

COMPETING RISK PERSPECTIVE OVER RELIABILITY DATABASES

Cornel Bunea and Roger Cooke
Delft University of Technology
Delft, The Netherlands
e-mail: c.bunea@its.tudelft.nl
e-mail: r.m.cooke@its.tudelft.nl

Bo Lindqvist
Norwegian University of Science and Technology
Trondheim, Norway
e-mail: bo@math.ntnu.no

Abstract

Modern reliability data bases are going to be discuss, taking into account competing risk models. A new competing risk model is presented.

1 Introduction

Modern RDB's consist of data fields relating to failure modes, failure causes, maintenance/repair actions, severity of failures, component characteristics, operating characteristics. For every specific field we can distinguish ten or more competing items, which can compete to terminate the service sojourn of the component. For example, a maintenance engineer is interested in "degraded" and "incipient" failures, as they are associated with preventive maintenance. He is trying also to take the least expensive maintenance action: "repair action" or "adjustment action" are favored above "replace action". "Critical" failures are of primary interest in risk and reliability calculations and a component designer is interested in the particular component function that is lost, in the failure mechanism and he wishes to prevent the failure of the most expensive components of the system.

The general theory of competing risk is advanced here as a proper mathematical language for modeling reliability data. To check the performance of different probabilistic models, we discuss data set coming from two identical Compressor Units from one Norsk Hydro ammonia plant, for the period of observation 2-10-68 up to 25-6-89.

The competing risk models available in the literature are developed especially for nuclear sector where strict regulations are imposed, are not appropriate for all the fields of Compressor Unit Data. Due to the fact that the compressor unit consists of more heterogeneous sub-components another competing risk model was developed to interpret the competing risk between different failure modes.

2 Competing Risk

In the competing risk approach we model the data as a renewal process, that is as a sequence of i.i.d. copies of a r.v. Z . Each observable Z is the minimum of two or more variables. In many cases we can reduce the problem to the analysis of two competing risks classes, described by two random variables X and Y . Hence we observed the least of X and Y , $Z = [\min(X, Y), I(X < Y)]$ and observe which it is. For simplicity we assume that $P(X = Y) = 0$.

From data we can estimate only the subsurvival functions, $S_X^*(t) = P(X > t, X < Y)$ and $S_Y^*(t) = P(Y > t, Y < X)$, but not the true survival functions of X and Y . Note that $S_X^*(t)$ depends on Y , though this fact is suppressed in the notation. Note also that $S_X^*(0) = P(X < Y)$ and $S_Y^*(0) = P(Y < X)$, hence $S_X^*(0) + S_Y^*(0) = 1$.

The conditional subsurvival function is the subsurvival function, conditioned on the event that the failure mode in question has occurred. With continuity of $S_X^*(t)$ and $S_Y^*(t)$ at zero,

$$CS_X(t) = P(X > t, X < Y | X < Y) = S_X^*(t)/S_X^*(0), CS_Y(t) = P(Y > t, Y < X | Y < X) = S_Y^*(t)/S_Y^*(0).$$

Closely related to the notion of subsurvival functions is the probability of censoring beyond time t ,

$$\Phi(t) = P(Y < X | Y \wedge X > t) = \frac{S_Y^*(t)}{S_X^*(t) + S_Y^*(t)}.$$

This function seems to have some diagnostic value, enabling us to choose the competing risk model which fit the data. Note that $\Phi(t) = P(Y < X) = S_Y^*(0)$.

Peterson derived bounds on the survival function $S_X(t)$ by noting that

$$P(X \leq t, X \leq Y) \leq P(X \leq t) \leq P(X \wedge Y \leq t),$$

which entails $S_X^*(t) + S_Y^*(0) \geq S_X(t) \geq S_X^*(t) + S_Y^*(t)$. Note that the quantities on the left and right hand sides are observable from data.

Using the relation between failure rate and survival function, the Peterson bounds become:

$$l_{min}(t) = -\ln(S_X^*(t) + S_Y^*(0))/t \leq (1/t) \int_0^1 r_X(u) du \leq -\ln(S_X^*(t) + S_Y^*(t)) = l_{max}(t).$$

A similar relation holds for $S_Y(t)$.

2.1 Independent Exponential Competing Risk

By observing independent copies of $Z \equiv [\min(X, Y), I(X < Y)]$ we can estimate the subsurvival functions. Assuming independence of X and Y we can determine uniquely the survival functions of X and Y ; in this case the survival functions of X and Y are said to be identifiable from censored data. A continuous subsurvival pair is always consistent with an independent model, but is not always consistent with an independent exponential model. Bedford and Cooke (2001) showed that for X and Y independent exponential life variables with failure rates λ and γ , the conditional subsurvival functions of X and Y are equal and exponential distributed with failure rate $\lambda + \gamma$ and the probability of censoring beyond time t is constant.

2.2 Random Signs Censoring

Perhaps the simplest dependent competing risk model which leads to identifiable marginal distributions without restricting their form is random sign censoring. Cooke (1996) proved that the random signs model is consistent given subsurvival functions if and only if the conditional subsurvival function of X is greater than the conditional subsurvival function of Y for all $t > 0$. In this case the probability of censoring beyond time t is maximum at the origin. This results suggests that if the random signs model holds then the independent exponential model is difficult to characterize data.

2.3 Conditional independence model

Another model from which we have identifiability is conditional independent model. This model considers the competing risk variables, X and Y , as sharing a common quantity V , and as being independent given V : $X = V + W$, $Y = V + U$, where V, U, W are mutually independent. Hokstadt (1997) derived explicit expressions for the case that V, U, W are exponential distributed and obtained the same result as in the case of Independent Exponential Competing Risks: equal conditional subsurvival functions and constant probability of censoring beyond time t .

3 Analyze of Competing Data Sets

The data set proposed for discussion of different competing risk models comes from one Norsk Hydro ammonia plant operating two identical compressor units, for the period of observation 2-10-68 up to 25-6-89. This yields 21 years of observation and more than 350 events. As every modern reliability data base, this data base has a detailed compressor unit history: time of component failure; failure mode: leakage, no start, unwanted start, vibration, warming, overhaul, little gas stream, great gas stream, others; degree of failure: critical, non-critical; down time of the component; failure at the compressor unit: 1 - first unit failed, 2 - second unit failed, 3 - both units failed; failure of System and Sub-System of Compressor unit; action taken: immediate reparation, immediate replacement, adjustment, planned overhaul, modification, others.

For the analysis of this data base a software tool was developed in the higher order programming languages Visual Basic and Excel. The operator can choose two classes of competing risk for the five fields of interest: failure mode, degree of failure, action taken, system/subsystem. The analyze can be

performed over the whole time of observation or for a certain time window by specifying the limiting dates of interest.

The upper graphs show the empirical subsurvival functions, conditional subsurvival functions and the probability of censoring above time t . The bottom graphs show the time-wise average failure rate bounds for both classes of risk. Selecting an appropriate competing risk model on the bases of empirical subsurvival functions, was simply made graphically (Cooke 1996). In (Bunea 2002) we proposed a statistical test to check if an independent exponential model is appropriate to interpret data against the alternative of random signs model. However we will discuss further the data analysis only in terms of graphically visualization.

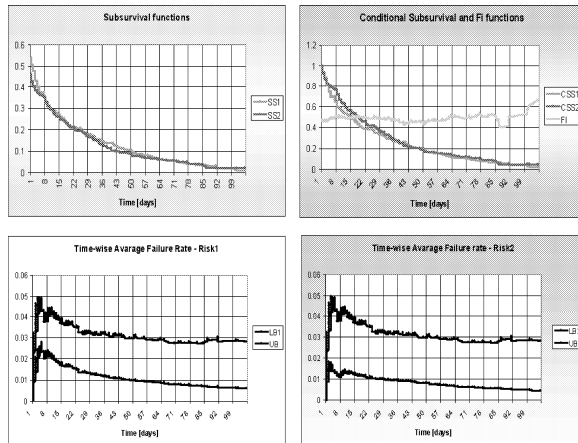


Figure 1. Analysis of “failure effect” field

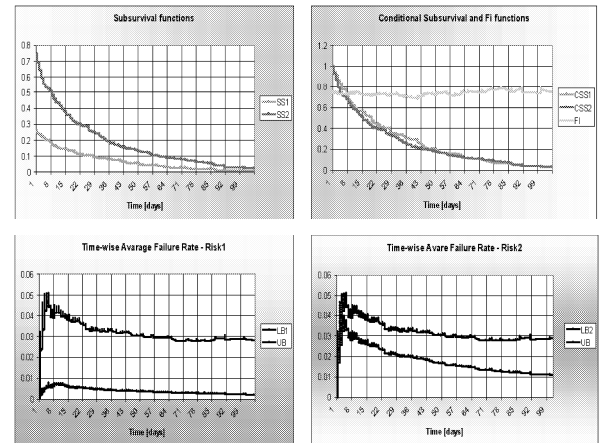


Figure 2. Analysis of “degree of failure” field

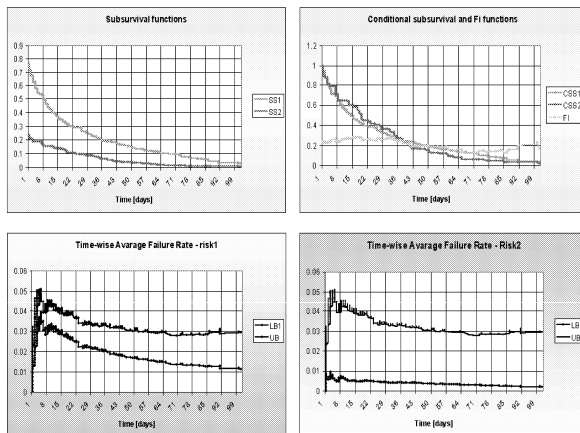


Figure 3. Analysis of “action taken” field

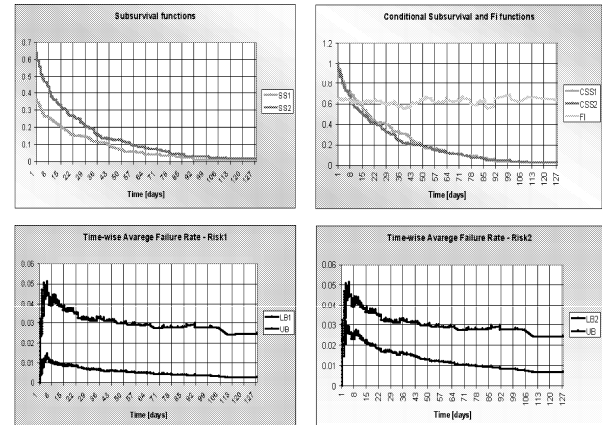


Figure 4. Analysis of “sub-system” field

In “failure effect” field analysis the class “leakage, no start, unwanted start and vibration” seems to have a negative influence on systems, hence the maintenance engineer is trying to avoid it. Figure 1. shows a slightly increasing probability of censoring after time t and non-constant time average failure rate, which is not consistent with any model available in the literature. A new model will be present in next section for this case.

The condition subsurvival function for critical and noncritical failures seem to be equal and the exponential independent model appropriate to fit data. But, the right bottom graph in Figure 2. indicates that the time average failure rate for the critical failure can not be constant and in this case the conditional independent model may be preferred.

Discussions with maintenance personnel revealed that the most expensive action that they will try to avoid is immediate replacement. The conditional subsurvival function of the two competing risk classes are crossing each other in one point. The literature provides no model for this case. The solution proposed is to modify the time window period of analysis.

In Figure 4. the most expensive components (as piston and pump) are grouped in one competing risk class. The conditional independent model is again favored against exponential independent model.

4 A new competing risk model

Let the survival function of the life time - $S_X(t)$ be a mixture of two exponential with parameters λ_1, λ_2 and the mixing coefficient p , and the censoring survival function - $S_Y(t)$ be an exponential with parameter λ_y : $S_X(t) = p \exp\{-\lambda_1 t\} + (1-p) \exp\{-\lambda_2 t\}$ and $S_Y(t) = \exp\{-\lambda_y t\}$.

If X and Z are independent, one can verify:

- $S_X^*(t) = p \frac{\lambda_1}{\lambda_y + \lambda_1} \exp\{-(\lambda_y + \lambda_1)t\} + (1-p) \frac{\lambda_2}{\lambda_y + \lambda_2} \exp\{-(\lambda_y + \lambda_2)t\}$
- $S_Y^*(t) = p \frac{\lambda_y}{\lambda_y + \lambda_1} \exp\{-(\lambda_y + \lambda_1)t\} + (1-p) \frac{\lambda_y}{\lambda_y + \lambda_2} \exp\{-(\lambda_y + \lambda_2)t\}$
- $\frac{S_X^*(t)}{S_X^*(0)} = \frac{\exp\{-(\lambda_y + \lambda_1)t\} + \frac{1-p}{p} \frac{\lambda_2}{\lambda_1} \frac{\lambda_y + \lambda_1}{\lambda_y + \lambda_2} \exp\{-(\lambda_y + \lambda_2)t\}}{1 + \frac{1-p}{p} \frac{\lambda_2}{\lambda_1} \frac{\lambda_y + \lambda_1}{\lambda_y + \lambda_2}}$
- $\frac{S_Y^*(t)}{S_Y^*(0)} = \frac{\exp\{-(\lambda_y + \lambda_1)t\} + \frac{1-p}{p} \frac{\lambda_y + \lambda_1}{\lambda_y + \lambda_2} \exp\{-(\lambda_y + \lambda_2)t\}}{1 + \frac{1-p}{p} \frac{\lambda_y + \lambda_1}{\lambda_y + \lambda_2}}$
- $\frac{S_X^*(t)}{S_X^*(0)} \leq \frac{S_Y^*(t)}{S_Y^*(0)}$, $\Phi(t)$ is minimum at the origin, and is continuously increasing

Note that $S_X^*(0) = p \frac{\lambda_1}{\lambda_y + \lambda_1} + (1-p) \frac{\lambda_2}{\lambda_y + \lambda_2}$, and $S_Y^*(0) = p \frac{\lambda_y}{\lambda_y + \lambda_1} + (1-p) \frac{\lambda_y}{\lambda_y + \lambda_2}$.

Recalling that Z is the minimum of two variables and the indicator of which variable was smaller, the survival function of Z becomes: $P(Z > t) = p \exp\{-(\lambda_1 + \lambda_y)t\} + (1-p) \exp\{-(\lambda_2 + \lambda_y)t\}$.

The expectation and variance of Z are: $E(Z) = p \frac{1}{\lambda_y + \lambda_1} + (1-p) \frac{1}{\lambda_y + \lambda_2}$ and $VAR(Z) = \frac{1}{(\lambda_y + \lambda_2)^2} + 2p \frac{1}{\lambda_y + \lambda_1} (\frac{1}{\lambda_y + \lambda_1} - \frac{1}{\lambda_y + \lambda_2}) - p^2 (\frac{1}{\lambda_y + \lambda_1} - \frac{1}{\lambda_y + \lambda_2})^2$.

Using the above equations and $S_Y^*(0)$ we can obtain an estimation for the unknown parameters. For example $\lambda_y = S_Y^*(0)/E(Z)$ and the values of p, λ_1 and λ_2 can be obtained numerically. Note that the same solution for λ_y is obtained using the maximum likelihood method.

5 Conclusions

A new simply and elegant model has been developed for the case when the conditional subsurvival functions of the censoring variable dominates the conditional subsurvival functions of the other risk. The model agrees also with empirical time average failure rate bounds in terms of time average failure rate shape: in “failure effect” field we can have decreasing time average failure rate for X and constant for Y . In the same case of “failure effect”, we obtain a quick estimation of parameters: $\lambda_y = 0.0159, p = 0.8543, \lambda_1 = 0.2617$ and $\lambda_2 = 0.852$. These values can be used as input parameters for the MLE method. As further work, it remains to find out a model which is appropriate for the “action taken” analysis case, when the conditional subsurvival function are crossing each others.

References

- Bunea, C. Cooke, R.M. Lindqvist, Bo (2002). Maintenance study for components under competing risks. *ESREL 2002*.
- Bedford, T.J. Cooke, R.M. (2001). *Probabilistic Risk Analysis: Foundations and Methods*. Cambridge University Press, 2001.
- Cooke, R.M. (1996). “The design of reliability data bases“ Part I and II. *Reliability Engineering and System Safety* 51(2),137-146 and 209-223, 1996.
- Cooke, R.M. Bedford, T.J. (2001). Reliability databases in perspective. Due to appear in *IEEE Transaction on. Reliability*, December 2001
- Dorrepaal, J.W. Hokstad, P. Cooke, R.M. Paulsen, J.L. (1997) The effect of preventive maintenance on component reliability *Advances in Safety and Reliability*, Ed.: C.G. Soares, Proceedings of the ESREL '97 conference, 1775-1781.