DTU Library

Analysis of contaminated field failure data for repairable systems

Hansen, Christian Kornerup; Thyregod, Poul

Analysis of Contaminated Field Failure Data
for Repairable Systems

Christian Kornerup Hansen □ IMSOR, The Technical University of Denmark □ Lyngby

Poul Thyregod □ IMSOR, The Technical University of Denmark □ Lyngby

SUMMARY

In this paper we present a simple model for electronic systems with repair, and a method for analysing recorded field failure data for such systems. The work performed has resulted in analytical results that may be used for assessing the product reliability. The method was originally developed for use under quite ideal circumstances, but in this paper the method is adapted for use also with "contaminated" data, i.e., data where the failure times are observed imbedded by noise. We propose a simple model for the noise that enables an analytical solution for the mean cumulative number of failures. The expression is compared with the expression for non-contaminated data, for the purpose of assessing the effect of the contamination, and for the purpose of correcting estimates in the parametric model. The method is illustrated by a real world example of industrial failure data recorded on field basis, and the effect caused by contamination in this data is investigated under a "Worst Case" assumption. The example indicates that the model is quite robust to contamination. Only one of the three key parameters is affected significantly by assuming the data to be contaminated.

INTRODUCTION

Product assurance for repairable electronic systems is naturally connected with the assessment of the reliability of the product, particularly in the early (warranty) period. The reliability is usually measured in terms of the expected number of failures as a function of time, or its time derivative, evaluated from either standard handbooks or from empirical experience. Since neither handbooks nor laboratory experiments (such as accellerated life tests) usually reflect the true behavior of the equipment when operating under field conditions, it has become increasingly common to base the reliability assessment on analyses of recorded field failure data. Such analyses, however, may be complicated due to contamination of the data. There are various sources for this contamination, some of them as listed below:

- *Inaccuracy in the recording of failure times (ages).*
  Failure reports are practically never complete. Dates and other information may be missing or erroneous, and have to be estimated.

- *Failures are not detected instantaneously.*
  The system is often assumed to have a simple coherent structure such as a series system, so that the system is assumed to fail immediately after one of its components have failed. In practice, the system may continue to operate for a period even after one of its components has failed, or the misbehavior of the system may not be noticed instantaneously by its user.

- *The systems are up and down randomly.*
  Very few electronic systems are operating continuously 24 hours a day, but more usually they are turned on and off randomly. These downperiods of the system should be ignored in the determination of the failure times.

Intuitively the above listed items, if ignored, will cause a bias in the estimation of parameters, and in worst cases we may under-estimate the expected costs through warranty agreements etc. However, very little attention has been addressed to the mathematical modeling and assessment of these effects.

In the paper we shall consider a series system, and a simple model for the component lifetime distribution that agrees with the well-known fact that electronic systems often fail more frequent in its early operating period than in the preceeding period. This behavior may be ascribed to "infant failure" for components, and may be considered as a consequence of heterogeneity in the component population. Even for components produced under efficient quality control usually a small proportion of the components have (invisible) defects or imperfections that make those components more prone to failure. This small population is usually referred to as a "freak population" (Refs. 1-2). In methods for improving the product reliability usually the objective is to reduce the size of this population through an appropriate "screening" strategy.

ASSUMPTIONS

1. All systems in the field are operating under equivalent conditions.

2. A system is considered as a series system.

3. Components have mixed exponential lifetimes.

4. The observation noise is exponential and independent of the lifetimes.

MATHEMATICAL NOTATION AND NOMENCLATURE

t		Time Reference (Component or System Age)
T		Lifetime of a Component (Time to Failure)
ϵ		Observation Noise
$T^{(c)} = T + \epsilon$		Contaminated Lifetime Observation
n		Number of Sockets in a System
$N.(t)$		Observed Number of Failures as Function of t
$M.(t) = E[N.(t)]$	MCNF	Mean Cumulative Number of Failures
$m.(t) = \frac{d}{dt}M.(t)$	ROCOF	Rate of Occurrence Of Failures
$\hat{M}.(t)$	ACNF	Average Cumulative Number of Failures
$R(t)$		Number of Systems Active at Age t
$\hat{V}.(t)$		Estimated Variance of the ACNF Estimate
	SRP	Superimposed Renewal Process
$f.(t)$	PDF	Probability Density Function
$F.(t)$	CDF	Cumulative Distribution Function
λ_0		Failure Intensity for a Standard Component [Days^{-1}]
λ_1		Failure Intensity for a Weak Component [Days^{-1}]
p		Proportion of Weak Components
λ_2		Intensity Parameter in the Noise Distribution [Days^{-1}]

REPAIRABLE SYSTEMS

The SRP Model

In the simplest (and most common) model for a *repairable system* we consider the system as a collection of (at least 2) *sockets* with *components* (parts) attached, see Refs. 3-4. The components are considered as non-repairable items that are discarded after failure, and the system fails if and only if one of its components fail (a series system). At failure the system is restored to completely satisfactory operation by replacing the failed component by a new one. With these assumptions the behavior of the system is completely determined by the lifetimes of the involved components in each socket.

Consider a system as described above with sockets numbered $1, \ldots, n$. In the ith socket the lifetime of the attached component has the cumulative distribution function (CDF)

$$F_i(t) = P(T \leq t) \qquad (1)$$

and the probability density function

$$f_i(t) = \frac{d}{dt}F_i(t) \qquad (2)$$

where T is the lifetime of the component. Since the components are always replaced by components that are "the same as new" it follows from the model that the counting process which counts the number of failures in the ith socket as a function of time

$$N_i(t) \qquad (3)$$

may be considered as a *renewal process* with interarrival times determined by the lifetimes for the components that successively were put in that socket, and the system may be considered as a *superimposed renewal process* (SRP), since the number of system failures $N(t)$ is determined simply as the sum of the n independent renewal processes $N_i(t)$ (Refs. 4-5).

$$N(t) = \sum_{i=1}^{n} N_i(t) \qquad (4)$$

The socket renewal process is usually characterized by either the *mean cumulative number of failures* (MCNF), $M_i(t)$

$$M_i(t) = E[N_i(t)] \qquad (5)$$

or the *rate of occurrence of failures* (ROCOF) $m_i(t)$

$$m_i(t) = \frac{d}{dt}M_i(t) \qquad (6)$$

and for the system

$$M(t) = E[N(t)] = \sum_{i=1}^{n} M_i(t) \qquad (7)$$

$$m(t) = \frac{d}{dt}M(t) \qquad (8)$$

It follows from the relations Eqs. (7) - (8), that the MCNF and ROCOF functions for the system are determined from the collection of MCNF and ROCOF functions for the sockets in the system. Therefore, we may without loss of generality restrict our discussion to a single socket, and consequently in the following, we shall omit the socket index i in the formulas.

The ROCOF, and the corresponding PDF associated with the socket are connected through the well-known formula

$$m^*(s) = \frac{f^*(s)}{1 - f^*(s)} \qquad (9)$$

where $m^*(s)$ and $f^*(s)$ denotes the Laplace transforms of the ROCOF and the PDF respectively (Ref. 5, p. 48). In simple cases this leads to an analytical expression for the MCNF and the ROCOF.

A Simple Component Lifetime Model

It is well-known that constant failure rate lifetime models for components (i.e., the component is assumed to have an exponentially distributed lifetime) implies a constant ROCOF function. This is a quite simple and convenient model, however, in practical applications this model will often be too primitive, and give a very poor fit to observed failure data, see e.g. Ref. 6. For repairable systems an observed ROCOF is usually ininitally decreasing from high level and then leveling off to an asymptotic constant low level – at least for the period before components starts to wear out.

As previously mentioned, the large amount of "early failures", see Ref. 7, may be considered as a simple consequence of heterogeneity among the components. A quite simple way to deal with this heterogeneity is to assume, that for a large population of similar components a small proportion p of the components are "weak", and thus more likely to fail early than the remaining proportion 1-p "standard" components. The resulting lifetime distribution (for a component randomly drawn from the population) is usually referred to as a *mixed distribution* (sometimes *bimodal*, see Ref. 1). Assuming an exponential distribution for both "standard" and "weak" components results in the CDF and PDF

$$F(t) = (1 - p)\{1 - e^{-\lambda_0 t}\} + p\{1 - e^{-\lambda_1 t}\} \qquad (10)$$

$$f(t) = (1 - p)\lambda_0 e^{-\lambda_0 t} + p\lambda_1 e^{-\lambda_1 t} \qquad (11)$$

where λ_0 and λ_1 are failure intensities (i.e., the reciprocal of the *mean time to failure*) for the exponential distributions corresponding to "standard" and "weak" type components.

By taking the Laplace transforms of Eq. (11) and applying Eq. (9), we obtain an analytical expression for the MCNF (Ref. 7)

$$M(t) = K_0 t + K_1\{1 - e^{-K_2 t}\} \qquad (12)$$

where

$$K_0 = \frac{\lambda_0 \lambda_1}{p\lambda_0 + (1 - p)\lambda_1} \approx \lambda_0 \qquad (13)$$

$$K_1 = \frac{(1 - p)\lambda_0 + p\lambda_1 - K_0}{K_2} \approx p \qquad (14)$$

$$K_2 = (1 - p)\lambda_1 + p\lambda_0 \approx \lambda_1 \qquad (15)$$

The approximations are valid if $\lambda_0 \ll \lambda_1$ and $p \ll 1$, which is usually the case.

It should be noted that although we are dealing with *repairable* systems, under the SRP model it is sufficient to consider a set of *non-repairable* items, viz. the components attached to a socket in the system. The lifetimes of the failed components determine the renewal process for the socket, and the renewal processes for the sockets determine the SRP for the system.
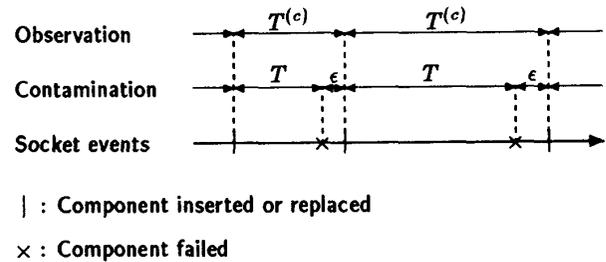
A MODEL FOR CONTAMINATED DATA

Figure 1: *Contaminated observations for a socket renewal process.*

Under ideal circumstances the lifetime of a failed component is observed without error. As previously mentioned, it is more likely that we observe not the actual lifetime, but the lifetime embedded by noise, and typically a lifetime that is larger than the actual one. We shall refer to a such observation as a contaminated observation, and denote it by $T^{(c)}$

$$T^{(c)} = T + \epsilon \qquad (16)$$

where T is the actual (true) lifetime and ϵ is the observation noise, see Figure 1. Since our statistics have to based on what we see, i.e., the observation $T^{(c)}$, we will have to consider the distribution of $T^{(c)}$ rather than that of T. We shall consider only the simple case where ϵ is modeled as an exponentially distributed random variable independent of T and with intensity parameter λ_2. This implies that ϵ is non-negative, but as mentioned earlier, this is often a reasonable assumption. In the time domain the PDF for $T^{(c)}$ is expressed as a convolution integral involving the PDFs of T and ϵ, but for the Laplace transforms we obtain the simple relation

$$f_c^*(s) = f(s) \cdot f_\epsilon^*(s) \qquad (17)$$

where $f^*(s)$ and $f_\epsilon^*(s)$ are the Laplace transforms of the PDFs corresponding to the random variables T and ϵ respectively. For the contaminated MCNF we then obtain

$$M_c(t) = K_0^{(c)}t + K_1^{(c)}(1 - e^{-K_2^{(c)}t}) + K_3^{(c)}(1 - e^{-K_4^{(c)}t}) \qquad (18)$$

The coefficients may be computed exact – or approximate – from

$$K_0^{(c)} = \frac{P}{\alpha\beta} \approx \lambda_0 \tag{19}$$

$$K_1^{(c)} = \frac{P - Q\alpha}{\alpha^2(\alpha - \beta)} \approx \frac{\lambda_2}{\lambda_2 - \lambda_1}p \tag{20}$$

$$K_2^{(c)} = \alpha \approx \lambda_1 \tag{21}$$

$$K_3^{(c)} = \frac{P - Q\beta}{\beta^2(\beta - \alpha)} \approx -\frac{\lambda_0}{\lambda_2} - \frac{\lambda_1}{\lambda_2 - \lambda_1}p \tag{22}$$

$$K_4^{(c)} = \beta \approx \lambda_2 \tag{23}$$

where

$$P = \lambda_0\lambda_1\lambda_2 \tag{24}$$

$$Q = \lambda_2\{(1 - p)\lambda_0 + p\lambda_1\} \tag{25}$$

and $-\alpha$ and $-\beta$ are the roots in the quadratic equation

$$s^2 + (\lambda_0 + \lambda_1 + \lambda_2)s + \lambda_0\lambda_1 + p\lambda_0\lambda_2 + (1 - p)\lambda_1\lambda_2 \tag{26}$$

It can be shown that Eq. (26) has two distinct, negative roots, i.e., $\alpha, \beta > 0$ and $\alpha \neq \beta$. The given approximations are quite good if $\lambda_0 \ll \lambda_1$, $\lambda_0 \ll \lambda_2$ and if $p \ll 1$. In the case where $\lambda_1 = \lambda_2$ the approximations in Eqs. (20) and (22) are undefined, but similar approximations in this case can be obtained.
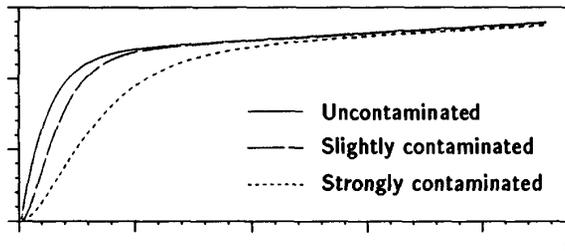

Figure 2: *Examples of contaminated MCNF curves, each curve corresponding to a different value of the contamination parameter* λ_2.

The approximations indicate that the contamination mainly affects the transient part of the MCNF function (Figure 2), whereas the steady state part is practically unaffected.

ANALYSIS OF FIELD FAILURE DATA

Data Collection and Failure Reporting

For the type of field failure data we consider we shall assume that the failures are observed from a number of systems that may be put into – and taken out of – operation at individual times. For each system we report its (censored) lifehistory which contains information of

- The time when the system was put into operation.
- The times when the system failed.
- The time periods when the system was temporarily put out of operation (down periods), e.g., because of service.
- The time when the data recording was terminated.

The data are assumed to be *socket specific*, i.e., the failures observed in each system are reported with reference to the socket in which the failure occurred. This information is necessary in order to retrieve the lifetimes of the components. Moreover, we assume that the down periods does not contribute to the system (or component) age, i.e., these periods are merely ignored in the determination of the system age.

Estimation of the Mean Cumulative Number of Failures

Introduce $T_1, T_2 \ldots T_N$ to denote the system ages corresponding to the observed failures and $R(t)$ to denote the number of systems active at age t, then a (non-parametric) estimate of the system MCNF is the *average cumulative number of failures* (ACNF), originally introduced by Nelson (Refs. 8-9)

$$\hat{M}(t) = \sum_{s \leq t} \frac{\Delta N(s)}{R(s)} \tag{27}$$

where the increment $\Delta N(t)$ is defined as

$$\Delta N(t) = \begin{cases} 1 & \text{if } t \in \{T_i\}_{i=1}^N \\ 0 & \text{otherwise} \end{cases} \tag{28}$$

Pointwise confidence intervals $[M^-(t), M^+(t)]$ can be computed approximately by methods adapted from Ref. 10

$$M^-(t) = \hat{M}(t) - U_{1-\alpha/2}\sqrt{\hat{V}(t)} \tag{29}$$

$$M^+(t) = \hat{M}(t) + U_{1-\alpha/2}\sqrt{\hat{V}(t)} \tag{30}$$

$$\hat{V}(t) = \text{Var}[\hat{M}(t)] \approx \sum_{s \leq t} \frac{\Delta N(s)}{R(s)^2} \tag{31}$$

where $U_{1-\alpha/2}$ denotes the $1-\alpha/2$ quantile in a standard $N(0, 1)$ normal distribution.

Both the ACNF function and the associated confidence intervals $M^-(t)$ and $M^+(t)$ are step functions with jumps at the failure times. The values at these jumps can be computed recursively, see Ref. 11.

The extent of this paper does not permit a detailed discussion of parametric estimation, i.e., estimation of the parameters λ_0, λ_1 and p. For non-contaminated data Ref. 7 gives a detailed discussion of the *Maximum Likelihood* (ML) method, which provides a parametric estimate of the MCNF and the associated confidence band. Alternatively Ref. 11 gives an approximate *Least Squares* (LS) method. In this method the slope and intercept of the steady state part of the ACNF are estimated together with a key point

in the transient part. The parameters are then estimated by use of Eqs. (13-15).

The LS-method for non-contaminated data can be adapted to use also for contaminated data by correcting the estimates in according to Eqs. (19-23).

A FIELD FAILURE DATA EXAMPLE

The system considered in this example is a printed circuit board with 560 sockets (component positions) installed at 3481 different industrial users, and under complete service from the manufacturer. Field failure data from these installations were collected over a 5-6 year period (1982-1987). The installed systems were put into – and taken out of – operation individually during the period; thus data from the different systems were censored individually with ages ranging from 0 to approximately 2000 days.
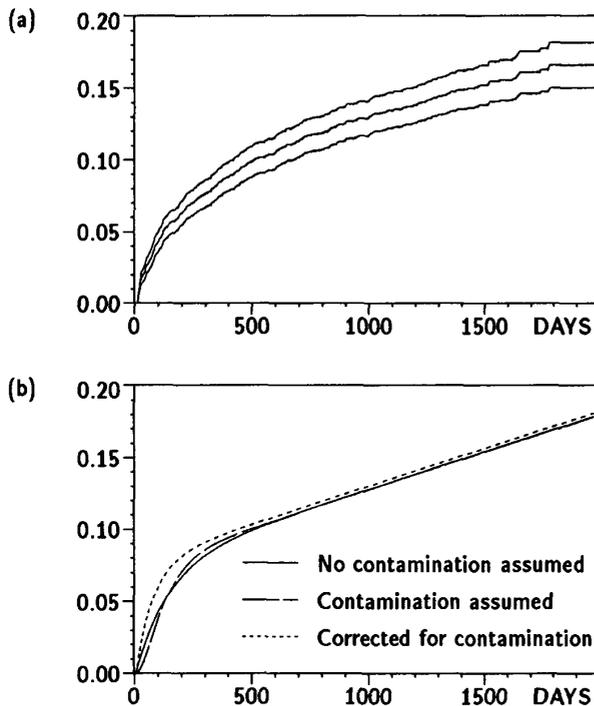
We here assume that all sockets are equivalent, i.e., the same parameters in the mixed exponential model is used for all type of components. This assumption is mainly because data are quite sparse, so that a stratification of the failures could not be empirically justified, see also Refs. 4, 7, 11. For stratified data the analysis is carried out for each stratum separately.

In Figure 3(a) the non-parametric ACNF plot, Eq. (27) is shown together with the confidence intervals, Eqs. (29-30). A visual inspection does not show any significant signs of contamination, since the plot is clearly in agreement with the theoretical shape of the (uncontaminated) MCNF function (Figure 2). However, in order to see what effect a possible contamination of these data has on the assessment of the reliability of this system, we shall consider these data as (possibly) contaminated.

From our sources we know that a contamination typically causes delays ranging from 10 - 50 days in average. We shall therefore evaluate the effect of contamination by assuming the contamination parameter λ_2 to have values $0.1, 0.05$ and 0.02; the latter choice may be thought of as a "Worst Case" choice.

The ML-method was used to obtain estimates of the parameters in the non-contaminated case $(\lambda_2 = \infty)$ and the asymptotic standard deviations of these parameter estimates. The (adapted) approximate LS-method was used in both the non-contaminated, and in the three contaminated cases. (see Table 1). Some of the corresponding parametric MCNF curves are shown i Figure 3.

(a)

(b)

Estimate	$\hat{\lambda}_0$ $(\times 10^{-9})$	$\hat{\lambda}_1$ $(\times 10^{-3})$	$\hat{p}$ $(\times 10^{-6})$
Uncorrected (ML) $\lambda_2 = \infty$	97.3	7.63	122
Uncorrected (LS) $\lambda_2 = \infty$	91.1	7.09	132
Corrected (LS) $\lambda_2 = 0.1$	91.1	7.71	133
Corrected (LS) $\lambda_2 = 0.02$	91.1	8.51	134
Corrected (LS) $\lambda_2 = 0.05$	91.1	12.5	136
Std.Dev. (ML) $\lambda_2 = 0.05$	9.9	1.28	12.9

Figure 3: *(a) The non-parametric ACNF plot with confidence intervals. (b) The estimated parametric MCNF curves i) with no contamination assumed, ii) with contamination ($\lambda_2 = 0.02$) assumed, and iii) corrected for contamination ($\lambda_2 = 0.02$).*

Table 1: *Corrected and uncorrected estimates for different values of the contamination parameter λ_2.*

We note that λ_0 and p are estimated with almost no influence from the contamination, i.e., the corrections are small in comparison with both the uncertainty in the estimation, and in comparison with the differences between the two estimation methods. The estimates of λ_1, in contrast, are sensitive to contamination.

CONCLUSION

In the paper we have investigated the effect of contamination of field failure data, by assuming a simple three-parameter model for the lifetimes of the components, and a simple one-parameter model for the observation noise. The advantage of the suggested approach is that the simple models enables analytical results which can be used to correct the estimates or analyse its sensitivity to contamination. In the given example, which is typical for the type of systems considered, it is seen that the contamination affects the estimation of the system behavior in the transient operating period, whereas the estimation in the steady state period is practically unaffected.

ACKNOWLEDGEMENTS

REFERENCES

1. F. Jensen and N.E. Petersen, *Burn-in: An engineering approach to the design and analysis of burn-in procedures*, John Wiley and Sons, 1982.

2. F. Jensen, "Component failures based on flaw distributions", in *Proceedings of the Annual Reliability and Maintainability Symposium*, 91-95, 1989.

3. H. Ascher and H. Feingold, *Repairable systems reliability. Modeling, inference, misconceptions and their causes*, Marcel Dekker, 1984

4. L. Rimestad, "Field failures of electronic systems", in *Proceedings of the Annual Reliability and Maintainability Symposium*, 528-534, 1990.

5. D.R. Cox, *Renewal theory*, Methuen & Co. Ltd., 1962

6. J.A. McLinn, "Is failure rate constant for a complex system?", in *ASQC Quality Congress Transactions*, 538-544, 1989.

7. E. Arjas, C.K. Hansen and P. Thyregod, "Heterogeneous part quality as a source of reliability improvement in repairable systems", to appear in *Technometrics*, 33, No. 1, 1991.

8. W. Nelson, "Hazard plotting for incomplete failure data", *Journal of Quality Technology*, 1, 27-51, 1969.

9. W. Nelson, "Graphical analysis of system repair data", *Journal of Quality Technology*, 20, 24-35, 1988.

10. O. Bie, Ø. Borgan and K. Liestøl, "Confidence intervals and confidence bands for the cumulative hazard rate function and their small sample properties", *Scand. J. Statist.*, 14, 221-233, 1987.

11. C.K. Hansen and P. Thyregod, "Estimation of the mean cumulative number of failures in a repairable system with mixed exponential component lifetimes", to appear in *Quality and Reliability Engineering International*, 6, No. 5, 1990.

BIOGRAPHIES

Christian Kornerup Hansen, MScEE
IMSOR
Building 321, The Technical University of Denmark
DK-2800 Lyngby, Denmark
Telephone +45 42 88 14 33
Telefax +45 42 88 13 97

Christian Kornerup Hansen recieved his Master of Science degree in electrical engineering from the Technical University of Denmark in 1988. He is currently a research fellow (Ph.D. candidate) at IMSOR, the Technical University of Denmark, where he is developing statistical methods for analyzing the reliability of repairable systems.

Poul Thyregod, PhD
IMSOR
Building 321, The Technical University of Denmark
DK-2800 Lyngby, Denmark
Telephone +45 42 88 14 33
Telefax +45 42 88 13 97

Poul Thyregod recieved his Ph.D. degree in statistics from the University of Copenhagen in 1970. He was visiting the University of Minnesota 1970-71 and is currently at IMSOR, the Technical University of Denmark, where he has been since 1971. He has been engaged in research on statistical methods in reliability and quality control over 25 years. He is a member of ASQC.