

1

Maintenance of Repairable Systems

Bo Henry Lindqvist

1.1 Introduction

A commonly used definition of a repairable system (Ascher and Feingold [3]) states that this is a system which, after failing to perform one or more of its functions satisfactorily, can be restored to fully satisfactory performance by any method other than replacement of the entire system. In order to cover more realistic applications, and to cover much recent literature on the subject, we need to extend this definition to include the possibility of additional maintenance actions which aim at servicing the system for better performance. This is referred to as preventive maintenance (PM), where one may further distinguish between condition based PM and planned PM. The former type of maintenance is due when the system exhibits inferior performance while the latter is performed at predetermined points in time.

Traditionally, the literature on repairable systems is concerned with modelling of the failure times only, using point process theory. A classical reference here is Ascher and Feingold [3]. The most commonly used models for the failure process of a repairable system are renewal processes (RP), including the homogeneous Poisson processes (HPP), and nonhomogeneous Poisson processes (NHPP). While such models often are sufficient for simple reliability studies, the need for more complex models is clear. In this chapter we consider some generalizations and extensions of the basic models, with the aim to arrive at more realistic models which give better fit to data. First we consider the Trend Renewal Process (TRP) introduced and studied in Lindqvist, Elvebakk and Heggland [22]. The TRP includes NHPP and RP as special cases, and the main new feature is to allow a trend in processes of non-Poisson (renewal) type.

As exemplified by some real data, in the case where several systems of the same kind are considered, there may be unobserved heterogeneity between the systems which, if overlooked, may lead to non-optimal or possibly completely wrong decisions. We will consider this in the framework of the TRP process, which in Lindqvist et al. [22] is extended to the so called HTRP model which includes the possibility of heterogeneity. Heterogeneity can be thought of as an effect of an unobserved covariate.

Another extension of the basic models is to allow the systems to be preventively maintained. We review some recent research in this direction, where this

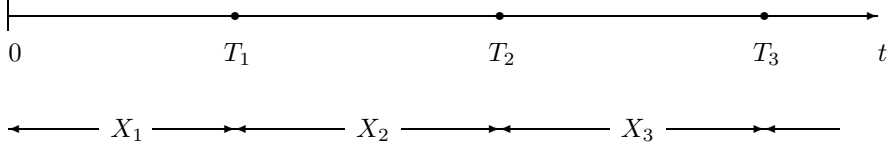


Figure 1: Event times (T_i) and sojourn times (X_i) of a repairable system.

situation is modelled as a competing risks problem between failure and PM. This leads to a need for combining the theory of competing risks with repair models and point process theory. Relevant statistical data for such analyses are found in most modern reliability databases. The book by Bedford and Cooke [4] contains a chapter related to this. A general reference to competing risks is the book by Crowder [10].

The last extension of the basic model considered in the present chapter, consists in using Markov models to model the behavior of periodically inspected systems in between inspections, with the use of separate Markov models for the maintenance tasks at inspections.

Recent review articles concerning repairable systems and maintenance include Pena [27] and Lindqvist [20]. A review of methods for analysis of recurrent events with a medical bias is given by Cook and Lawless [7]. General books on statistical models and methods in reliability, covering much of the topics considered here, are Meeker and Escobar [25] and Rausand and Høyland [30].

1.2 Point Process Approach

1.2.1 Notation and Basic Definitions

Consider a repairable system where time usually runs from $t = 0$ and events occur at ordered times T_1, T_2, \dots . Here time is not necessarily calendar time, but can be for example operation time, number of cycles, number of kilometers run, length of a crack etc. In the present treatment we shall disregard time durations of repair and maintenance, and assume that the system is always restarted immediately after failure or maintenance action. The inter-event, or inter-failure, times will be denoted X_1, X_2, \dots . Here $X_i = T_i - T_{i-1}$, $i = 1, 2, \dots$, where for convenience we define $T_0 \equiv 0$. Figure 1 illustrates the notation. We also make use of the counting process representation $N(t) = \text{number of events in } (0, t]$.

In order to describe probability models for repairable systems we use some notation from the theory of point processes. A key reference is Andersen, Borgan, Gill and Keiding [2]. Let \mathcal{H}_t denote the history of the failure process up to, but not including, time t .

The conditional intensity of the process at time t is now defined as

$$\gamma(t) = \lim_{\Delta t \downarrow 0} \frac{Pr(\text{event of type } j \text{ in } [t, t + \Delta t) | \mathcal{H}_t)}{\Delta t}, \quad (1)$$

For statistical inference we need an expression for the likelihood function. Suppose that a single system as described above is observed from time 0 to time τ , resulting in observations $T_1, T_2, \dots, T_{N(\tau)}$. The likelihood function is then given by ([2, Section II.7]),

$$L = \left\{ \prod_{i=1}^{N(\tau)} \gamma(T_i) \right\} \exp \left\{ - \int_0^\tau \gamma(u) du \right\}. \quad (2)$$

1.2.2 Perfect and Minimal Repair Models

Consider a system or component with failure rate $z(t)$. Suppose first that after each failure, the system is repaired to a condition as good as new. In this case the failure process can be modelled by a renewal process with inter-event time distribution F , denoted $RP(F)$. Clearly, the conditional intensity defined in (1) is given by

$$\gamma(t) = z(t - T_{N(t-)})$$

where $t - T_{N(t-)}$ is the time since the last failure strictly before time t .

Suppose instead that after a failure, the system is repaired only to the state it had immediately before the failure, called a minimal repair. This means that the conditional intensity of the failure process immediately after the failure is the same as it was immediately before the failure, and hence is exactly as it would be if no failure had ever occurred. Thus we must have

$$\gamma(t) = z(t)$$

and the process is an NHPP with intensity $z(t)$, denoted $NHPP(z(\cdot))$. In practice a minimal repair usually corresponds to repairing or replacing only a minor part of the system.

If $z(t) = \lambda$ does not depend on t , then $NHPP(z(\cdot))$ is a homogeneous Poisson process which we denote by $HPP(\lambda)$. Note that an HPP is at the same time an RP with exponential inter-failure times.

1.2.3 The Trend-Renewal Process

The idea behind the trend-renewal process is to generalize the following well known property of the NHPP. First let the cumulative intensity function corresponding to an intensity $\lambda(\cdot)$ be defined by $\Lambda(t) = \int_0^t \lambda(u) du$. Then if T_1, T_2, \dots is an $NHPP(\lambda(\cdot))$, the time-transformed stochastic process $\Lambda(T_1), \Lambda(T_2), \dots$ is $HPP(1)$.

The trend-renewal process (TRP) is defined simply by allowing the above $HPP(1)$ to be any renewal process $RP(F)$. Thus, in addition to the intensity

function $\lambda(t)$, for a TRP we need to specify a distribution function F of the inter-arrival times of this renewal process. Formally we can define the process $\text{TRP}(F, \lambda(\cdot))$ as follows:

Let $\lambda(t)$ be a nonnegative function defined for $t \geq 0$, and let $\Lambda(t) = \int_0^t \lambda(u) du$. The process T_1, T_2, \dots is called $\text{TRP}(F, \lambda(\cdot))$ if the transformed process $\Lambda(T_1), \Lambda(T_2), \dots$ is $\text{RP}(F)$, that is if the $\Lambda(T_i) - \Lambda(T_{i-1})$; $i = 1, 2, \dots$ are i.i.d. with distribution function F . The function $\lambda(\cdot)$ is called the trend function, while F is called the renewal distribution. In order to have uniqueness of the model it is usually assumed that F has expected value 1.

Figure 2 illustrates the definition. For the cited property of the NHPP, the lower axis would be an HPP with unit intensity, $\text{HPP}(1)$. For the TRP, this process is instead taken to be any renewal process, $\text{RP}(F)$, where F has expectation 1. This shows that the TRP includes the NHPP as a special case. Further, if $\lambda(t) \equiv 1$ is the constant 1, then $\Lambda(T_i) = T_i$, and so T_1, T_2, \dots is $\text{RP}(F)$.

For an $\text{NHPP}(\lambda(\cdot))$, the $\text{RP}(F)$ would be $\text{HPP}(1)$. Thus $\text{TRP}(1 - e^{-x}, \lambda(\cdot)) = \text{NHPP}(\lambda(\cdot))$. Also, $\text{TRP}(F, 1) = \text{RP}(F)$, which shows that the TRP class includes both the RP and NHPP classes.

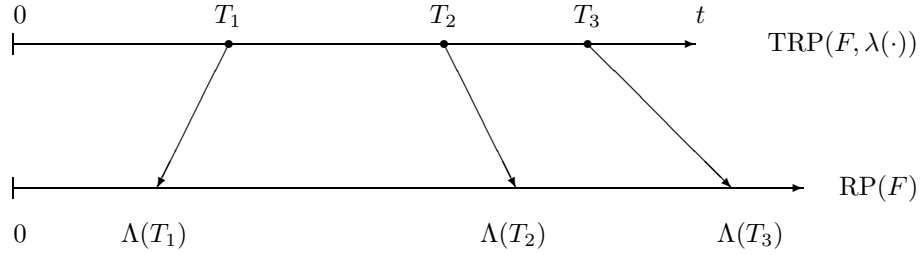


Figure 2: The defining property of the trend-renewal process.

It can be shown [22] that the conditional intensity function, given the history \mathcal{H}_t , for the $\text{TRP}(F, \lambda(\cdot))$ is

$$\gamma(t) = z(\Lambda(t) - \Lambda(T_{N(t-)}))\lambda(t) \quad (3)$$

where $z(\cdot)$ is the hazard rate corresponding to F . This is a product of one factor, $\lambda(t)$, which depends on the age t of the system and one factor which depends on a transformed time from the last previous failure.

Suppose now that a single system has been observed in $[0, \tau]$, with failures at $T_1, T_2, \dots, T_{N(\tau)}$. If a $\text{TRP}(F, \lambda(\cdot))$ is used as a model, then substitution of (3) into (2) gives the likelihood

$$L = \left\{ \prod_{i=1}^{N(\tau)} z[\Lambda(T_i) - \Lambda(T_{i-1})]\lambda(T_i) \right\} \exp\left\{ - \int_0^{\tau} z[\Lambda(u) - \Lambda(T_{N(u-)})]\lambda(u) du \right\}. \quad (4)$$

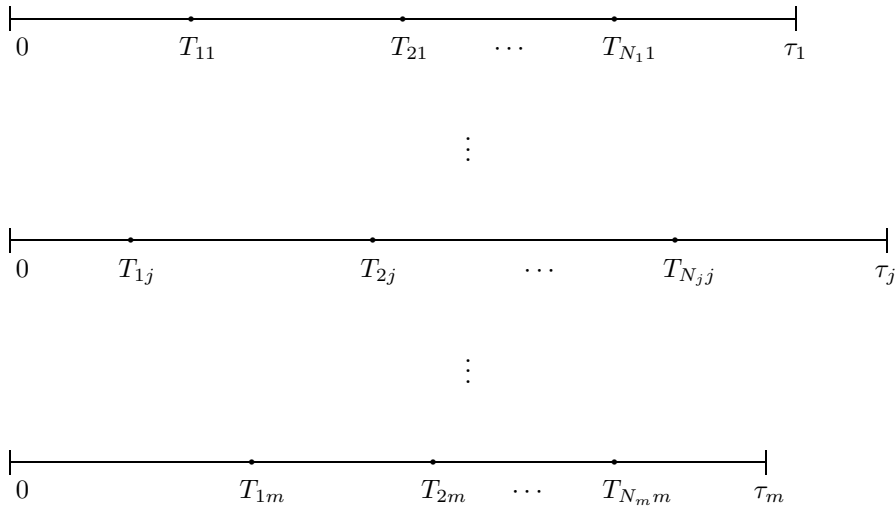


Figure 3: Observation of failure times of m systems. The j th system is observed over the time interval $[0, \tau_j]$, with $N_j \geq 0$ observed failures.

For the NHPP($\lambda(\cdot)$) we have $z(t) \equiv 1$, so the likelihood simplifies to the well known expression (Crowder et al. [12])

$$L = \left\{ \prod_{i=1}^{N(\tau)} \lambda(T_i) \right\} \exp\left\{-\int_0^{\tau} \lambda(u) du\right\}.$$

Returning to the general case, if f is the density function corresponding to F , then we can write the likelihood (4) as

$$L = \left\{ \prod_{i=1}^{N(\tau)} f[\Lambda(T_i) - \Lambda(T_{i-1})] \lambda(T_i) \right\} \{1 - F[\Lambda(\tau) - \Lambda(T_{N(\tau)})]\}. \quad (5)$$

This latter form of the likelihood of the TRP follows directly from the definition, since the conditional density of T_i given $T_1 = t_1, \dots, T_{i-1} = t_{i-1}$ is $f[\Lambda(t_i) - \Lambda(t_{i-1})] \lambda(t_i)$, and the probability of no failures in the time interval $(T_{N(\tau)}, \tau]$, given $T_1, \dots, T_{N(\tau)}$, is $1 - F[\Lambda(\tau) - \Lambda(T_{N(\tau)})]$.

This again simplifies if $\lambda(t) \equiv 1$ in which case it gives the likelihood of an RP(F) observed on $[0, \tau]$.

1.2.4 Observations from Several Similar Systems

Suppose that m systems of the same kind are observed, where the j th system ($j = 1, 2, \dots, m$) is observed in the time interval $[0, \tau_j]$. For the j th system, let N_j denote the number of failures that occur during the observation period,

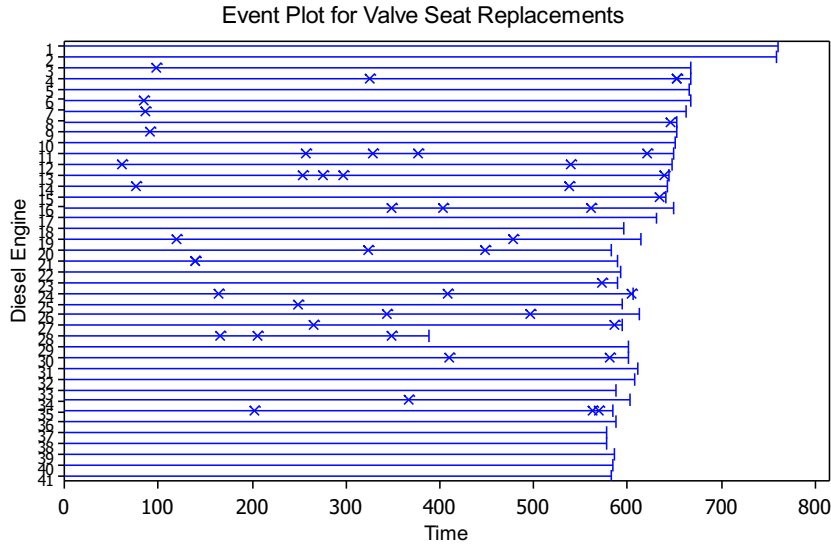


Figure 4: Event plot for times of valve seat replacements for 41 diesel engines, taken from Nelson [26].

and let the specific failure times be denoted $T_{1j} < T_{2j} < \dots < T_{N_j j}$. Fig. 3 illustrates the notation and explains the information given in a so-called event plot which is provided by computer packages for analysis of this kind of data (see examples below).

Example 1. Nelson [26] presented data for the data on times of valve-seat replacements in a fleet of $m = 41$ diesel engines. Figure 4 shows an event plot of the complete dataset.

Example 2. Bhattacharjee et al. [5] presented failure data for motor operated closing valves in safety systems at two boiling water reactor plants in Finland. Failures of the type “External Leakage” were considered for 104 valves with a follow-up time of 9 years. An event plot for the 16 valves which experienced at least one failure, is given in Figure 5. The remaining 88 valves had no failures.

When data are available for m systems as described above, one will typically assume that the systems behave independently but with the same probability laws (“i.i.d. rules”). The total likelihood for the data will then be the product of likelihoods (4) or (5), one factor for each of the m systems.

However, even if the m systems are considered to be of the same type, they may well exhibit different probability failure mechanisms. For example, systems may be used under varying environmental or operational conditions.

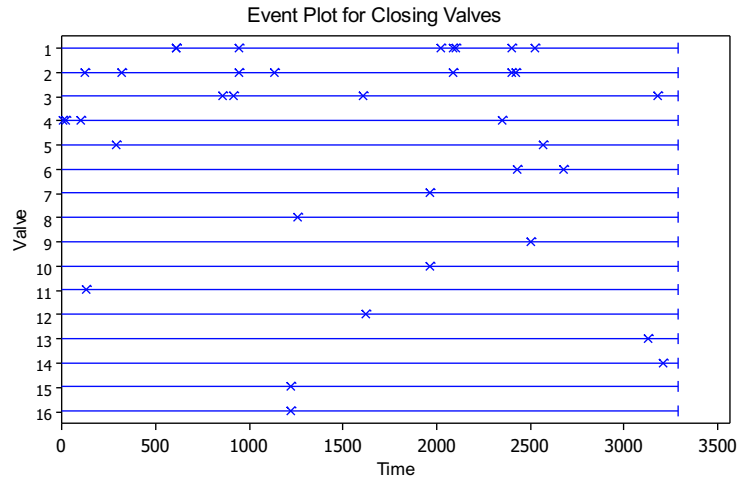


Figure 5: Event plot for times of external leakage from nuclear plant valves, taken from Bhattacharjee et al. [5]. In addition, 88 valves had no failures in 3286 days (9 years).

To cover such cases we shall assume that failures of the j th system follow the process TRP($F, \lambda_j(\cdot)$), $j = 1, \dots, m$, where the renewal function F is fixed and differences between systems are modelled by letting the trend functions $\lambda_j(t)$ vary from system to system. The assumption of a fixed F parallels the NHPP case, where F is the unit exponential distribution.

Assuming that systems work independently of each other, we obtain from (5) the full likelihood $L \equiv \prod_{j=1}^m L_j$ where

$$L_j = \left\{ \prod_{i=1}^{N_j} f[\Lambda_j(T_{ij}) - \Lambda_j(T_{i-1,j})] \lambda_j(T_{ij}) \right\} \{1 - F[\Lambda_j(\tau_j) - \Lambda_j(T_{N_j j})]\}. \quad (6)$$

As an example of the use of (6), assume that differences between system performances can be attributed to an observable covariate vector \mathbf{x} , and that the trend $\lambda_j(t)$ for system j is represented by a proportional trend model with

$$\lambda_j(t) = g(\mathbf{x}_j) \lambda(t), \quad j = 1, \dots, m \quad (7)$$

Here $\lambda(\cdot)$ is a basic trend function common to all systems, while g is a function of the covariate vector \mathbf{x}_j of system j . The special cases of this model corresponding to NHPP and RP are studied, respectively, by Lawless [19], and Follmann and Goldberg [14].

1.2.5 The Heterogeneous Trend Renewal Process

As noted in the introduction, in addition to observable differences there may be an unobserved heterogeneity between systems. A common way of incorporating such heterogeneity is to modify (7) to $\lambda_j(t) = a_j g(\mathbf{x}_j) \lambda(t)$ where the a_j are unobservable (positive) random variables taking values independently across systems (Andersen et al. 1993, Chapter IX).

For simplicity we shall in this chapter restrict attention to the case with no observed covariates, and instead concentrate on unobserved heterogeneity. In the following we thus assume the model

$$\lambda_j(t) = a_j \lambda(t) \quad (8)$$

where the a_j are independently distributed according to a common probability distribution H , say, and where for convenience we assume that the expected value of a_j equals 1. Thus in (8), $\lambda(\cdot)$ is regarded as a basic trend function, while the a_j represent a possibly different failure intensity “level” for each system, averaging to 1. The special case when $a_j = 1$ with probability 1 will be referred to as the “no heterogeneity” case.

For given values of the a_j the likelihood for the j th system is, by (6),

$$L_j(a_j) = \left\{ \prod_{i=1}^{N_j} f[a_j(\Lambda(T_{ij}) - \Lambda(T_{i-1,j}))] a_j \lambda(T_{ij}) \right\} \{1 - F[a_j(\Lambda(\tau_j) - \Lambda(T_{N_j j}))]\}$$

However, since the a_j are unobservable, we need to take the expectation with respect to the a_j , giving

$$L_j = E[L_j(a_j)] = \int L_j(a_j) dH(a_j)$$

as the contribution to the likelihood from the j th system. The total likelihood is then the product

$$L = \prod_{j=1}^m L_j \quad (9)$$

We shall use the notation $\text{HTRP}(F, \lambda(\cdot), H)$ for the model with likelihood (9). Here the renewal distribution F and the heterogeneity distribution H are distributions corresponding to positive random variables with expected value 1, while the basic trend function $\lambda(t)$ is a positive function defined for $t \geq 0$.

A useful feature of the HTRP model is that several important models for repairable systems are easily represented as submodels. With the notation HPP, NHPP, RP and TRP used as before, we define corresponding models with heterogeneity (8) by putting an H in front of the abbreviations. Specifically, from a full model, $\text{HTRP}(F, \lambda(\cdot), H)$, we can identify the seven submodels described in Table 1.

The HTRP and the seven submodels may also be represented in a cube, as illustrated in Figures 6 and 7. Each vertex of the cube represents a model, and

Submodel	HTRP-formulation
HPP(ν)	HTRP(exp, ν , 1)
RP(F, ν)	HTRP(F, ν , 1)
NHPP($\lambda(\cdot)$)	HTRP(exp, $\lambda(\cdot)$, 1)
TRP($F, \lambda(\cdot)$)	HTRP($F, \lambda(\cdot)$, 1)
HHPP(ν, H)	HTRP(exp, ν, H)
HRP(F, ν, H)	HTRP(F, ν, H)
HNHPP($\lambda(\cdot), H$)	HTRP(exp, $\lambda(\cdot), H$)

Table 1: The seven submodels of HTRP($F, \lambda(\cdot), H$). 'exp' means the unit exponential distribution, '1' means the distribution degenerate at 1. The third column contains references to work on the corresponding models or special cases of them.

the lines connecting them correspond to changing one of the three “coordinates” in the HTRP-notation. Going to the right corresponds to introducing a time trend, going upwards corresponds to entering a non-Poisson case, and going backwards (inwards) corresponds to introducing heterogeneity. In analyzing data by parametric HTRP models we shall see below how we use the cube to facilitate the presentation of maximum log-likelihood values for the different models in a convenient, visual manner. The log-likelihood cube was introduced in Lindqvist et al. [22].

Example 1 (continued). Figure 6 shows the log-likelihood cube of the valve-seat data. It should be noted that each arrow points in a direction where exactly one parameter is added (see text of Figure6 for definitions of parameters). Using standard asymptotic likelihood theory we know that if this parameter has no influence in the model, then twice the difference in log likelihood is approximately chi-square distributed with 1 degree of freedom. For example, if twice the difference is larger than 3.84, then the p-value of no significant difference is less than 5% and we have an indication that the extra parameter in fact has some relevance. Note that adding an extra parameter will always lead to a larger value of the maximum log likelihood, but from what we just argued, the difference needs to be more than, say, $3.84/2 = 1.92$ to be of real interest.

Looking at the valve-seat data cube we note first that going from a vertex of the front face to the corresponding vertex of the back face (adding “H” in front of the model acronym) there is never much to gain (1.17 at most from HPP to HHPP). This indicates no apparent heterogeneity between the various engines.

By comparing the left and right faces we conclude, however, that there seems to be a gain in including a time trend. Having already excluded heterogeneity we are thus faced with the possibilities of either NHPP or TRP. Here the latter model “wins”, since the difference in log-likelihood is as large as $(-343.66) - (-346.49) = 2.83$ and twice the difference equal to 5.66 corresponding to an approximate p-value of 0.017.

The resulting estimated TRP is seen to have a renewal distribution which is Weibull with shape parameter 0.6806 which implies a decreasing failure rate.

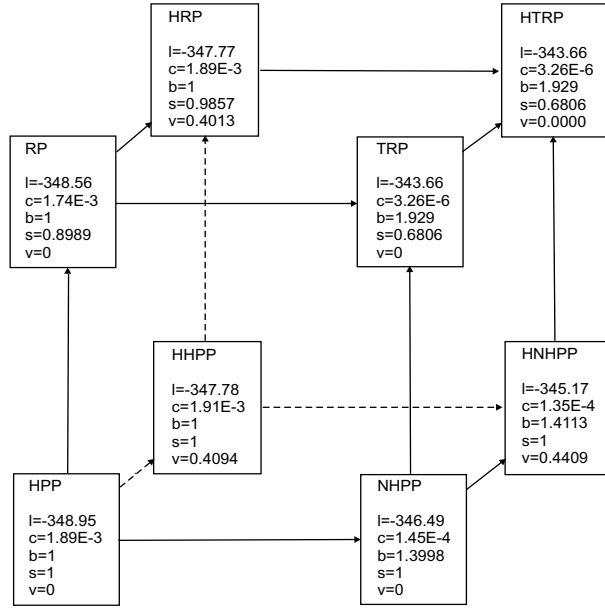


Figure 6: The log-likelihood cube for the Nelson valve seat data of [26], fitted with a parametric HTRP($F, \lambda(\cdot), H$) model and its sub-models. Here F is a Weibull-distribution with expected value 1 and shape parameter s , $\lambda(t) = cbt^{b-1}$ is a power function of t , and H is a gamma-distribution with expected value 1 and variance v . The maximum value of the log likelihood is denoted l .

This means that the conditional intensity function will jump upward at each failure, which may be explained by burn-in problems at each valve-seat replacement. Further, there will be an estimated time trend of the form $\hat{\lambda}(t) = 3.26 \cdot 10^{-6} \cdot 1.929 \cdot t^{0.929} = 6.29 \cdot 10^{-6} \cdot t^{0.929}$ which increases with t so that replacements are becoming more and more frequent.

Example 2 (continued). For the closing valve failures considered by Bhattacharjee et al. [5], previous studies had shown significant variations in the number of failures of each valve, suggesting a heterogeneity between valves. Bhattacharjee et al. [5] thus stressed the importance of taking heterogeneity into consideration and concluded that even very simple models may describe the heterogeneous behavior successfully. In particular they considered a model where heterogeneity was represented by assuming that each valve is either “good” or “bad”.

While Bhattacharjee et al. [5] used hierarchical Bayes-models, we fitted an HTRP model and its sub-models, with a trend function of power law type as for the valve-seat data, $\lambda(t) = cbt^{b-1}$, but now with a heterogeneity distribution H being a two-point distribution with values $a_1 = \text{“good”}$, $a_2 = \text{“bad”}$ (so $a_1 \leq a_2$ by assumption) and $P(\text{“good”}) = p$. In order to have uniqueness of parameters we imposed the restriction of expected value 1 for the distribution H , leading to

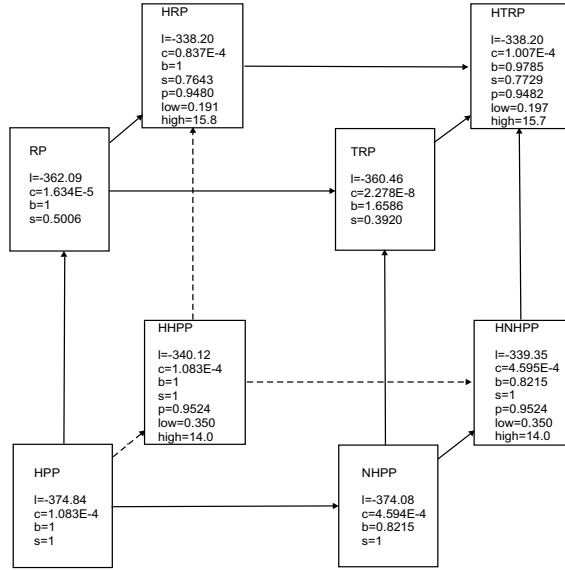


Figure 7: The log-likelihood cube for the data of Bhattacharjee et al. [5] concerning failures of motor operated closing valves in nuclear reactor plants in Finland, fitted with a parametric HTRP($F, \lambda(\cdot), H$) model and its sub-models. Here F is a Weibull-distribution with expected value 1 and shape parameter s , $\lambda(t) = cbt^{b-1}$ is a power function of t , and H is a two-point distribution with unit expectation, giving probability p for the value “low” and $1 - p$ for the value “high”. The maximum value of the log likelihood is denoted l .

$pa_1 + (1 - p)a_2 = 1$. The results are given in the log-likelihood cube of Figure 7.

By comparing the front and back faces of Figure 7 it is clear that there is a considerable heterogeneity present leaving us with the back face. Thus we continue by investigating whether we have Poisson-behavior or renewal-behavior at failures. This is done by comparing the bottom and top faces, in other words (HHPP, HNHPP) versus (HRP, HTRP). The difference from HHPP to HRP happens to be 1.92 so the p-value is 5%. Thus we might prefer the HRP model. However, in order to obtain a simple model with a simple interpretation we might go for the HHPP which gives that the closing valve is a “good” one with probability 0.9524, with failures following an HPP with rate

$$1.083 \cdot 10^{-4} \cdot 0.35 = 3.79 \cdot 10^{-5} \text{ (per day)}$$

or a “bad” one with probability 0.0476 and rate

$$1.083 \cdot 10^{-4} \cdot 14.0 = 1.52 \cdot 10^{-3} \text{ (per day.)}$$

The expected number of failures in 3286 days are hence 0.125 and 4.99, respectively.

1.3 A Competing Risks Model for Failure vs Preventive Maintenance

1.3.1 A General Setup

Consider again the situation illustrated in Figure 1, where the sojourns X_1, X_2, \dots are times to failure of a system which was repaired immediately before the start of the sojourn. In the present section we consider the case when the failure expected at the end of the sojourn X_i may be avoided by a preventive maintenance (PM) after a time Z_i in the sojourn. The experienced sojourn time will in this case be $Y_i = \min(X_i, Z_i)$, and it will result in either a failure or a PM according to whether $Y_i = X_i$ or $Y_i = Z_i$. We thus have a competing risks situation with two risks, corresponding to failure and PM.

Doyen and Gaudoin [13] recently presented a point process approach for modelling of such competing risks situations between failure and PM. A general setup for this kind of processes is furthermore suggested by Lindqvist in the review paper [20].

For simplicity we shall in this chapter consider only the case where the component or system is perfectly repaired or maintained at the end of each sojourn. This will lead to the observation of independent copies of the competing risks situation in the same way as for a renewal process. We will therefore in the following consider only a single sojourn and hence suppress the subscripts of the observed times. Thus we let X and Z be, respectively, the potential times to failure and time to PM of a single sojourn. Then $Y = \min(X, Z)$ is the observed sojourn, and in addition we observe the indicator variable δ which we define to be 1 if there is a PM ($Y = Z$) and 0 if there is a failure ($Y = X$). This situation has been extensively studied by Cooke [8, 9], Bedford and Cooke [4], Langseth and Lindqvist [17, 18], Lindqvist et al. [24], Lindqvist and Langseth [23].

Thus note that the observable result is the pair (Y, δ) , rather than the underlying times X and Z , which may often be the times of interest. For example, knowing the distribution of X would be important as a basis for maintenance optimization. It is well known (see Crowder, 2001, Ch. 7), however, that in a competing risks case as described here, the marginal distributions of X and Z are not identifiable from observation of (Y, δ) alone unless specific assumptions are made on the dependence between X and Z . The most used assumption of this kind is to let X and Z be independent, in which case identifiability follows. This assumption is not reasonable in our application, however, since the maintenance crew is likely to have some information regarding the system's state during operation. This insight is used to perform maintenance in order to avoid failures. We are thus in practice usually faced with a situation of dependent competing risks between X and Z .

1.3.2 Random Signs Censoring

Cooke [8, 9] suggested that the competing risks situation between failure and PM will often satisfy what he called the random signs censoring property. The important features of random signs censoring are that the marginal distribution

of X is always identifiable, and that an indication of the validity of this type of censoring could be found from data plotting.

A lifetime Z is said to be a random signs censoring of X if the event $\{Z < X\}$ is stochastically independent of X , i.e. if the event of having a PM before failure is not influenced by the time X at which the system fails or would have failed without PM. The idea is that the system emits some kind of signal before failure, and that this signal is discovered with a probability which does not depend on the age of the system.

We introduce now some notation. Below we assume without further mentioning that X, Z are positive and continuous random variables, with $P(X = Z) = 0$. We let $F_X(t) = P(X \leq t)$ and $F_Z(t) = P(Z \leq t)$ be the cumulative distribution functions of X and Z , respectively. The subdistribution functions of X and Z are defined as, respectively, $F_X^*(t) = P(X \leq t, X < Z)$ and $F_Z^*(t) = P(Z \leq t, Z < X)$.

Note that the functions F_X^* and F_Z^* are nondecreasing with $F_X^*(0) = 0$ and $F_Z^*(0) = 0$. Moreover, we have $F_X^*(\infty) + F_Z^*(\infty) = 1$.

We will also use the notion of conditional distribution functions, defined by $\tilde{F}_X(t) = P(X \leq t | X < Z)$ and $\tilde{F}_Z(t) = P(Z \leq t | Z < X)$. Note then that $\tilde{F}_X(t) = F_X^*(t)/F_X^*(\infty)$, $\tilde{F}_Z(t) = F_Z^*(t)/F_Z^*(\infty)$.

It is important to note that the functions $F_X^*, F_Z^*, \tilde{F}_X, \tilde{F}_Z$ are identifiable from data of the form (Y, δ) , since they are given in terms of probabilities of events that can be expressed by (Y, δ) . For example, $F_X^*(t) = P(Y \leq t, \delta = 0)$ and can hence be estimated consistently from a sample of values of (Y, δ) .

On the other hand, as already mentioned, the marginal distribution functions F_X, F_Z are not identifiable in general since they are not probabilities of events that can be expressed by (Y, δ) .

We now show that the marginal distribution of X is identifiable under random signs censoring. In fact this follows directly from the definition, since we must have

$$\tilde{F}_X(t) = P(X \leq t | X < Z) = P(X \leq t) = F_X(t) \quad (10)$$

by independence of X and the event $X < Z$. As verified above, $\tilde{F}_X(t)$ can always be estimated consistently from data, and thus this holds for $F_X(t)$ as well by (10). Hence we have the somewhat surprising result under random signs censoring that the marginal distribution of X is the same as the distribution of the observed occurrences of X .

Cooke [8] showed that under random signs censoring we have

$$\tilde{F}_X(t) < \tilde{F}_Z(t) \text{ for all } t > 0. \quad (11)$$

Moreover, he showed the kind of inverse statement that whenever (11) holds, there exists a joint distribution of (X, Z) satisfying the requirements of random signs censoring and giving the same sub-distribution functions.

On the other hand, if $\tilde{F}_X(t) \geq \tilde{F}_Z(t)$ for some t , then there is no joint distribution of (X, Z) for which the random signs requirement holds. For more discussion on random signs censoring and its applications we refer to Cooke [8, 9] and Bedford and Cooke [4, Ch. 9]. One idea is to estimate the functions

$\tilde{F}_X(t)$ and $\tilde{F}_Z(t)$ to check whether (11) may possibly hold and when this is the case to suggest a model satisfying the random signs property.

1.3.3 The Repair Alert Model

Lindqvist et al. [24] introduced the so called repair alert model which extends the idea of random signs censoring by defining an additional repair alert function which describes the “alertness” of the maintenance crew as a function of time. The definition can be given as follows:

The pair (X, Z) of life variables satisfies the requirements of the repair alert model provided the following two conditions both hold:

- (i) Z is a random signs censoring of X
- (ii) There exists an increasing function G defined on $[0, \infty)$ with $G(0) = 0$, such that for all $x > 0$,

$$P(Z \leq z | Z < X, X = x) = \frac{G(z)}{G(x)}, \quad 0 < z \leq x$$

The function G is called the cumulative repair alert function. Its derivative g (when it exists) is called the repair alert function. The repair alert model is hence a specialization of random signs censoring, obtained by introducing the repair alert function G .

Part (ii) of the above definition means that, given that there would be a failure at time $X = x$, and given that the maintenance crew will perform a PM before that time (i.e. given that $Z < X$), the conditional density of the time Z of this PM is proportional to the repair alert function g .

Lindqvist et al. [24] showed that whenever (11) holds there is a unique repair alert model giving the same sub-distribution functions. Thus, restricting to repair alert models we are able to strengthen the corresponding result for random signs censoring which does not guarantee uniqueness.

The repair alert function is meant to reflect the reaction of the maintenance crew. More precisely, $g(t)$ ought to be high at times t for which failures are expected and the alert therefore should be high. Langseth and Lindqvist [17] simply put $g(t) = \lambda(t)$ where $\lambda(t)$ is the failure rate of the marginal distribution of X . This property of $g(t)$ of course simplifies analyses since it reduces the number of parameters, but at the same time it seems fairly reasonable given a competent maintenance crew. In a subsequent paper, Langseth and Lindqvist [18] present ways to test whether $g(t)$ can be assumed equal to the hazard function $\lambda(t)$.

It follows from the construction in Lindqvist et al. [24] that the repair alert model is completely determined by the marginal distribution function F_X of X , the (cumulative) repair alert function G , the probability $q \equiv P(Z < X)$, and the assumption that X is independent of the event $\{Z < X\}$ (i.e. random signs censoring). Thus, given statistical data, the inference problem consists in

estimating $F_X(t)$ (possibly on parametric form), the repair alert function g (or G), and the probability q of PM. We refer to Lindqvist et al. [24] and Lindqvist and Langseth [23] for details on such statistical inferences.

The following is a simple example of a repair alert model.

Example 3. Let (X, Z) be a pair of life variables with joint density parameterized by $\lambda > 0$ and $0 < q < 1$,

$$f_{XZ}(x, z; \lambda, q) = (q/x)\lambda e^{-\lambda x} \text{ for } x > 0, 0 < z < x/q.$$

The marginal distribution of X is the exponential distribution with density $f_X(x) = \lambda e^{-\lambda x}$, while the conditional distribution of Z given $X = x$ is the uniform distribution on $(0, x/q)$. From this we obtain $P(Z < X|X = x) = q$ for all $x > 0$. Thus the event $Z < X$ is independent of X and condition (i) of the definition is satisfied. The following computation shows that condition (ii) holds as well. Let $0 < z < x$. Then

$$\begin{aligned} P(Z \leq z|Z < X, X = x) &= \frac{P(Z \leq z, Z < X|X = x)}{P(Z < X|X = x)} \\ &= \frac{P(Z \leq z|X = x)}{q} \\ &= \frac{z(q/x)}{q} = \frac{z}{x}, \end{aligned}$$

which implies condition (ii) of Definition 2 with $G(t) = t$.

The practical interpretation of this example is as follows. We consider a component or system with lifetime X which is exponentially distributed with failure rate λ . With probability q a PM is performed before X , at a time which for given $X = x$ is uniformly distributed on the interval from 0 to x .

1.3.4 Further Properties of The Repair Alert Model

The following formula (taken from Lindqvist et al., 2006) shows in particular why (11) holds. Note that for random signs and hence for the repair alert model we have $\tilde{F}_X(t) = F_X(t)$.

$$\tilde{F}_Z(t) = F_X(t) + G(t) \int_t^\infty \frac{f_X(y)}{G(y)} dy \quad (12)$$

$$(13)$$

We next discuss some implications of the repair alert model, in particular how the parameters q and G influence the observed performance of PM and failures. In order to help intuition, we sometimes consider the power version $G(t) = t^\beta$ where $\beta > 0$ is a parameter. Then $g(t) = \beta t^{\beta-1}$ so $\beta = 1$ means a constant repair alert function, while $\beta < 1$ and $\beta > 1$ correspond to, respectively, a decreasing and increasing repair alert function.

Under the random signs assumption, the parameter $q = P(Z < X)$ is connected to the ability to discover “signals” regarding a possibly approaching failure. More precisely, q is understood as the probability that a failure is avoided by a preceding PM.

Given that there will be a PM, one should ideally have the time of PM immediately before the failure. It is seen that this issue is connected to the function G . For example, large values of β will correspond to distributions with most of its mass near x .

Moreover, it follows from equation (12) that

$$E(Z|Z < X) = \int_0^\infty (1 - \tilde{F}_Z(z))dz = E(X) - E\left[\frac{M(X)}{G(X)}\right]$$

where $M(x) = \int_0^x G(t)dt$. For the special case when $G(t) = t^\beta$, we obtain the simple result

$$E(Z|Z < X) = \frac{\beta}{\beta + 1}E(X) \quad (14)$$

which clearly indicates that good PM performance corresponds to large values of β . An interesting observation is, furthermore, that (14) can be used to estimate β from a sample of (Y, δ) . In fact, $E(Z|Z < X)$ can be estimated simply by the average of the observed Z , and since $E(X) = E(X|X < Z)$ for random signs censoring, we can estimate $E(X)$ similarly by the average of the observed X . An estimate of the quotient $\beta/(\beta + 1)$ and hence of β follow.

Instead of merely considering the conditional expectation $E(Z|Z < X)$ one may more generally study the conditional distribution of Z given $Z < X$, or the conditional distribution of $X - Z$ given $Z = z, Z < X$. A good PM performance would then mean that the former distribution is stochastically as large as possible, while the latter distribution should be small (stochastically). For precise results in this direction we refer to Lindqvist et al. [24].

Consider next $Y = \min(X, Z)$, which is the actual sojourn time. The following results are hence of practical interest, and may in addition shed light on the influence of the parameters of the repair alert model:

$$\begin{aligned} P(Y \leq t) &= F_X(t) + qG(t) \int_t^\infty \frac{f_X(y)}{G(y)} dy \\ E(Y) &= E(X) - qE\left[\frac{M(X)}{G(X)}\right], \text{ where } M(x) = \int_0^x G(t)dt. \end{aligned}$$

Furthermore, if $G(t) = t^\beta$, then

$$E(Y) = E(X) \left(1 - \frac{q}{\beta + 1}\right) \quad (15)$$

We finally give a simple illustration of how the parameters q and β (assuming $G(t) = t^\beta$ for simplicity) influence the long run cost per time unit under the repair alert model. Let C_{PM}, C_F be costs of PM and failure, respectively, for

a single sojourn. Assume now that following an event (PM or failure), the operation is restarted with a system assumed to be as good as new, and that this process continues. This leads to a sequence of observations of (Y, δ) , which we shall assume are independent and identically distributed. The theory of renewal reward processes (e.g. Ross [31], p. 78) implies that the expected cost per unit time in the long run equals the expected cost per sojourn divided by the expected length of a sojourn, i.e.

$$\frac{qC_{PM} + (1 - q)C_F}{E(X) \left(1 - \frac{q}{\beta+1}\right)}$$

where we used (15).

This is a decreasing function of β , which seems reasonable. On the other hand, it is a decreasing function of q provided $\beta > C_{PM}/(C_F - C_{PM})$. This last inequality is likely to hold in many practical cases since the right hand side will usually be much less than 1, while β should for a competent maintenance crew be larger than 1. Thus a high value of q is usually preferable.

1.4 Periodically Tested Systems

Certain systems, for example alarm systems, are tested only at fixed times which are usually periodic. If the system is found in a failed state, then it is repaired or replaced. Thus repair is usually not done at the same time as the failure, and the situation is hence not covered by the methods considered earlier in this chapter. A simple model of this situation was suggested by Hokstad and Frøvig [15] and further studied and extended by Lindqvist and Amundrustad [21] which is the main source for the present section.

The approach of Lindqvist and Amundrustad [21] involves a continuous time Markov model for the system state when time runs between testing epochs, and in addition two discrete time Markov chains for the states of the system reported immediately before and after each test, respectively. As will be seen, the given framework also allows in an easy manner the potentially useful extension to modelling of incomplete repairs or maintenance actions.

We consider a standby system observed from time 0, with testing and preventive maintenance, abbreviated PM, performed periodically at times

$$\tau, 2\tau, 3\tau, \dots,$$

called PM epochs. Here $0 < \tau \leq \infty$ is the length of what we shall call the PM interval.

1.4.1 The Markov Model

Let $X(t) \in S$ denote the state of the system at time t , where the set S of possible states is finite. It is assumed that $X(t)$ behaves like a time homogeneous Markov chain as long as time runs inside PM intervals, i.e. inside time intervals $n\tau \leq t < (n+1)\tau$ for $n = 0, 1, \dots$. This Markov chain is governed by an

infinitesimal intensity matrix A , where the entry a_{jk} of A for $j \neq k$ is the transition intensity from state j to state k , see for example Taylor and Karlin [32, p. 254]. An example of an intensity matrix A is given by (16), an illustration of which is provided by the state diagram in Figure 9. Let

$$P_{jk}(t) = P(X(t) = k | X(0) = j); j, k \in S, t > 0$$

denote transition probabilities for the Markov chain governed by A and let

$$P(t) = (P_{jk}(t); j, k \in S)$$

be the corresponding transition matrix.

In order to specify the effect of maintenance and repair at PM epochs, we next introduce for $n = 1, 2, \dots$,

$$Y_n = X(n\tau-) \equiv \lim_{t \uparrow n\tau} X(t)$$

which is the state of the system immediately before the n th PM epoch. The effect of PM at time $n\tau$ is to change the state of the system from Y_n to Z_n according to a transition matrix $R = (R_{jk})$, where

$$P(Z_n = k | Y_n = j) = R_{jk}; j, k \in S.$$

Moreover, given Y_n it is assumed that Z_n is independent of all transitions of the system state before time $n\tau$.

The definitions of the Y_n and Z_n are illustrated in Figure 8.

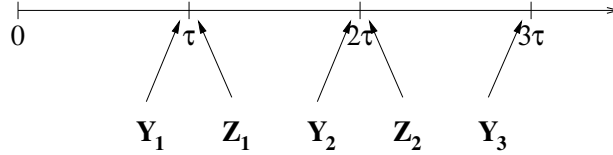


Figure 8: The definition of Y_n and Z_n .

The model description is completed by defining the initial state of the Markov chain $X(t)$ running inside the PM interval $[n\tau, (n+1)\tau)$ to be $X(n\tau) \equiv Z_n$ ($n = 0, 1, \dots$), where Z_0 is the initial state of the system, usually the perfect state in S . It is, furthermore, assumed that the Markov chain $X(t)$ on $[n\tau, (n+1)\tau)$, given its initial state Z_n , is independent of all transitions occurring before time $n\tau$.

Let the distribution of $Z_0 \equiv X(0)$ be denoted $\rho = (\rho_j; j \in S)$, where $\rho_j = P(Z_0 = j)$. Then for any $k \in S$,

$$\begin{aligned} P(Y_1 = k) &= P(X(\tau-) = k) \\ &= \sum_{j \in S} P(X(\tau-) = k | X(0) = j) P(X(0) = j) \end{aligned}$$

$$= \sum_{j \in S} \rho_j P_{jk}(\tau) = [\rho P(\tau)]_k$$

Thus the distribution of Y_1 is given by the vector-matrix product $\rho P(\tau)$. Further, for $n \geq 1$,

$$\begin{aligned} & P(Y_{n+1} = k | Y_n = j) \\ &= \sum_{\ell \in S} P(Y_{n+1} = k | Z_n = \ell, Y_n = j) P(Z_n = \ell | Y_n = j) \\ &= \sum_{\ell \in S} P_{\ell k}(\tau) R_{j\ell} = [RP(\tau)]_{jk}. \end{aligned}$$

It follows that Y_1, Y_2, \dots is a discrete time Markov chain on S with transition matrix

$$Q = RP(\tau).$$

On the other hand,

$$\begin{aligned} & P(Z_{n+1} = k | Z_n = j) \\ &= \sum_{\ell \in S} P(Z_{n+1} = k | Y_{n+1} = \ell, Z_n = j) \\ &\quad \times P(Y_{n+1} = \ell | Z_n = j) \\ &= \sum_{\ell \in S} P_{j\ell}(\tau) R_{\ell k} = [P(\tau)R]_{jk}. \end{aligned}$$

Thus, our assumptions imply that Z_0, Z_1, \dots is a discrete time Markov chain on S with transition matrix

$$T = P(\tau)R.$$

1.4.2 Reliability Measures

The approach may now be used to compute interesting reliability measures.

Average rate of critical failures. Let $\pi = (\pi_j, j \in S)$ be the stationary distribution of the Markov chain Y_1, Y_2, \dots , i.e. π is the unique probability vector satisfying the equation

$$\pi Q \equiv \pi RP(\tau) = \pi.$$

For any subset $G \subset S$, define $\pi_G = \sum_{j \in G} \pi_j$. This is the expected relative number of PM epochs, in the long run, where the system is found to be in G . Moreover, $1/\pi_G$ is the mean time, in the long run, between visits to G (measured with time unit τ). These facts are well known from the theory of Markov chains (Taylor and Karlin, 1984).

Let in the following G be the subset of S defining the critical failure states of the system. Then as in HF we define the mean time between critical failures to be

$$MTBF_{crit} = \tau / \pi_G$$

and the average rate of critical failures to be

$$\lambda_{crit} = 1/MTBF_{crit} = \pi_G/\tau.$$

Critical safety unavailability. Consider a PM interval $[n\tau, (n+1)\tau)$. The expected relative amount of time in this interval that the system is in a critical state, i.e. in G , is

$$U_n = \frac{1}{\tau} \int_{n\tau}^{(n+1)\tau} P(X(t) \in G) dt$$

By our assumptions, $X(t)$ behaves in the interval $[n\tau, (n+1)\tau)$ in the same manner as if it was run in the interval $[0, \tau)$ and started in state Z_n . Thus

$$U_n = \frac{1}{\tau} \int_0^\tau \sum_{j \in S} P_{jG}(t) P(Z_n = j) dt$$

where $P_{jG}(t) = \sum_{k \in G} P_{jk}(t)$.

Letting n tend to infinity, the $P(Z_n = j)$ tend to the limiting values γ_j defined from the stationary distribution $\gamma = (\gamma_j)$ of the Markov chain Z_0, Z_1, \dots . This distribution is found by solving the equations

$$\gamma T \equiv \gamma P(\tau) R = \gamma.$$

Following Hokstad and Frøvig [15] we shall define the Critical Safety Unavailability (CSU) of the system by

$$\begin{aligned} CSU &= \lim_{n \rightarrow \infty} U_n \\ &= \frac{1}{\tau} \int_0^\tau \sum_{j \in S} P_{jG}(t) \gamma_j dt = \sum_{j \in S} \gamma_j Q_j \end{aligned}$$

where

$$Q_j = \frac{1}{\tau} \int_0^\tau P_{jG}(t) dt$$

is the critical safety unavailability given that the system state is j at the beginning of the PM interval.

1.4.3 The Failure Model of Hokstad and Frøvig [15]

As an illustration we shall reconsider the most general failure model of Hokstad and Frøvig [15], namely their Failure Mechanism III. Here the state space is

$$S = \{O, D, K^I, K^{II}\},$$

where O = the system is as good as new, D = the system has a failure classified as degraded (noncritical), K^I = the system has a failure classified as critical,

caused by a sudden shock, K^{II} = the system has a failure classified as critical, caused by the degradation process.

It is assumed that the Markov chain $X(t)$ is defined by the state diagram of Figure 9, and thus has infinitesimal transition matrix

$$A = \begin{bmatrix} -\lambda_d - \lambda_k & \lambda_d & \lambda_k & 0 \\ 0 & -\lambda_k - \lambda_{dk} & \lambda_k & \lambda_{dk} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad (16)$$

Note that both K^I and K^{II} are absorbing states.

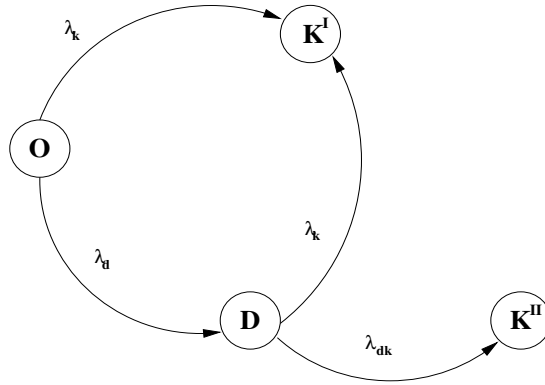


Figure 9: State diagram for the failure mechanism of HF.

The model assumes that no repairs are done in the time intervals between PM epochs. Moreover, since A is upper triangular, we can obtain $P(t) = e^{tA}$ rather easily. It is clear that $P(t)$ can be written

$$\begin{bmatrix} P_{OO}(t) & P_{OD}(t) & P_{OK^I}(t) & P_{OK^{II}}(t) \\ 0 & P_{DD}(t) & P_{DK^I}(t) & P_{DK^{II}}(t) \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

where expressions for the entries are found in Lindqvist and Amunrustad [21].

In practice it is of interest to quantify the effect of various forms of preventive maintenance. This can be done in the presented framework by means of the repair matrix R . Some examples are given below.

If all failures are repaired at PM epochs, then the PM always returns the system back to state O , and we have

$$R = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

Next, if only critical failures are repaired at PM epochs, then the appropriate R matrix is

$$R = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

More generally one may consider an extension of this by assuming that all critical failures are repaired, while degraded failures are repaired with probability $1 - r$ and remain unrepaired with probability r , $0 \leq r \leq 1$. The repair strategy is thus determined by the parameter r .

This clearly leads to the matrix

$$R = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 - r & r & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

A more general imperfect repair model can be defined by

$$R = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 - r & r & 0 & 0 \\ 1 - r_{k1} & 0 & r_{k1} & 0 \\ 1 - r_{k2} & 0 & 0 & r_{k2} \end{bmatrix}$$

Here r has the same meaning as before, while $1 - r_{k1}$ is the probability of successful repair of a K^I failure and $1 - r_{k2}$ is the similar for K^{II} .

Note that imperfect repair models are well known in the literature of repairable systems. For example the model of Brown and Proschan [6] specifies a probability p of the occurrence of a perfect repair and $1 - p$ of a minimal repair.

1.5 Discussion

In the present chapter we have considered some aspects of the modelling and analysis of repaired and maintained systems. Rather than giving a comprehensive review of the field we have concentrated on a few points, partly chosen by the interest of the author. It is believed, however, that the chapter touches some topics that have to a certain degree been overlooked in much of reliability practice.

The first point concerns the use of the NHPP as the single model for repairable systems with trend. Although this is appropriate in perhaps most cases, there are cases where renewal effects caused by repair or maintenance destroy the randomness associated with Poisson processes. One way of checking NHPP models is to embed in a larger model, and here the TRP can serve as a means of model checking (see for example the consideration of maximum log likelihoods in the Examples of Section 1.2.5 of the present chapter). Another way of extending the NHPP processes is via the large class of imperfect repair models. An introduction to the literature here is found in the review paper

Lindqvist [20]. These are models which combine two basic ingredients, a hazard rate $z(t)$ of a new system together with a particular repair strategy which governs a so called virtual age process. The idea is that the virtual age of the system is reduced at repairs by a certain amount which depends on the repair strategy. The extreme cases are the perfect repair or renewal models where the virtual age is set to 0 after each repair, and the minimal repair or NHPP models where the virtual age is not reduced at repairs and hence always equals the actual age.

Secondly, we have put some emphasis on the consideration of possible heterogeneity between systems of the same kind. Recall our Example 2 based on data from Bhattacharjee et al. [5]. The authors write in their conclusion: “The heterogeneity of failure behaviour of safety related components, such as valves in our case study, may have important implications for reliability analysis of safety systems. If such heterogeneity is not identified and taken into account, the decisions made to maintain or to enhance safety can be non-optimal or even erroneous. This non-optimality is more serious if the safety related decisions are made on the basis of failure histories of the components”. Still it is believed that heterogeneity has been neglected in many reliability applications. In fact, analyses of reliability data will often lead to an apparent decreasing failure rate which is counterintuitive in view of wear and ageing effects. Proschan [29] pointed out that such observed decreasing rates could be caused by unobserved heterogeneity. Proschan presented failure data from 17 air conditioner systems on Boeing 720 airplanes, concluding that an HPP model was appropriate for each plane, but that the rates differed from plane to plane. This is a classical example of heterogeneity in reliability. If times between failures had been treated as independent and identically distributed across planes, the conclusion would have been that these times between failures had a decreasing failure rate.

It has long been known in biostatistics that neglecting individual heterogeneity may lead to severe bias in estimates of lifetime distributions. The idea is that individuals have different “frailties”, and that those who are most “frail” will die or fail earlier than the others. This in turn leads to a decreasing population hazard, which has often been misinterpreted in the same manner as mentioned for the reliability applications. Important references on heterogeneity in the biostatistics literature are Vaupel et al. [33], Hougaard [16] and Aalen [1]. It should be noted that heterogeneity is in general unidentifiable if being considered an individual quantity. For identifiability it is necessary that frailty is common to several individuals, for example in family studies in biostatistics, or if several events are observed for each individual, such as for the repairable systems considered in this paper. The presence of heterogeneity is often apparent for data from repairable systems if there is a large variation in the number of events per system. However, it is not really possible to distinguish between heterogeneity and dependence of the intensity on past events for a single process.

The third point to be mentioned regards the use, or lack of use, of methods for competing risks in reliability applications. The following is a citation from the article on Competing Risks in Encyclopedia of Actuarial Sciences from 2003 (Crowder [11]): “If something can fail, it can often fail in one of several ways and

sometimes in more than one way at a time. In the real world, the cause, mode, or type of failure is usually just as important as the time to failure. It is therefore remarkable that in most of the published work to date in reliability and survival analysis there is no mention of competing risks. The situation hitherto might be referred to as a lost case". Fortunately, there has been done some work recently in order to include competing risks in the study of repaired and maintained systems. Much of this work, reviewed in Section 1.3, has been motivated by the work of Cooke [9] and his collaborators. His point of departure was formulated in the conclusion of [9]: "The main themes of parts I and II of this article are that current RDB (Reliability Data Bank) designs: 1. are not giving RDB users what they need 2. are not doing a good job of analyzing competing risk data 3. are not doing a good job in handling uncertainty. Improvements in all these areas are possible. However, it must be acknowledged that the models and methods presented here merely scratch the surface. It is therefore appropriate to conclude with a summary of open issues..."

The final section of the chapter considers an example of an approach which in some sense generalizes the competing risks issue, namely using Markov chains to model failure mechanisms of various equipment.

The present chapter has mostly considered the modelling of repairable systems, with less emphasis on statistical methods. There are, however, certainly large statistical challenges connected to the models. This point was touched by Cooke in the citation above, and it is clear that the information in reliability databases could and should be handled by more sophisticated methods than the ones that are traditionally used. Here there is much to learn from the biostatistics literature where there is a large emphasis on nonparametric methods and on regression methods using covariate information. As an example, Peña [27] reviews a class of models suggested in Peña and Hollander [28]. This is a virtual age model which includes the possibility of heterogeneity between systems, time-dependent covariates, and for which in addition the conditional intensities may depend on the number of previous events. This last feature adds an interesting flexibility to the model. In particular it enables modelling of certain load sharing processes and software failure processes.

References

- [1] Aalen OO, (1988) Heterogeneity in survival analysis. *Statistics in Medicine* 7:1121–1137.
- [2] Andersen P, Borgan O, Gill R, Keiding, N, (1992) *Statistical Models Based on Counting Processes*. Springer, New York.
- [3] Ascher H, Feingold H, (1984) *Repairable Systems – Modeling, inference, misconceptions and their causes*. Marcel Dekker, New York.
- [4] Bedford T, Cooke RM, (2001) *Probabilistic Risk Analysis: Foundations and Methods*; Cambridge University Press: Cambridge.

- [5] Bhattacharjee M, Arjas E, Pulkkinen, U, (2003) Modelling heterogeneity in nuclear power plant valve failure data. In: *Mathematical and Statistical Methods in Reliability* (Lindqvist BH, Doksum KA, eds.) World Scientific Publishing, Singapore, pp 341–353.
- [6] Brown M, Proschan F, (1983) Imperfect repair. *Journal of Applied Probability* 20:851–859.
- [7] Cook RJ, Lawless JF, (2002) Analysis of repeated events. *Statistical Methods in Medical Research* 11:141–166.
- [8] Cooke RM, (1993) The total time on test statistics and age-dependent censoring. *Statistics and Probability Letters* 18:307–312.
- [9] Cooke RM, (1996). The design of reliability databases, Part I and II. *Reliability Engineering and System Safety* 51:137–146 and 209–223.
- [10] Crowder MJ, (2001) *Classical competing risks*. Chapman & Hall/CRC, Boca Raton.
- [11] Crowder MJ, (2004) Competing risks. In: *Encyclopedia of actuarial science* (Teugels JL, Sundt B, eds.) Wiley, Chichester, pp. 305–313.
- [12] Crowder MJ, Kimber AC, Smith RL, Sweeting TJ, (1991) *Statistical Analysis of Reliability Data*. Chapman & Hall, Great Britain.
- [13] Doyen L, Gaudoin O, (2005) Imperfect maintenance in a generalized competing risk framework. To appear in *Journal of Applied Probability*.
- [14] Follmann DA, Goldberg MS, (1988) Distinguishing heterogeneity from decreasing hazard rate. *Technometrics* 30:389–396.
- [15] Hokstad P, Frøvig AT, (1996) The modelling of degraded and critical failures for components with dormant failures. *Reliability Engineering and System Safety* 51:189–199.
- [16] Hougaard P, (1984) Life table methods for heterogeneous populations: Distributions describing the heterogeneity. *Biometrika* 71:75–83.
- [17] Langseth H, Lindqvist BH, (2003) A maintenance model for components exposed to several failure mechanisms and imperfect repair. In: *Mathematical and Statistical Methods in Reliability* (Lindqvist BH, Doksum KA, eds.). World Scientific Publishing, Singapore, pp 415–430.
- [18] Langseth H, Lindqvist BH, (2006) Competing risks for repairable systems: A data study. *Journal of Statistical Planning and Inference* 136:1687–1700.
- [19] Lawless JF, (1987) Regression methods for Poisson process data. *Journal of American Statistical Association* 82:808–815.

- [20] Lindqvist BH, (2006) On the statistical modelling and analysis of repairable systems. To appear in *Statistical Science*.
- [21] Lindqvist BH, Amundrustad H, (1998) Markov models for periodically tested components. In: *Safety and Reliability. Proceedings of the European Conference on Safety and Reliability - ESREL '98* (Lydersen S, Hansen GK, Sandtorv HA). AA Balkema, Rotterdam, pp 191–197.
- [22] Lindqvist BH, Elvebakk G, Heggland K, (2003) The trend-renewal process for statistical analysis of repairable systems. *Technometrics* 45:31–44.
- [23] Lindqvist BH, Langseth H, (2005) Statistical modelling and inference for component failure times under preventive maintenance and independent censoring. In: *Modern Statistical and Mathematical Methods in Reliability* (Wilson A, Limnios N, Keller-McNulty S, Armijo Y). World Scientific Publishing, Singapore, pp. 323–337.
- [24] Lindqvist BH, Støve B, Langseth H, (2006) Modelling of dependence between critical failure and preventive maintenance: The repair alert model. *Journal of Statistical Planning and Inference* 136:1701–1717.
- [25] Meeker WQ, Escobar LA, (1998) *Statistical methods for reliability data*. Wiley, New York.
- [26] Nelson W, (1995) Confidence limits for recurrence data - applied to cost or number of product reappear. *Technometrics* 37:147–157.
- [27] Peña EA, (2006) Dynamic modelling and statistical analysis of event times. To appear in *Statistical Science*.
- [28] Peña EA, Hollander M, (2004) Models for recurrent events in reliability and survival analysis. In: *Mathematical Reliability: An Expository Perspective* (Soyer R, Mazzuchi T, Singpurwalla N, eds.) Kluwer Academic Publishers, Dordrecht, The Netherlands.
- [29] Proschan F, (1963) Theoretical explanation of observed decreasing failure rates. *Technometrics* 5:375–383.
- [30] Rausand M, Høyland A, (2004) *System reliability theory: Models, statistical methods, and applications*. 2nd ed. Wiley-Interscience, Hoboken, N.J.
- [31] Ross SM, (1983) *Stochastic Processes*. Wiley, New York.
- [32] Taylor HM, Karlin S, (1984) *An introduction to stochastic modeling*. Academic Press, Orlando.
- [33] Vaupel JW, Manton KG, Stallard E, (1979) The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* 16:439–454.