

Modeling failures of repairable systems under "worse than old" assumption

Génia Babykina
Cemagref
50 avenue de Verdun
33620 Cestas
France
genia.babykina@cemagref.fr

Vincent Couallier
IMB, University of Bordeaux
145 rue Léo Saignat
33076 Bordeaux
France
couallier@sm.u-bordeaux2.fr

Yves Le Gat
Cemagref
50 avenue de Verdun
33620 Cestas
France
yves.legat@cemagref.fr

Abstract

The objective of the work is to model the failure process of a repairable system under "worse than old", or harmful repairs, assumption. The proposed model is founded on the counting process probabilistic approach and interprets harmful repairs as the accumulation of failures on the same system. The explicit form of likelihood function is given. Fixed and time-varying covariates are taken into account. Monte-Carlo simulations are carried out to assess the quality of the MLE. The method of data generation is detailed.

1 Introduction

Recurrent failures of a repairable system are commonly modeled by the non-homogeneous Poisson process (NHPP) or the renewal process (RP), implying "as bad as old" and "as good as new" repair effects respectively. Other after-repair states of a system ("better than old but worse than new", "worse than old", "better than new") can be modeled within the framework of the generalized renewal process (GRP), using for example the concept of virtual age. Refer to (Krivtsov 2007), (Lindqvist 2006) for the review of models and references and, for example, to (Lawless 1987), (Yáñez et al. 2002), (Doyen and Gaudoin 2004) for statistical inference details.

The counting process theory as defined in (Andersen et al. 1993) and as reviewed below, is widely used to construct the probability models for the mentioned processes. Let us denote $(T_i)_{i \geq 1}$ the failure times (assumed to be confounded with repair times) and $N(t) = \sum_{i \geq 1} \mathbb{1}_{T_i \leq t}$ the corresponding counting process ($N(0) = 0$). Modeling is then based on the conditional failure intensity (Andersen et al. 1993)

$$\lambda(t) = \lim_{dt \rightarrow 0} \frac{1}{dt} \mathbb{P}(N(t+dt) - N(t) = 1 \mid \mathcal{H}_{t-})$$

with \mathcal{H}_{t-} the history of the process up to time t .

The corresponding likelihood function for m failures of a system, depending on a parameter vector θ , used for statistical inference, is defined by (Andersen et al. 1993)

$$L(\theta) = \left(\prod_{j=1}^m \lambda(t_j) \right) \times \exp \left(- \sum_{j=0}^m \int_{t_j}^{t_{(j+1)}} \lambda(u) du \right) \quad (1)$$

($t_j, j = 1, \dots, m$) being the times of failures.

The "worse than old" after-repair state can be captured, for instance, by a particular value of the maintenance effect parameter in intensity-reduction models (Doyen 2005), or by stratifying the NHPP model according to the failure rank (Røstum 2000).

In this paper we propose an alternative approach to model the failure process under the "worse than old" assumption. The general framework is presented in Section 2, Section 3 deals with simulation tools. Section 4 presents the results of Monte Carlo simulations.

2 An alternative model for harmful repairs

Harmful repairs, i.e. the after-repair state of a system is worse than that just before failure, can result in accumulation of failures on a system. Starting from this assumption Le Gat (2009) proposes to link the conditional failure intensity function to the number of previous failures of a system. The resulting model is derived from the birth process theory (see (Ross 1983)) and is defined by

$$\lambda(t) = (1 + \alpha j) \delta t^{\delta-1} e^{Z(t)'\beta}, \quad (\alpha > 0)$$

with j the number of failures occurred up to time t , $Z(t)$ a $1 \times p$ covariate vector (eventually time-varying) and $\theta = \{\alpha, \delta, \beta_0, \dots, \beta_p\}$ the vector of parameters to estimate. We note $\lambda_0(t) = \delta t^{\delta-1} e^{Z(t)'\beta}$ (baseline intensity) and $\Lambda_0(s) = \int_0^s \delta u^{\delta-1} e^{Z(u)'\beta} du$ (cumulative baseline intensity).

The proposed model implies the following:

- $\alpha > 0$: instantaneous failure intensity augments with the number of previous failures (note that for $\alpha = 0$ the model is reduced to an NHPP);
- $\delta > 1$: system wears-out;
- $\exists k \neq 0 \beta_k \neq 0$: external risk factors and/or system's individual characteristics influence the failure intensity.

Le Gat (2009) shows that for a system with a known number of observed failures during a time-period $[a, b)$ (a not necessarily being the system installation time), the number of failures to occur in the future period $[b, c]$ follows the Negative Binomial distribution (NBD)

$$[N(c) - N(b) \mid N(b-) - N(a) = k] \sim \mathcal{NB} \left(\alpha^{-1} + k, \frac{e^{\alpha\Lambda_0(b)} - e^{\alpha\Lambda_0(a)} + 1}{e^{\alpha\Lambda_0(c)} - e^{\alpha\Lambda_0(a)} + 1} \right) \quad (2)$$

This result enables us to give an explicit expression for the likelihood function (1), and to estimate the parameters of the model. For example, the log-likelihood for m failures of a system during the period $[a, b]$ is as follows

$$\begin{aligned} \ln L(\theta) &= m \ln \alpha + \ln \Gamma(\alpha^{-1} + m) - \ln \Gamma(\alpha^{-1}) \\ &- (\alpha^{-1} + m) \ln(e^{\alpha\Lambda_0(b)} - e^{\alpha\Lambda_0(a)} + 1) \\ &+ \sum_{j=1}^m (\ln \lambda_0(t_j) + \alpha\Lambda_0(t_j)) \end{aligned} \quad (3)$$

where Γ is the gamma function resulting from the Negative Binomial distribution.

3 Simulation of recurrent event times

Numerical study of the MLE properties requires generation of times of events and of covariate vector. Whilst the latter is quite easy to obtain using standard probability distributions (exponential, normal, uniform, binomial, etc.), simulation of event times is less straightforward. The employed method is detailed in the present section.

The times of events $\{t_j, j \in \mathbb{N}\}$ are simulated using the inverse transform algorithm (see for example (Ross 1997)).

Consider random variable T_j , the time of j -th event, X_{j+1} the holding time between the j -th and the $(j+1)$ -th events, and u the uniform $(0, 1)$ random variable. The conditional distribution function F of X_{j+1} can be used for inversion. Specifically,

$$\begin{aligned} F(X_{j+1}) &= \mathbb{P}[X_{j+1} > x \mid T_j = t_j] \\ &= \mathbb{P}[N(t_{j+1}) - N(t_j) = 0 \mid N(t_j) = j] && \text{since } \{T_j = t_j\} \equiv \{N(t_j) = j\} \\ &= \exp(-(1 + \alpha j)[\Lambda(t_j + x) - \Lambda(t_j)]) && \text{using Eq. 2 and NBD properties} \end{aligned}$$

Successively solving $F(X_{j+1}) = u_{j+1}$ for t_{j+1} with $t_0 = 0$ we obtain the event times.

For a time-varying covariate vector $Z(t)$, $\Lambda(\cdot)$ is a piecewise function (see Fig. 1 for schematic representation) and its inversion becomes more involved. The solution for piecewise constant covariate vector is to divide the time into sub-intervals $[\tau_h, \tau_{h+1})$, where $Z(t)$ remains constant at value Z_{h+1} , and evaluate $\Lambda_k = \Lambda(\tau_k)$ at each predefined point τ_k . Generated random number u_j at each iteration j gives the interval for $\Lambda(t_j)$ and, thus, the interval for the constant covariate vector, allowing the inversion of $\Lambda(t_j)$. Specifically, the time of $(j + 1)$ -th event is calculated by

$$t_{j+1} = \left(\frac{\Lambda(t_j) - \frac{\ln(u_{j+1})}{1+\alpha_j} - \Lambda(\tau_h)}{e^{Z'_{h+1}\beta}} + \tau_h^\delta \right)^{1/\delta} \quad (4)$$

The data-generation algorithm for time-varying covariate vector is detailed below (refer to Fig. 1 for graphical support)

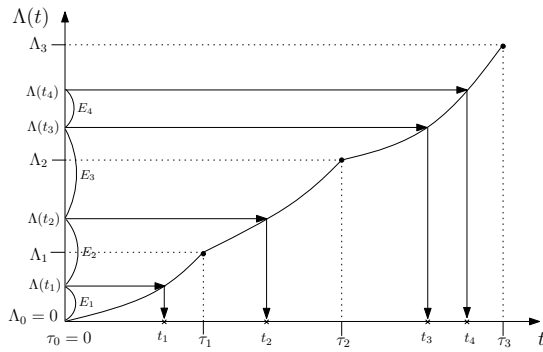


Figure 1: Data simulation schema.

1. Calculate Λ_k at predefined points $\{\tau_k\}$.
2. Set iteration counter: $j = 0$ ($t_0 = 0, \Lambda_0 = 0$).
3. Generate uniform $(0, 1)$ random variable u_{j+1} .
4. Calculate $E_{j+1} = -\frac{\ln(u_{j+1})}{(1+\alpha_j)}$, $\Lambda(t_{j+1}) = \Lambda(t_j) + E_{j+1}$.
5. Identify h for which $\Lambda(t_{j+1}) \in [\Lambda_h, \Lambda_{h+1}) \Rightarrow t_{j+1} \in [\tau_h, \tau_{h+1})$. Use Z_{h+1} in Eq. 4 to inverse $\Lambda(\cdot)$.
6. Reset counter: $j = j + 1$.
7. Repeat steps 3–6 until the end of observational period.

4 Monte Carlo study of the MLE

The simulated data consists of 10 000 simultaneously observed systems over the time period $[a, b]$. Time unit is year. Installation moments are uniformly distributed over $[a - 10, b]$. One fixed exponential covariate is considered. The time-dependent covariate is a dummy variable for the period $[b - 6, b - 5]$.

The times of events are simulated according to (4). The likelihood defined by (3) is maximized to estimate the parameter vector $\theta = \{\alpha, \delta, \beta_0, \beta_1, \beta_2\}$, β_0 corresponding to the intercept, β_1 and β_2 being the coefficients of the fixed and time-dependent covariates respectively. The initial parameter vector is $\theta = \{1.5, 1.2, -6, 0.2, 0.5\}$.

The 100-runs Monte-Carlo simulations are carried out. For comparison purposes the 10-, 20-, 30- and 40-year observational periods are considered.

The quality of the MLE for θ is assessed by means of the empirical mean square error, $\widehat{\text{MSE}}$, and the relative square error, RSE. The $\widehat{\text{MSE}}$ is calculated as the sum of the empirical variance and the squared empirical biases of θ . The RSE_j for each parameter θ_j over K Monte-Carlo runs is:

$$\text{RSE}_j = \frac{1}{K} \sum_{j=1}^K \left(\frac{\hat{\theta}_j - \theta_j}{\theta_j} \right)^2.$$

In general the estimators are better for longer observational periods, and thus, longer failure history (refer to Fig. 2 and Fig. 3 for the $\widehat{\text{MSE}}$ and the RSE respectively).

The Shapiro-Wilks test is used to empirically verify the asymptotic normality of the estimator. The p-values are presented in Table 1.

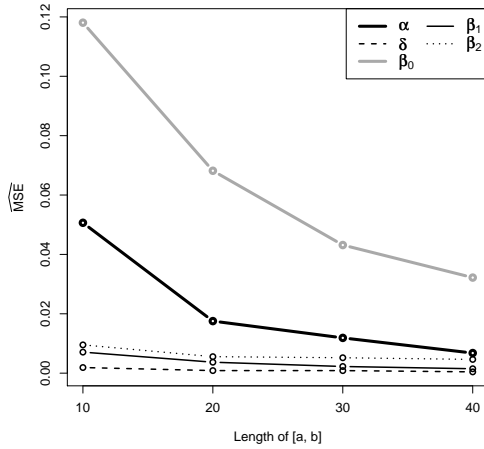


Figure 2: Empirical mean square error.

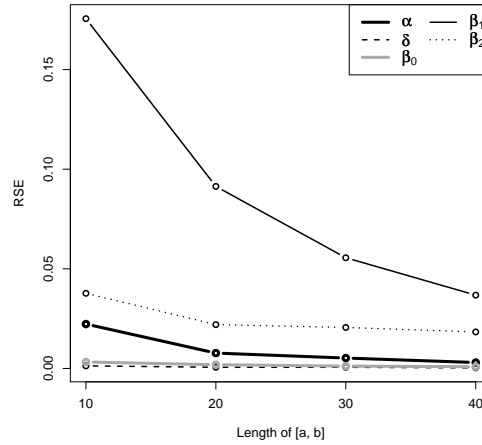


Figure 3: Relative square error.

Table 1: P-values for Shapiro-Wilks normality test.

Length of $[a, b]$	$\hat{\alpha}$	$\hat{\delta}$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$
10	0.8887	0.5286	0.0797	0.0222	0.6668
20	0.2251	0.5095	0.4498	0.7890	0.2354
30	0.0965	0.3514	0.7361	0.9396	0.5519
40	0.7057	0.8197	0.9691	0.1629	0.7938

References

- Andersen, P., O. Borgan, R. Gill, and N. Keiding (1993). *Statistical models based on counting process*. Springer-Verlag.
- Doyen, L. (2005). Repair efficiency estimation in the ARII imperfect repair model. *Modern statistical and mathematical methods in reliability*, World Scientific, Singapore, 153–168.
- Doyen, L. and O. Gaudoin (2004). Classes of imperfect repair models based on reduction of failure intensity or virtual age. *Reliability Engineering & System Safety* 84(1), 45–56.
- Krivtsov, V. (2007). Recent advances in theory and applications of stochastic point process models in reliability engineering. *Reliability Engineering & System Safety* 92, 549–551.
- Lawless, J. (1987). Regression methods for Poisson process data. *Journal of the American Statistical Association*, 808–815.
- Le Gat, Y. (2009). *Une extension du processus de Yule pour la modélisation stochastique des événements récurrents*. Ph. D. thesis, ENGREF.
- Lindqvist, B. H. (2006). On the statistical modeling and analysis of repairable systems. *Statistical Science* 21(4), 532–551.
- Ross, S. (1983). *Stochastic Processes*. John Wiley and Sons, New York.
- Ross, S. (1997). *Simulation*. Academic Press.
- Røstum, J. (2000). *Statistical Modelling of Pipe Failures in Water Networks*. Ph. D. thesis, Norwegian University of Science and Technology, Faculty of Civil Engineering.
- Yáñez, M., F. Joglar, and M. Modarres (2002). Generalized renewal process for analysis of repairable systems with limited failure experience. *Reliability Engineering & System Safety* 77(2), 167 – 180.