



A dynamic graphical model to represent complex survival distributions

Roland Donat, Laurent Bouillaut, Patrice Aknin, Philippe Leray

► To cite this version:

Roland Donat, Laurent Bouillaut, Patrice Aknin, Philippe Leray. A dynamic graphical model to represent complex survival distributions. *Advances in Mathematical Modeling for Reliability*, IOS Press, pp.17-24, 2008. hal-00412259

HAL Id: hal-00412259

<https://hal.science/hal-00412259>

Submitted on 15 Apr 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Dynamic Graphical Model to Represent Complex Survival Distributions

Roland DONAT^{a,c,1}, Laurent BOUILLAUT^a,
Patrice AKNIN^a and Philippe LERAY^b

^a *Laboratory of New Technologies, INRETS, France*

^b *Laboratory of Computer Science, Polytech' Nantes, France*

^c *Laboratory of Computer Science, Information Processing, and Systems, INSA Rouen, France*

Abstract. Reliability analysis has become an integral part of system design and operating. This is especially true for systems performing critical tasks. Moreover, recent works in reliability involving the use of probabilistic graphical models, also known as bayesian networks, have been proved relevant. This paper describes a specific dynamic graphical model, named graphical duration model (GDM), to represent complex stochastic degradation processes with any kind of state sojourn time distributions. We give qualitative and quantitative descriptions of the proposed model and detail a simple algorithm to estimate the system reliability. Finally, we illustrate our approach with a three-states system subjected to one context variable and non-exponential sojourn time distributions.

Keywords. Probabilistic graphical models, Graphical duration models, Reliability

Introduction

Reliability analysis has become an integral part of system design and operating. This is especially true for systems performing critical applications. Typically, the results of such analysis are given as inputs to a decision support tool in order to optimise the maintenance operations. Unfortunately, in most of cases, the system state cannot be evaluated exactly. This is one of the reasons which has led to the important development of probabilistic methods in reliability.

A wide range of works about reliability analysis is available in the literature. For instance in numerous applications, the aim is to model a multi-state system and therefore to capture how the system state changes over time. This problematic can be partially solved using the Markov framework. The major drawback of this approach comes from the constraint on state sojourn times which are necessarily exponentially distributed. This issue can be overcome by the use of semi-Markov models [1] which allow to consider any kind of sojourn time distributions. On the other hand, one can be interested in modeling the context impacting on the system degradation. A classic manner to address such an

¹Corresponding Author: Institut National de Recherche sur les Transports et leur Sécurité, Laboratoire des Technologies Nouvelles, 2 avenue du Général Malleret-Joinville 94114 Arcueil Cedex, France; E-mail: roland.donat@inrets.fr.

issue consists in using a Cox model or a more general proportional hazard model [2]. Nevertheless, as far as we know, it is unusual to find works considering both approaches at the same time.

Moreover, recent works in reliability involving the use of Probabilistic Graphical Models (PGMs), also known as Bayesian Networks (BNs), have been proved relevant. For instance, the authors in [3] show how to model a complex system dependability by means of PGMs. In [4], the authors explain how fault trees can be represented by PGMs. Finally in [5], the authors explain how to exploit Dynamic PGMs (DPGMs) to study the reliability of a dynamic system represented by a Markov chain. Our work aims to describe a general methodology to model the stochastic degradation process of a system, allowing any kind of state sojourn time distributions along with an accurate context description. We achieve to meet these objectives using a specific DPGM called Graphical Duration Model (GDM).

This paper is divided into four sections. Section 1 briefly describes the PGMs and DPGMs formalism. Then, section 2 introduces the GDMs by defining both its structure and its quantitative part. Section 3 depicts a simple iterative method to compute the reliability of a system represented by a GDM. Finally to illustrate our methodology, we propose to study in section 4 a three states system subjected to one context variable and non exponential duration distributions.

1. Probabilistic Graphical Models

Probabilistic Graphical Models (PGMs), also known as Bayesian Networks (BNs) [6], provide a formalism for reasoning about partial belief under conditions of uncertainty. This formalism relies on the probability theory and the graph theory. Indeed, PGMs are defined by a Directed Acyclic Graph (DAG) $\mathcal{G} = (\mathbf{X}, \mathcal{E})$ over a sequence of nodes $\mathbf{X} = (X_1, \dots, X_N)$ representing random variables that takes value from given domains $\mathcal{X}_1, \dots, \mathcal{X}_N$. The set of edges $\mathcal{E}$ encodes the existence of correlations between the linked variables. The strength of these correlations are quantified by conditional probabilities.

A PGM is a pair $(\mathcal{G}, \{P_n\}_{1 \leq n \leq N})$, where $\mathcal{G} = (\mathbf{X}, \mathcal{E})$ is a DAG and $\{P_n\}_{1 \leq n \leq N}$ denotes the set of Conditional Probability Distributions (CPDs) associated to each variable X_n and its parents. We refer to the sequence of random variables $\mathbf{X}_{\text{pa}_n}$ as the "parents" of X_n in the graph $\mathcal{G}$. Exploiting the conditional independence relationships introduced by the edges of $\mathcal{G}$, the joint probability over $\mathbf{X}$ can be economically rewritten with the product form

$$P(x_1, \dots, x_N) = \prod_{n=1}^N P_n(x_n | \mathbf{x}_{\text{pa}_n}), \quad (1)$$

where the general notation $\mathbf{x}_S$ (resp. $\mathbf{X}_S$) denotes the projection of a sequence $\mathbf{x}$ (resp. $\mathbf{X}$) over a subset of indices S .

Besides, both the DAG and the CPDs of a PGM can be automatically learnt [7] if some data or experts' opinions are available. Using PGMs is also particularly interesting because of the possibility to propagate knowledge through the network. Indeed, various inference algorithms can be used to compute marginal probability distributions over the system variables. The most classical one relies on the use of a junction tree [8]. In ad-

dition, inference in PGMs allows to take into account any variable observations (also called evidence) so as to update the marginal distributions of the other variables.

On the other hand, the classic PGM formalism is not able to represent temporal stochastic processes. Thereby the Dynamic Probabilistic Graphical Models (DPGMs, a.k.a. DBN) have been developed [9]. Strictly speaking, a DPGM is a way to extend PGM to model probability distributions over a collection of random variables $(\mathbf{X}_t)_{t \in \mathbb{N}^*} = (X_{1,t}, \dots, X_{N,t})_{t \in \mathbb{N}^*}$ indexed by the discrete-time t . A DPGM is defined as a pair $(\mathcal{M}_1, \mathcal{M}_\rightarrow)$. $\mathcal{M}_1$ is a PGM representing the prior distribution $P_1(\mathbf{X}_1) = \prod_{n=1}^N P_{n,1}(X_{n,1} | \mathbf{X}_{\text{pa}_{n,1}})$. $\mathcal{M}_\rightarrow$ is a particular PGM, called s -slice Temporal Probabilistic Graphical Model (s -TPGM) aiming to define the distribution of $\mathbf{X}_t$ given $(\mathbf{X}_\tau)_{t-(s+1) \leq \tau \leq t-1}$, where $s \geq 2$ denotes the temporal dependency order of the model. In this paper, we set $s = 2$ such that $\mathcal{M}_\rightarrow$ is a 2-TPGM representing the distribution $P_\rightarrow(\mathbf{X}_t | \mathbf{X}_{t-1}) = \prod_{n=1}^N P_{n,\rightarrow}(X_{n,t} | \mathbf{X}_{\text{pa}_{n,t}})$.

Consequently, it is possible to compute the joint distribution over random variables $(\mathbf{X}_t)_{1 \leq t \leq T}$ by simply "unrolling" the 2-TPGM until we have a sequence of length T as follows

$$\begin{aligned} P((\mathbf{X}_t)_{1 \leq t \leq T}) &= P((X_{1,t}, \dots, X_{N,t})_{1 \leq t \leq T}) = P_1 \prod_{t=2}^T P_\rightarrow(\mathbf{X}_t | \mathbf{X}_{t-1}) \\ &= \prod_{n=1}^N P_{n,1}(X_{n,1} | \mathbf{X}_{\text{pa}_{n,1}}) \prod_{t=2}^T \prod_{n=1}^N P_{n,\rightarrow}(X_{n,t} | \mathbf{X}_{\text{pa}_{n,t}}). \end{aligned}$$

Finally, as it is possible to consider a DPGM as a big unrolled PGM, these dynamic models inherit some of the convenient properties of static PGMs. On the other hand, performing inference in such models can raise some computation problems if the sequence length is too large. Consequently, specific methods have been developed to partially solve this issue [9]. In section 3, a simple inference method is described to compute the reliability of a system represented by a Graphical Duration Model (GDM).

2. Introducing the Graphical Duration Models

2.1. Graphical structure

In this article, we propose to extend the variable duration models introduced in [9] to build a comprehensive model for complex survival distributions. The 2-TPGM associated to the underlying model, called Graphical Duration Model (GDM), is depicted in figure 1. It allows to describe in a flexible and accurate way the discrete survival function of a system given its context. It relies on the two following variables : the system state X_t ; and the duration variable X_t^D describing the time spent in any system states. Moreover, a binary transition variable J_t is added to explicitly characterise the system transitions from one state to another. A collection of context variables (covariates) $\mathbf{Z}_t = (Z_{1,t}, \dots, Z_{P,t})$ can also be used to model the system context.

As shown in figure 1, the current system state X_t depends on the previous duration X_{t-1}^D and the previous system state X_{t-1} . Thus, the process generated by a GDM is then

similar to a discrete semi-markovian process [10]. Indeed, thanks to the variable X_t^D , it is possible to specify any kind of state sojourn time distributions. Consequently exploiting the powerful modelling properties of PGMs, the GDMs account for convenient, intuitive and easily generalisable tools to represent complex degradation processes.

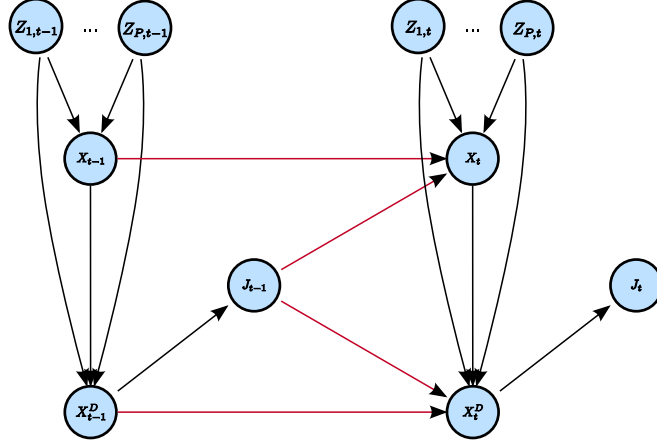


Figure 1. Representation of a GDM. The $Z_{p,t}$'s represent the system covariates. X_t is the system state and X_t^D is the duration variable in the current state. J_t is the explicit transition variable of the system.

2.2. CPDs

The following paragraphs address the specification of each CPD involved in a GDM except those about the distribution of $\mathbf{Z}_t$ since the system context modelling is strongly dependant of the application. Hence in the sequel, we suppose that the probability distribution of $\mathbf{Z}_t$ is known. In addition as it is usually the case in PGMs, all the CPDs are assumed to be finite and discrete, i.e. can be shown in the form of a multidimensional table.

First of all, we refer to $\mathcal{X} = \{x^1, \dots, x^K\}$ and $\mathcal{Z} = \{z^1, \dots, z^L\}$ as the domain associated to the system state variable and the context variable respectively. Let begin with the CPD of the initial system state given its context, namely $P(X_1|\mathbf{Z}_1)$. This CPD describe the probability for the system to start in a given state $x \in \mathcal{X}$ and given a particular context configuration $\mathbf{z} \in \mathcal{Z}$.

Then, it is necessary to define the transition CPD from one state to another. When a transition occurs at time t , i.e. if the variable $J_{t-1} = 1$, the probability that the system goes to state x^l from state x^k conditionally to the context $\mathbf{z}$ is given by the homogeneous transition matrix $A(\cdot, \mathbf{z}, \cdot)$. On the other hand, while there is no transition, i.e. if $J_{t-1} = 0$, the system deterministically remains in the previous state x^k . Therefore, the corresponding static transition matrix is reduced to identity whatever the context. As a result the CPD of $X_t, t \geq 2$ is

$$P(\underbrace{X_t = x^l}_{\text{current state}} | \underbrace{X_{t-1} = x^k}_{\text{previous state}}, \underbrace{J_{t-1} = j}_{\text{transition at time } t}, \underbrace{\mathbf{Z}_t = \mathbf{z}}_{\text{current context}}) = \begin{cases} A(k, \mathbf{z}, l) & \text{if } j = 1, \\ \delta(k = l) & \text{if } j = 0. \end{cases}$$

where δ is the characteristic function.

The initial duration CPD encodes the sojourn time distributions for each state given the context $\mathbf{z}$, such that $P(X_1^D = d | \mathbf{Z}_1 = \mathbf{z}, X_1 = x^l) = \varphi(\mathbf{z}, l, d)$, where $\varphi(\mathbf{z}, l, d)$ gives the probability to remain d time units in each state x^l given the context $\mathbf{z}$. Besides as we made the discrete and finite assumption for all the CPDs, the domain of $X_t^D, t \geq 1$ has to be discrete and finite which is not natural for a duration distribution. Basically, we overcome this issue by setting an upper time bound $D_{\max}$ large enough compared to the dynamic of the studied system. Consequently, X_t^D takes its values in the set $\mathcal{X}^D = \{1, \dots, D_{\max}\}$.

The CPD of $X_t^D, t \geq 2$ plays an analogous role except it has also to update the remaining time to spend in the current state at each sequence step. Indeed, while the remaining previous duration is greater than one (i.e. $X_{t-1}^D > 1$), the remaining sojourn time is deterministically counted down. On the other hand, when the previous remaining duration reaches the value one, a transition is triggered to occur at time t , then a sojourn time for the new current state x^l is drawn according to $\varphi(\mathbf{z}, l, \cdot)$. In other words, the CPD of $X_t^D, t \geq 2$ is defined by

$$P(\underbrace{X_t^D = d}_{\text{current remaining time}} | X_{t-1} = x^k, \underbrace{X_{t-1}^D = d'}_{\text{previous remaining time}}, \underbrace{J_{t-1} = j}_{\text{transition at time } t}, \mathbf{Z}_t = \mathbf{z}) = \begin{cases} \delta(d = d' - 1) & \text{if } j = 0 \\ \varphi(\mathbf{z}, l, d) & \text{if } j = 1 \end{cases}.$$

Note that the discrete-time assumption laid on by the DPGM formalism can be easily overcome. Indeed, authors in [11] present a survey of discrete lifetime distributions and explain how to derive usual continuous ones (e.g. exponential, Weibull, ...) in the discrete case.

Finally, J_t is the random variable characterising transitions between two different system states. More precisely, when $J_t = 1$, a transition is triggered at time t and the system state changes at time $t + 1$. The system state remains unchanged while $J_t = 0$. Besides, a transition is triggered at time t if and only if the current remaining duration reaches the value one. Consequently, the CPD of J_t is deterministic and merely defined by $P(J_t = 1 | X_t^D = d) = \delta(d = 1)$.

3. Reliability Estimation using GDM

Let assume that the set of system states $\mathcal{X}$ is partitioned into two sets $\mathcal{U}$ and $\mathcal{D}$ (i.e. $\mathcal{X} = \mathcal{U} \cup \mathcal{D}$ with $\mathcal{U} \cap \mathcal{D} = \emptyset$), respectively for "up" states and for "down" states (i.e. OK and failure situations). The discrete-time system reliability is then define as the function $R : \mathbb{N}^* \mapsto [0, 1]$ where $R(t)$ represents the probability that the system has always stayed in an up state until moment t , i.e. $R(t) = P(X_1 \in \mathcal{U}, \dots, X_t \in \mathcal{U})$. In addition, it is possible to derive some interesting metrics such as the failure rate or the MTTF (cf. [11] for details) from the reliability definition.

As the reliability estimation boils down to a probability computation, we proposed the following inference algorithm to compute $R(t)$:

- 1: Compute $P(X_1, X_1^D)$ and find out $P(X_1) = \sum_{X_1^D} P(X_1, X_1^D)$
- 2: **for** $t = 2$ to T **do**
- 3: Compute

$$P(X_t|X_{t-1}) = \sum_{X_{t-1}^D, J_{t-1}, \mathbf{Z}_t} P(X_{t-1}, X_{t-1}^D) P(J_{t-1}|X_{t-1}^D) P(\mathbf{Z}_t) P(X_t|X_{t-1}, J_{t-1}, \mathbf{Z}_t)$$

- 4: Compute $P(X_t, X_t^D)$
- 5: Find out $R(t) = P(X_1 \in \mathcal{U}) \prod_{\tau=2}^t P(X_\tau \in \mathcal{U}|X_{\tau-1} \in \mathcal{U})$
- 6: **end for**

Note that it is possible to show that the computation of the distribution $P(X_t, X_t^D)$ can be achieved by means of any classic PGM inference algorithms. Some details about this simple inference method are given in [12].

4. Application

To illustrate our approach, we use a GDM to model the behaviour of a 3-states system representing a production machine. This machine is supposed to be subjected to one covariate, namely its production speed. Hence, the resulting GDM consists of : one covariate $Z_{1,t}$ representing the speed level, "low" or "high"; the system state X_t which can be "nominal" (N) or "degraded" (D) for the up states and "failed" (F) for the down state. ; and the duration variable X_t^D where we arbitrarily set $D_{\max} = 150$ months which is large enough since our analysis is performed over only 100 months. the transition matrix A and the survival distribution φ for each state and each context are given in tables 1. Note in this example that all the sojourn times are assumed to have discrete right censored Weibull distributions denoted by $\mathcal{W}^r(\mu, \gamma)$ where μ and γ are the classic scale and shape parameters and r is the right censoring time bound assuring the finiteness of the distribution. In other words for each context z_1 and each state l , the probability of a sojourn time d is given by

$$\varphi(z_1, l, d) = [F(d-1) - F(d)] \delta(1 \leq d \leq r-1) + [1 - F(r-1)] \delta(d = r),$$

where F is the cumulative distribution function of the well-known continuous Weibull distribution with scale parameter $\mu_{z_1, l}$ and shape parameter $\gamma_{z_1, l}$.

The GDM used in this example has been implemented in MATLAB[®] environment, using the free Bayes Net Toolbox (BNT). The corresponding reliability estimations are presented in figure 2(a). The associated failure rate and MTTF are depicted in figures 2(b) and 2(c) respectively. These figures allow to characterise the behaviour of the studied system for different functioning policies controlled by the percentage of high speed production per time unit. As a consequence, useful information about the covariate effects can be deduced from such analysis and such survival analysis can be essential inputs for reliability-based maintenance decision support tools.

(a) $A(k, \text{low}, l)$				(b) $A(k, \text{high}, l)$				(c) Conditional duration distributions φ		
	N	D	F		N	D	F		state	
N	0	9/10	1/10	N	0	3/10	7/10	speed	N	D
D	0	0	1	D	0	0	1	low	$\mathcal{W}^r(30, 1)$	$\mathcal{W}^r(20, 3)$
F	0	0	1	F	0	0	1	high	$\mathcal{W}^r(20, 1)$	$\mathcal{W}^r(10, 3)$

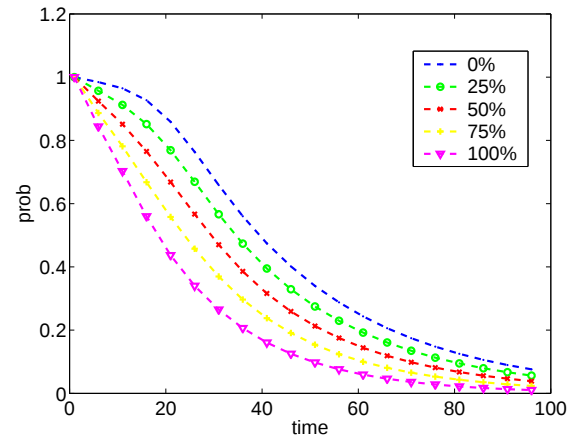
Table 1. Parameters of the GDM. $\mathcal{W}^r(\mu, \gamma)$ denotes the discrete right censored Weibull distribution. The right censoring parameter r is set to $D_{\max} = 150$ months.

5. Conclusion

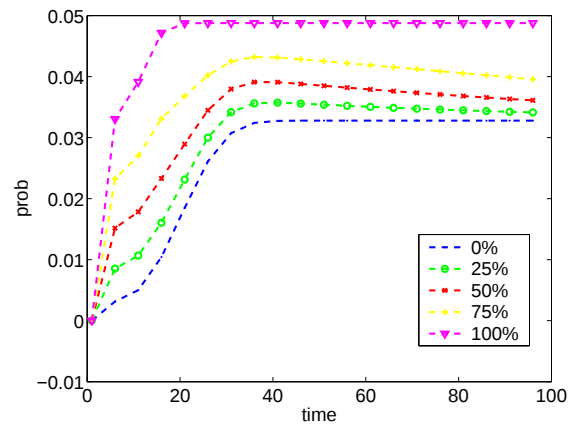
The proposed method based on the GDMs aims to study the behaviour of a complex system. Our approach turns to be a satisfying and a comprehensive solution to model and estimate the reliability of a complex system. Indeed, the proposed modelling is generic since it is possible to take into account the context of the system along with an accurate description of its survival distributions. In addition as this work is based on graphical models, the underlying approach is intuitive and easily generalisable. The encouraging results presented in this paper confirm that GDMs are competitive reliability analysis tools for practical problems. Finally in future works, we will address the problem of maintenance modelling based on system represented by GDMs.

References

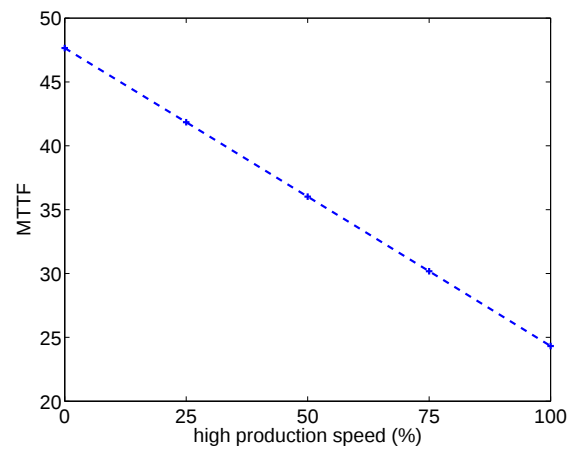
- [1] Nikolaos Limnios and Gheorghe Oprisan. *Semi-Markov Processes and Reliability*. Statistics for Industry & Technology. Springer, 2001.
- [2] R. Kay. Proportional hazard regression models and the analysis of censored survival data. *Applied Statistics*, 26(3):227–237, (1977).
- [3] H. Boudali and J. B. Dugan. A discrete-time bayesian network reliability modeling and analysis framework. *Reliability Engineering & System Safety*, 87(3):337–349, March (2005).
- [4] Helge Langseth and Luigi Portinale. Bayesian networks in reliability. *Reliability Engineering & System Safety*, 92(1):92–108, (2007).
- [5] Philippe Weber and Lionel Jouffe. Reliability modelling with dynamic bayesian networks. In *5th IFAC Symposium on fault Detection, Supervision and Safety of Technical Processes*, Washington D.C., USA, (2003).
- [6] J. Pearl. *Probabilistic Reasoning in Intelligent Systems : Networks of Plausible Inference*. Morgan Kaufmann, September 1988.
- [7] Richard E. Neapolitan. *Learning Bayesian Networks*. Prentice Hall, April 2003.
- [8] S. L. Lauritzen and D. J. Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society*, 50(2):157–224, (1988).
- [9] Kevin P. Murphy. *Dynamic Bayesian Networks : Representation, Inference and Learning*. PhD thesis, University of California, Berkeley, 2002.
- [10] V. Barbu, M. Boussemart, and N. Limnios. Discrete time semi-markov processes for reliability and survival analysis. *Communication in Statistics - Theory and Methods*, 33(11):2833–2868, (2004).
- [11] C. Bracquemond and O. Gaudoin. A survey on discrete lifetime distributions. *International Journal on Reliability, Quality, and Safety Engineering*, 10(1):69–98, (2003).
- [12] Roland Donat, Laurent Bouillaut, Patrice Aknin, Philippe Leray, and Didier Levy. A generic approach to model complex system reliability using graphical duration models. In *Proceedings of the Fifth International Mathematical Methods in Reliability Conference*, Glasgow, (2007).



(a) Reliability.



(b) Failure rate.



(c) MTTF.

Figure 2. Reliability and related metrics over time (in months) for different functioning policies.