>377.5</td>
<td>160.1</td>
<td>502.7</td>
</tr>
<tr>
<td>(\hat{\beta})</td>
<td>1.9</td>
<td>10.7</td>
<td>72.5</td>
</tr>
<tr>
<td>(\hat{b})</td>
<td>0.2</td>
<td>0.7</td>
<td>0.2</td>
</tr>
<tr>
<td>s.d. (\hat{x})</td>
<td>346.6</td>
<td>21.0</td>
<td>197.6</td>
</tr>
<tr>
<td>s.d. (\hat{\beta})</td>
<td>1.2</td>
<td>3.5</td>
<td>100.7</td>
</tr>
<tr>
<td>s.d. (\hat{b})</td>
<td>0.2</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>\text{corr}(\hat{x}, \hat{\beta})</td>
<td>-0.8</td>
<td>0.7</td>
<td>0.97</td>
</tr>
<tr>
<td>\text{corr}(\hat{\beta}, \hat{b})</td>
<td>-0.9</td>
<td>-0.6</td>
<td>-0.9</td>
</tr>
<tr>
<td>\text{corr}(\hat{b}, \hat{\beta})</td>
<td>0.9</td>
<td>-0.7</td>
<td>-0.96</td>
</tr>
<tr>
<td>Kolmogorov–Smirnov statistic</td>
<td>0.1889</td>
<td>0.2004</td>
<td>0.1999</td>
</tr>
</tbody>
</table>
A mechanical system

Downton (1969) studied the failure pattern of a fleet of buses, from data originally published by Davis (1952); see Table 2. Indeed, until now, this data seems to have resisted statistical analysis. Pijnenburg (1991) showed that, on the basis of a graphical analysis, an additive-hazards model seemed plausible with the sequence number of the failure as an explanatory variable. Pijnenburg failed to find estimators for the additive-hazards model

\[ h_j(x) = (\beta_1 + \beta_x j) + h_0(x). \]

The reasons for the lack of estimators are now clear from the above results.

In the light of the difficulties with the additive-hazards model, a virtual-age model seems a reasonable alternative. The model is formulated precisely as in the accelerated-failure-time version above with a Weibull as the underlying distribution. The model

\[ \phi_j(x) = \phi_{\alpha} + \phi_{\beta} \log(x). \]

### Table 2

**Downton's data**

<table>
<thead>
<tr>
<th>Distance (1000 miles)</th>
<th>( X_1 )</th>
<th>( X_2 )</th>
<th>( X_3 )</th>
<th>( X_4 )</th>
<th>( X_5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–10</td>
<td>3</td>
<td>8</td>
<td>12</td>
<td>15</td>
<td>19</td>
</tr>
<tr>
<td>10–20</td>
<td>3</td>
<td>11</td>
<td>15</td>
<td>19</td>
<td>15</td>
</tr>
<tr>
<td>20–30</td>
<td>6</td>
<td>9</td>
<td>13</td>
<td>9</td>
<td>17</td>
</tr>
<tr>
<td>30–40</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>11</td>
<td>15</td>
</tr>
<tr>
<td>40–50</td>
<td>6</td>
<td>8</td>
<td>9</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>50–60</td>
<td>9</td>
<td>9</td>
<td>6</td>
<td>13</td>
<td>6</td>
</tr>
<tr>
<td>60–70</td>
<td>8</td>
<td>9</td>
<td>6</td>
<td>12</td>
<td>4</td>
</tr>
<tr>
<td>70–80</td>
<td>16</td>
<td>9</td>
<td>7</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>80–90</td>
<td>14</td>
<td>7</td>
<td>5</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>90–100</td>
<td>20</td>
<td>8</td>
<td>6</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>100–110</td>
<td>25</td>
<td>2</td>
<td>9</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>110–120</td>
<td>21</td>
<td>6</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>120–130</td>
<td>23</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>130–140</td>
<td>10</td>
<td>4</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>140–150</td>
<td>11</td>
<td>2</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>150–160</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>160–170</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>170–180</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>180–190</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>190–200</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>200–210</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>210–220</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 3**

**Estimated parameters and likelihoods**

<table>
<thead>
<tr>
<th>( \phi )</th>
<th>( \psi )</th>
<th>( \phi_j )</th>
<th>( \psi_j )</th>
<th>( \phi_j )</th>
<th>( \psi_j )</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.70</td>
<td>0.72</td>
<td>0.63</td>
<td>0.72</td>
<td>0.63</td>
<td>0.72</td>
</tr>
<tr>
<td>5.04</td>
<td>0.74</td>
<td>0.78</td>
<td>0.74</td>
<td>0.78</td>
<td>0.74</td>
</tr>
<tr>
<td>5.50</td>
<td>0.76</td>
<td>0.80</td>
<td>0.76</td>
<td>0.80</td>
<td>0.76</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Period</th>
<th>( \phi_j )</th>
<th>( \psi_j )</th>
<th>( \phi_j )</th>
<th>( \psi_j )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>173.65</td>
<td>063.47</td>
<td>179.32</td>
<td>068.10</td>
</tr>
<tr>
<td>2</td>
<td>319.85</td>
<td>260.52</td>
<td>258.59</td>
<td>198.58</td>
</tr>
<tr>
<td>3</td>
<td>352.57</td>
<td>340.47</td>
<td>370.03</td>
<td>355.61</td>
</tr>
<tr>
<td>4</td>
<td>283.19</td>
<td>278.20</td>
<td>401.38</td>
<td>435.58</td>
</tr>
<tr>
<td>5</td>
<td>289.69</td>
<td>314.45</td>
<td>308.52</td>
<td>331.89</td>
</tr>
</tbody>
</table>

Log-likelihood: -1506.54, -1513.73, -1506.33

**Fig. 1.** Histograms and fitted density functions.
is given by

\[ F_0(x | z) = 1 - e^{-x^z}, \quad f_0(x | z) = xz^{-1} e^{-x^z}. \]

As with many other data sets no covariates were reported and so the simplest version of the model was taken with interval hazard rate

\[ h_0(x) = \frac{1}{\psi_j} h_0 \left( \frac{x + \tau_i}{\psi_j} \right). \]

A grouped likelihood was used. The likelihood function appears to have multiple maxima, and the two parameters \( \tau_i \) and \( \psi_j \) are also highly correlated. The likelihood was maximized using the PC-MATLAB package (Grace 1990). Two different sets of parameters produced equal values of the likelihood, but only one of these sets was not rejected on the basis of chi-square tests. Thus there are, once more, indications problems something not usually discussed in reliability analysis—and the unsatisfactory behaviour of the likelihood function and estimators. Remarkably, the underlying process and may even seem unlikely in certain situations.

**14. Summary**

The discussion and examples given here show that, although the parametric models can be fitted to failure data from repairable systems, there are frequently identification problems—something not usually discussed in reliability analysis—and the unsatisfactory behaviour of the likelihood function and estimators. Remarkably, the conclusions seem to be that data analysis based on nonhomogeneous Poisson processes or proportional hazards is likely to yield most of the information available in the data, even though these models do not necessarily truly represent the underlying process and may even seem unlikely in certain situations. In particular, the proportional-hazards model appears very robust under these circumstances.

**REFERENCES**


