ON THE OPTIMAL IMPORTANCE PROCESS FOR PIECEWISE DETERMINISTIC MARKOV PROCESS

H. Chraibi[1], A. Dutfoy[1], T. Galtier[2,*] and J. Garnier[3]

Abstract. In order to assess the reliability of a complex industrial system by simulation, and in reasonable time, variance reduction methods such as importance sampling can be used. We propose an adaptation of this method for a class of multi-component dynamical systems which are modeled by piecewise deterministic Markovian processes (PDMP). We show how to adapt the importance sampling method to PDMP, by introducing a reference measure on the trajectory space. This reference measure makes it possible to identify the admissible importance processes. Then we derive the characteristics of an optimal importance process, and present a convenient and explicit way to build an importance process based on theses characteristics. A simulation study compares our importance sampling method to the crude Monte-Carlo method on a three-component systems. The variance reduction obtained in the simulation study is quite spectacular.

Mathematics Subject Classification. 60K10, 90B25, 62N05.

1. Introduction

For safety and regulatory issues, nuclear or hydraulic industries must assess the reliability of their power generation systems. To do so, they can resort to probabilistic safety assessment. In recent years, dynamic reliability methods have been gaining interest, because they avoid conservative static approximations of the systems and they better capture the dynamics involved in the systems. When dealing with complex industrial systems, this kind of reliability analysis faces two main challenges: the first challenge is related to the modeling of such complex systems, the second one concerns the quantification of the reliability. Indeed as we refine the model the estimation of the reliability requires more efforts and is often challenging.

1.1. A model based on a PDMP

Due to the complexity of the systems the reliability analysis is often done through an event tree analysis [4] which requires static approximations of the system, and relies on conservative approximations. With the development of computational capacities, it is now possible to consider more accurate tools for reliability assessment.

[1] PERICLES Department, EDF lab saclay, 7 Bd Gaspard Monge, 91120 Palaiseau, France.
[2] LPSM (Laboratoire de probabilités, statistique et modélisation), Université Paris Diderot, 75205 Paris Cedex 13, France.
[3] CMAP (Centre de mathématiques appliquées), École polytechnique, 91128 Palaiseau Cedex, France.

* Corresponding author: tgaltier@gmail.com

Several attempts have been proposed to better model the dynamical processes involved in the systems. In this article, we focus on the option proposed in [15, 29], which consists in modeling the system using a piecewise deterministic Markovian process (PDMP) with boundaries.

In many industrial systems, and in particular in power generation systems, failure corresponds to a physical variable of the system (such as temperature, pressure, water level) entering a critical region. The physical variables can enter this region only if a sufficient number of the basic components of the system are damaged. In order to estimate the reliability we need an accurate model of the trajectories of the physical variables. In industrial systems, the physics of the system is often determined by ordinary or partial differential equations which depend on the statuses of the components within the system (on, off or failed). Therefore the dynamics of the physical variables changes whenever the statuses of the components are altered. Such alteration can be caused by automatic control mechanisms within the systems or failures or repairs. It is also possible that the values of physical variables impact the statuses of the components, because the failure and repair rates of the components depend on the physical conditions. In order to deal with this interplay between the physical variables and the statuses of components, we need to model their joint evolution. The vector gathering these variables is called the state of the system. To address the challenge of modeling the trajectory of the state of the system, we model the evolution of the state of the system by a piecewise deterministic Markovian process (PDMP) with boundaries. PDMPs were introduced by M.H.A Davis in [11, 12], they benefit from high modeling capacity, as they are meant to represent the largest class of Markovian processes that do not include diffusion. These processes can easily incorporate component aging, failure on demand, and delays before repairs.

For a given system, we denote its state at time t by $Z_t = (X_t, M_t)$, where X_t is the vector of the values of the physical variables, and M_t the vector gathering the statuses of all the components in the system. Throughout the paper we call X_t the position of the system, and M_t the mode of the system. $\mathbf{Z} = (Z_t)_{t \in [0, t_f]}$ represents a trajectory of the state of the system up to a final observation time t_f. We consider that the trajectories are all initiated in a state z_o.

Recall the system fails when the physical variables enter a critical region. We denote by D the corresponding region of the state space, and we denote by $\mathcal{D}$ the set of the trajectories of $\mathbf{Z}$ that pass through D. In order to estimate the reliability on the observation time t_f, we want to estimate the probability of system failure defined by

$$p = \mathbb{P}\big(\mathbf{Z} \in \mathcal{D} | Z_0 = z_o\big) = \mathbb{P}_{z_o}\big(\mathbf{Z} \in \mathcal{D}\big).$$

1.2. Accelerate reliability assessment by using importance sampling

The second challenge is that the reliability of a complex industrial system can rarely be assessed analytically, so reliability analysis often relies on simulations techniques. The company Électricité de France (EDF) has recently developed the PyCATSHOO toolbox [8, 10], which allows the simulation and the modeling of dynamic hybrid systems. PyCATSHOO bases its modeling on PDMPs. Thanks to Monte-Carlo simulation, it evaluates dependability criteria, among which is the reliability of the system. The method we present in this article is used to accelerate the reliability assessment within the PyCATSHOO toolbox.

In the context of reliable systems, crude Monte-Carlo techniques perform poorly because the system failure is a rare event. Indeed, with the Monte-Carlo method, when the probability of failure approaches zero, the number of simulations to get a reasonable precision on the relative error increases dramatically, and so does the computational time. To reduce this computational burden, one option is to reduce the number of simulations needed by using a variance reduction method. Among variance reduction techniques [3, 24], we may think of multilevel splitting techniques [5, 14] and of importance sampling techniques [2, 13, 15, 30]. A variance reduction method, inspired from particle filtering can be used on a particular case of PDMP that is a PDMP whithout boundary [28]. Unfortunately the industrial systems are often modeled by a PDMP with boundaries, and other variance reduction methods need to be designed for these cases. We choose to focus on the importance sampling technique, because: (1) the importance sampling strategy that we propose can easily be implemented

(in particular in the PyCATSHOO toolbox); (2) the results derived in this paper (in particular the reference measure and the expressions of the densities and likelihood ratios) should be useful to study multilevel splitting.

In this paper we present how to adapt the importance sampling technique for PDMP. By doing so we generalize the use of importance sampling, not only for many power generation systems, but also for any phenomenon that can be modeled by a PDMP. As PDMP generalizes numerous kinds of processes (among which are discrete Markov chains, continuous time Markov chains, compound Poisson processes or queuing systems), the scope of our work goes way beyond the study of power generation systems.

1.2.1. Prerequisite for importance sampling on PDMPs

Remember that we want to apply importance sampling to estimate the probability $p = \mathbb{P}_{z_o}(\mathbf{Z} \in \mathcal{D})$ that the system fails. In our case, importance sampling would consist in simulating from a more fragile system, while weighting the simulation outputs by the appropriate likelihood ratio. The issue is that the random variable we are considering is a trajectory of a PDMP, so we need to clarify what is the density (or the likelihood) for a trajectory of PDMP. Namely we need to introduce a reference measure for PDMP trajectories, and to identify its related densities.

In simple cases of dynamical importance sampling, this issue of the reference measure is often eluded, because the reference measure has an obvious form: it is often a product of Lebesgue measures, or a product of discrete measures. But PDMPs are very degenerate processes, their laws involve hybrid random variables which have continuous and discrete parts. In this context, it is important to ensure that we do have a reference measure that is sigma-finite to define properly the densities and the likelihood ratios.

Suppose ζ is a reference measure for $\mathbb{P}_{z_o}(\mathbf{Z} \in .)$, we denote by f the density of $\mathbf{Z}$ with respect to ζ, and we denote by g the density of an importance process $\mathbf{Z}'$ with respect to ζ. If ζ exists, and f and g satisfy $\forall \mathbf{z} \in \mathcal{D}, \ f(\mathbf{z}) \neq 0 \Rightarrow g(\mathbf{z}) \neq 0$, then we can write:

$$\mathbb{P}_{z_o}(\mathbf{Z} \in \mathcal{D}) = \mathbb{E}_f\big[\mathbb{1}_{\mathcal{D}}(\mathbf{Z})\big] = \int_{\mathcal{D}} f(\mathbf{z})\,\mathrm{d}\zeta(\mathbf{z}) = \int_{\mathcal{D}} \frac{f(\mathbf{z})}{g(\mathbf{z})} g(\mathbf{z})\,\mathrm{d}\zeta(\mathbf{z}) = \mathbb{E}_g\bigg[\mathbb{1}_{\mathcal{D}}(\mathbf{Z})\frac{f(\mathbf{Z})}{g(\mathbf{Z})}\bigg]. \tag{1.1}$$

If $\big(\mathbf{Z}'_1, \ldots, \mathbf{Z}'_{N_{\mathrm{sim}}}\big)$ is a sample of independent trajectories simulated according to an importance process with density g, then $\mathbb{P}_{z_o}(\mathbf{Z} \in \mathcal{D})$ can be estimated without bias by:

$$\hat{p}_{\mathrm{IS}} = \frac{1}{N_{\mathrm{sim}}} \sum_{i=1}^{N_{\mathrm{sim}}} \mathbb{1}_{\mathcal{D}}(\mathbf{Z}'_i)\frac{f(\mathbf{Z}'_i)}{g(\mathbf{Z}'_i)} \qquad \text{with} \quad \mathbb{V}\mathrm{ar}(\hat{p}_{\mathrm{IS}}) = \frac{\mathbb{E}_f\left[\mathbb{1}_{\mathcal{D}}(\mathbf{Z})\frac{f(\mathbf{Z})}{g(\mathbf{Z})}\right] - p^2}{N_{\mathrm{sim}}} \tag{1.2}$$

When $\mathbb{E}_f\big[\mathbb{1}_{\mathcal{D}}(\mathbf{Z})\frac{f(\mathbf{Z})}{g(\mathbf{Z})}\big] < \infty$ and the conditions above are verified, we have a central limit theorem on $\hat{p}_{\mathrm{IS}}$:

$$\sqrt{N_{\mathrm{sim}}}(\hat{p}_{\mathrm{IS}} - p) \longrightarrow \mathcal{N}(0, \sigma^2_{\mathrm{IS}}) \quad \text{where} \quad \sigma^2_{\mathrm{IS}} = \mathbb{E}_f\big[\mathbb{1}_{\mathcal{D}}(\mathbf{Z})\frac{f(\mathbf{Z})}{g(\mathbf{Z})}\big] - p^2. \tag{1.3}$$

Thus the use of importance sampling on PDMP trajectories requires the following three conditions:

(C1) We have a measure ζ on the trajectory space, and the trajectory $\mathbf{Z}$ of the system state has density f with respect to ζ

(C2) We are able to simulate trajectories according to an importance process $\mathbf{Z}'$ which has density g with respect to ζ on $\mathcal{D}$ such that $\mathbb{E}_f\big[\mathbb{1}_{\mathcal{D}}(\mathbf{Z})\frac{f(\mathbf{Z})}{g(\mathbf{Z})}\big] < \infty$.

(C3) ζ-almost everywhere in $\mathcal{D}$ we have $f(\mathbf{z}) \neq 0 \Rightarrow g(\mathbf{z}) \neq 0$

The existence of a reference measure is an important theoretical argument, but it can also be used to characterize the admissible importance processes. Knowing the reference measure ζ tells us what we can modify in the law of $\mathbf{Z}$ to obtain an importance process $\mathbf{Z}'$ with a well-defined likelihood ratio. It is a valuable information to

know to which extent we can modify the density f to get the density g, because the variance σ_{IS}^2 depends on the density g.

It is theoretically possible to design an importance sampling strategy with zero variance, indeed, it suffices to use an importance process with a density

$$g^*(\mathbf{z}) = \frac{\mathbb{1}_{\mathcal{D}}(\mathbf{z})}{p} f(\mathbf{z}). \tag{1.4}$$

In practice, however, we cannot reach this zero variance, as we do not know the value of p. The expression of g^* rather serves as a guide to build an efficient and explicit density g. Indeed we can try to choose a density g as close as possible from g^* in order to get a strong variance reduction.

1.2.2. Our contributions to the literature

Many authors have used importance sampling on particular cases of PDMP sometimes without noting it was PDMPs, see [19–22]. Sometimes, the authors using PDMPs avoid considering automatic control mechanisms which activate and deactivate components depending on the values of physical variables. Such automatic control mechanisms play an important part in power generation systems, and therefore that can not be avoided in our case. Also, the modeling of control mechanisms implies to work with a special kind of PDMPs, which are the PDMPs with boundaries. These PDMP are typically the kind for which the reference measure is complex. In [22, 25], importance sampling is used on PDMP while taking into account automatic control mechanisms but the reference measure is not clearly identified, and so far we have not found a proof that likelihood ratios involved in importance sampling on PDMP are always defined. In Section 3 we provide a reference measure for PDMP trajectories. This allows to define the likelihood ratios for PDMP trajectories and to use the importance sampling method, but also to identify the admissible importance processes. Our major contributions are presented in Section 4 where the characteristics of the optimal importance process are identified, and used to propose a convenient way to build the importance process in practice. Note that the characteristics of the optimal importance process are identified for the general case of PDMP, therefore our result can be generalized to any subclass of the PDMP process, like Markov chains, or continuous time Markov chains, or queuing systems.

1.2.3. Optimization of the variance reduction

Finding the optimal importance process is equivalent to solving the following minimization problem:

$$g^* = \underset{g}{\operatorname{argmin}} \ \mathbb{E}_f \left[\mathbb{1}_{\mathcal{D}}(\mathbf{Z}) \frac{f(\mathbf{Z})}{g(\mathbf{Z})} \right]$$

Minimizing a quantity on a density space being difficult, we usually consider a parametric family of importance densities $\{g_\alpha\}$ and look for a parameter α which yields an estimator with the smallest possible variance. Under favorable circumstances the form of the family can be determined by a large deviation analysis [16, 17, 27], but the large deviation method is difficult to adapt to PDMP with boundaries which are degenerate processes with state spaces with complicated topologies. Therefore we focus on other methods which rather try to minimize an approximation of the distance between the importance density g and the optimal one g^*. For instance, if the approximated distance happens to be $D(g, g^*) = \mathbb{E}_f\left[\frac{g^*(\mathbf{Z})}{g(\mathbf{Z})}\right]$ it is equivalent to minimize the variance of the estimator, and if we consider the Kullback-Leibler divergence so that $D(g, g^*) = \mathbb{E}_{g^*}\left[\log\left(\frac{g^*(\mathbf{Z})}{g(\mathbf{Z})}\right)\right]$, we would be using the Cross-Entropy method [13, 30]. These two options have been compared on a set of standard cases in [7]. They yielded similar results, though results obtained with the Cross-Entropy seemed slightly more stable than with the other option. In [31] the Cross-Entropy method was applied to a model equivalent to a PDMP without boundaries and showed good efficiency. Therefore we choose this method to select the parameters of

the importance process in our paper. Of course, the efficiency of this procedure strongly depends on the choice of the parametric family of importance densities.

The rest of the paper is organized as follows. Section 2 introduces our model of multi-component system based on a Piecewise deterministic Markovian process. In Section 3, we introduce a reference measure on the space of the PDMP trajectories and study the admissible importance processes. In Section 4 we present an optimal process and a clever way to build the importance process in practice. In Section 5 we apply our adaptation of the importance sampling technique on a three-component system and compare its efficiency with the Monte-Carlo technique.

2. A MODEL FOR MULTI-COMPONENT SYSTEMS BASED ON PDMP

2.1. State space of the system

We consider a system with N_c components and d physical variables. Remember we call position the vector $X \in \mathbb{R}^d$ which represents the physical variables of the system, and we call mode the vector $M = (M^1, M^2, \ldots, M^{N_c})$ gathering the statuses of the N_c components. The state of the system Z includes the position and the mode: $Z = (X, M)$.

For ease of the presentation, we consider the status of a component can be ON, or OFF, or out-of-order (noted F), so that the set of modes is $\mathbb{M} = \{ON, OFF, F\}^{N_c}$, but as long as $\mathbb{M}$ stays countable, it is possible to consider more options for the statuses of the components. For instance, one could consider different regimes of activity instead of the simple status ON, or different types of failure instead of the status F. Note that we can also deal with continuous degradations, like the size of a breach in a pipe for instance: the presence of the degradation can be included in the mode and its size in the position.

Generally, there are some components in the system which are programmed to activate or deactivate when the position crosses some thresholds. For instance, it is typically what happens with a safety valve: when the pressure rises above a safety limit, the valves opens. To take into account these automatic control mechanisms, within a mode m the physical variables are restricted to a set $\Omega_m \subset \mathbb{R}^d$, which is assumed open. We set $E_m = \{(x, m), x \in \Omega_m\}$, so that the state space is:

$$E = \bigcup_{m \in \mathbb{M}} E_m = \bigcup_{m \in \mathbb{M}} \left\{ (x, m), x \in \Omega_m \right\} \tag{2.1}$$

2.2. Flow functions

In a given mode m, *i.e.* a given combination of statuses of components, the evolution of the position is determined by an ordinary differential equation. We denote by ϕ_x^m the solution of that equation initiated in x. If we consider a position state Z_t at time t, there exists a random time $T > 0$ such that $\forall s \in [0, T)$, $X_{t+s} = \phi_{X_t}^{M_t}(s)$ and $M_{t+s} = M_t$. For an initial state $z \in E$, we can introduce the flow function Φ_z with values in E. Regarding the evolution of the trajectory after a state $Z_t = (X_t, M_t)$, the next states are locally given by the function Φ_{Z_t}:

$$\exists T > 0, \ \forall s \in [0, T),$$
$$Z_{t+s} = \Phi_{Z_t}(s) = \left(\phi_{X_t}^{M_t}(s), M_t \right) = \left(X_{t+s}, M_t \right) \tag{2.2}$$

In practice an approximation of the function ϕ_x^m can be obtained by using a numerical method solving the ordinary differential equations. For instance the PyCATSHOO toolbox can use, among others, the Runge-Kutta methods up to the fourth order.

2.3. Jumps

The trajectory of the state can also evolve by jumping. This typically happens because of control mechanisms, failures, repairs, or natural discontinuities in the physical variables. When such a jump is triggered, the current

state moves to another one by changing its mode and/or its position. We denote by $\overline{E}$ the closure of E, and $\mathcal{B}(E)$ the Borelian σ-algebra on E. We define T as the time until the next jump after t, such that the next jump occurs at time $t + T$. The destination of this jump is determined according to a transition kernel $\mathcal{K}_{Z_{t+T^-}}$ where $Z_{t+T^-} \in \overline{E}$ is the departure state of the jump. This kernel is defined by:

$$\forall B \in \mathcal{B}(E), \quad \mathbb{P}\left(Z_{t+T} \in B | Z_{t+T^-} = z^-\right) = \mathcal{K}_{z^-}(B), \tag{2.3}$$

where $\mathcal{B}(\cdot)$ indicates the Borelians of a set. Let $\forall z^- \in \overline{E}$, ν_{z^-} be a σ-finite measure on E, such that $\mathcal{K}_{z^-} \ll \nu_{z^-}$. $\forall z^- \in \overline{E}$, we define K_{z^-} as the density of $\mathcal{K}_{z^-}$ with respect to ν_{z^-}, so:

$$\forall B \in \mathcal{B}(E), \quad \mathbb{P}\left(Z_{t+T} \in B | Z_{t+T^-} = z^-\right) = \int_B K_{z^-}(z)\,\mathrm{d}\nu_{z^-}(z), \tag{2.4}$$

The kernel density must satisfy $K_{z^-}(z^-) = 0$, so that even if ν_{z^-} has a Dirac point in z^-, jumping on the departure state is impossible. Note that with this setting, the law of the arrival state of a jump can depend on the departure point z^-. For instance, if the physical variables are all continuous, then the reference measure of the transition kernel ν_{z^-} could be defined by:

$$\forall B \in \mathcal{B}(E), \qquad \nu_{z^-}(B) = \sum_{\substack{w \in \mathbb{M} \backslash \{m^-\}, \\ (x^-, w) \in E}} \delta_{(x^-, w)}(B). \tag{2.5}$$

In this example, the jump kernel is discrete:

$$\forall B \in \mathcal{B}(E), \qquad \mathcal{K}_{z^-}(B) = \sum_{\substack{w \in \mathbb{M} \backslash \{m^-\}, \\ (x^-, w) \in E}} \mathbb{P}\left(Z_{t+T} = (x^-, w) | Z_{t+T^-} = (x^-, m^-)\right) \delta_{(x^-, w)}(B), \tag{2.6}$$

and it is generally the case, but one can imagine some cases where the kernel include a continuous part. For instance consider that the physical variables have two dimensions, the first corresponding to pressure on a concrete structure, and the second to the size of a crack in the structure. One can consider that the crack length increase in a jerky way, and that the amplitude of the increase is random and has a continuous law. For a jump triggered from a state $z^- = \left((x_1^-, x_2^-), m^-\right) \in E$ we could have:

$$\forall B \in \mathcal{B}(E), \qquad \nu_{z^-}(B) = \int_{\left\{y_2 > 0 \,\middle|\, \left((x_1^-, y_2), m^-\right) \in B\right\}} \mathrm{d}\mu_{Leb}(y_2) \tag{2.7}$$

where $\mu_{Leb}(\cdot)$ corresponds to the Lebesgue measure, and

$$\forall B \in \mathcal{B}(E), \qquad \mathcal{K}_{z^-}(B) = \int_B K_{z^-}(z)\,\mathrm{d}\nu_{z^-}(z) = \int_{\left\{y_2 > 0 \,\middle|\, \left((x_1^-, y_2), m^-\right) \in B\right\}} K_{z^-}\left(\left((x_1^-, y_2), m^-\right)\right) \mathrm{d}\mu_{Leb}(y_2) \tag{2.8}$$

with $K_{z^-}\left(\left((x_1, x_2), m\right)\right) = 0$. We think, that the cases of non-discrete jump kernel should be rather rare in the reliability analysis field, but PDMPs are also used in other fields, like finance, where non-discrete jump kernel could be more common and for which the use of importance sampling can be of interest too [26]. That is why we keep the most general form of PDMP, which can handle any type of jump kernel.
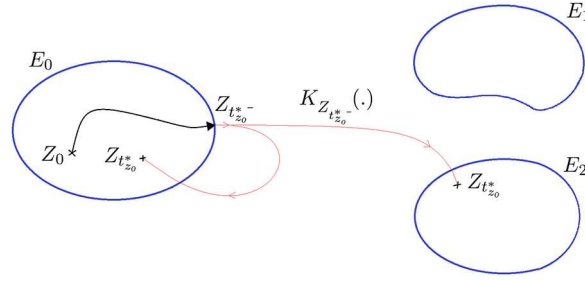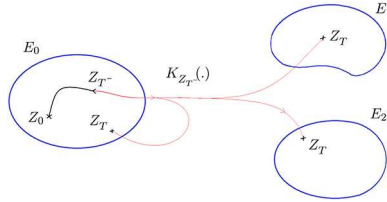
FIGURE 1. A jump at boundary.

FIGURE 2. A spontaneous jump.

2.4. Jump times

Now, assuming that $Z_t = z$, we present the law of the time until the next jump after t, which is denoted by T.

2.4.1. Jumps at boundaries

For $m \in \mathbb{M}$, let $\partial \Omega_m$ be the boundary of Ω_m. The boundary of the set E_m is the set $\partial E_m = \{(x, m), x \in \partial \Omega_m\}$. For $z = (x, m) \in E$, we define $t_z^* = \inf\{s > 0, \Phi_z(s) \in \partial E_m\}$ the time until the flow hits the boundary. We take the convention $t_z^* = +\infty$ if $\{s > 0, \Phi_z(s) \notin E_m\} = \emptyset$. Assume that the system starts in state $z = (x, m) \in E$. When the flow leads the position out of its restricted set Ω_m, *i.e.* the state touches ∂E_m, an automatic jump is triggered (see the scheme in 1), and $T = t_z^*$.

Boundaries can be used to model automatic control mechanisms, or any automatic change in the status of a component. For instance in a dam, if the water level X reaches a given threshold $x_{\max}$ the evacuation valve automatically opens to avoid overflow. If $M = C$, O, F represent respectively the modes where the valve is closed, or opened, or failed, this control system could be modeled by setting $\Omega_C = (0, x_{\max})$ and $K_{(x_{\max}, C)}(\{(x_{\max}, O)\}) = 1$.

2.4.2. Spontaneous jumps

The trajectory can also jump to another state when a random failure or a repair occurs (see Fig. 2). The distribution of the random time at which it happens is usually modeled through a state-related intensity function $\lambda : E \to \mathbb{R}_+$. For $z \in E$, $\lambda(z)$ represents the instantaneous probability (also called hazard rate) of having a failure or a repair at state z. If $Z_t = z$ and T is the duration until the next jump, $\forall s < T$ we have $Z_{t+s} = \Phi_z(s)$. To simplify the notations in the following, we introduce the time-related intensity λ_z such that $\lambda_z(s) = \lambda(\Phi_z(s))$ and $\Lambda_z(s) = \int_0^s \lambda(\Phi_z(u)) \mathrm{d}u$. If $\mathbb{P}_z(\cdot)$ is the probability of an event knowing $Z_t = z$, we have:

$$\mathbb{P}_z(T \leq s) = \begin{cases} 1 - \exp\left[-\Lambda_z(s)\right] & \text{if } s < t_z^*, \\ 1 & \text{if } s \geq t_z^*. \end{cases} \tag{2.9}$$
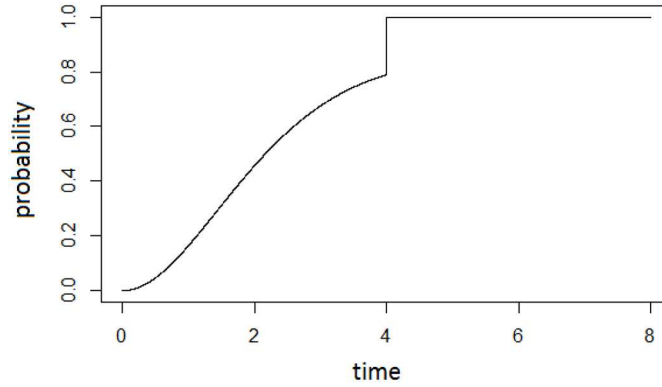
FIGURE 3. An example of the cumulative distribution function of T, where $x \in \mathbb{R}^+$, $z = (x, m) \in E$, $\Phi_z(t) = (x + t, m)$, $\lambda(z) = \frac{x(5-x)}{12}$, and $t_z^* = 4$.

The law of T has a continuous and a discrete part (see Fig. 3). As there is a discontinuity at t_z^* in the cumulative distribution function of T, its reference measure T must include a Dirac point at t_z^* and therefore depends on z. We denote μ_z the reference measure of T such that:

$$\forall B \in \mathcal{B}\big([0, t_z^*]\big), \qquad \mu_z\big(B\big) = \mu_{Leb}\big(B \cup [0, t_z^*)\big) + \delta_{t_z^*}\big(B\big). \tag{2.10}$$

This measure will be useful to define the dominant measure ζ in Section 3. It also allows to reformulate the law of T with an integral form:

$$\mathbb{P}_z(T \leq s) = \int_{(0, s]} \Big(\lambda_z(u)\Big)^{\mathbb{1}_{u < t_z^*}} \exp\Big[-\Lambda_z(u)\Big] \mathrm{d}\mu_z(u). \tag{2.11}$$

2.5. A link between jump rate and the hazard rates of the possible transitions

Note that the equation (2.9), or (2.11), gives the time of the next jump, but it does not tell whether the transition is a failure, or a repair, or an automatic control mechanism. The type of the transition triggered is determined by the transition Kernel $\mathcal{K}_{Z_{t+T}^-}$. For each jump, we consider that there can be a countable number of possible transitions. Each type of transition is indexed by a number in the countable set J. When $Z_{t+T}^- \in \partial E$, $\mathcal{K}_{Z_{t+T}^-}$ can take an arbitrary form, but when $Z_{t+T}^- \in E$, the density of the kernel $K_{Z_{t+T}^-}$ is linked to the hazard rates of the possible transitions, as shown in the following. If a transition is indexed by $j \in J$, we denote by T^j the time between t and the next occurrence of this transition, taking by convention $T^j = +\infty$ if the transition does not occur. This way the time of the next jump satisfies:

$$T = \min\big[\{T^j, \forall j \in J\} \cup \{t_z^*, \}\big]. \tag{2.12}$$

Let $\lambda^j : E \to \mathbb{R}_+$ be its associated state-related intensity function, such that:

$$\forall s < t_z^*, \quad \mathbb{P}_z(T^j \leq s) = 1 - \exp\Big[-\int_0^s \lambda^j\big(\Phi_z(u)\big)\mathrm{d}u\Big]. \tag{2.13}$$

For instance, if the transition j corresponds to a failure the function λ^j is the associated failure rate, and respectively if the transition j corresponds to a repair, the function λ^j is the associated repair rate. Knowing $Z_t = z_t = (x, m)$, and therefore, knowing the path given by ϕ_x^m that the positions are following, we make the

assumption that the times T^j are independent. This assumption is true if the position gathers all the variables affecting the different types of transitions when the system is in mode m. According to the equation (2.12), this conditional independence implies that:

$$\forall z^- \in E, \qquad \lambda(z^-) = \sum_{j \in J} \lambda^j(z^-). \tag{2.14}$$

Note the equations (2.14) is only valid when the departure state z^- is not on a boundary. We denote by $B^j_{z^-}$ the possible arrival states of a jump when the transition j is triggered and when the departure state is $z^- \in E$. We assume that the different types of transition are exclusive, meaning that for $i \neq j$ we have $B^i_{z^-} \cap B^j_{z^-} = \emptyset$. Then the probability of triggering the transition i from the departure state z^- is $\mathcal{K}_{z^-}(B^i_{z^-})$, and we have:

$$\forall z^- \in E, \qquad \mathcal{K}_{z^-}(B^i_{z^-}) = \frac{\lambda^i(z^-)}{\lambda(z^-)}. \tag{2.15}$$

Similarly the equation (2.15) is also valid only when the departure state z^- is not on a boundary. When $z^- \in \partial E$ there is no link between λ and $\mathcal{K}_{z^-}$.

2.6. Generate a trajectory

In order to generate a realization of the PDMP, one can follow these steps [11, 12, 15]:

1. Start at $t = 0$ with state $Z_t = z_t$
2. Generate T the time until the next jump using (2.9) or (2.11), and (2.14)
3. Follow the flow Φ until T using (2.2)
4. Generate $Z_{t+T} = z_{t+T}$ the arrival state of the jump knowing the departure state is $Z_{t+T^-} = \Phi_z(T)$ using (2.4)
5. Taking $t := t + T$, repeat steps 1 to 5 until a trajectory of size t_f is obtained

2.7. Example

As an example of a system, we consider a room heated by three identical heaters. X_t represents the temperature of the room at time t. x_e is the exterior temperature. β_1 is the rate of the heat transition with the exterior. β_2 is the heating power of each heater. The differential equation giving the evolution of the position (*i.e.* the temperature of the room) has the following form:

$$\frac{\mathrm{d}\,X_t}{\mathrm{d}t} = \beta_1(x_e - X_t) + \beta_2 \mathbb{1}_{M^1_t \text{ or } M^2_t \text{ or } M^3_t = ON}.$$

The heaters are programmed to maintain the temperature within an interval $(x_{\min}, x_{\max})$ where $x_e < 0 < x_{\min}$. Heaters can be on, off, or out-of-order, so $\mathbb{M} = \{ON, OFF, F\}^3$. We consider that the three heaters are in passive redundancy in the sense that: when $X \leq x_{\min}$ the second heater activates only if the first one is failed, and the third one activates only if the two other heaters are failed. When a repair of a heater occurs, if $X \leq x_{\min}$ and all other heaters are failed the heater status is set to ON, else the heater status is set to OFF. To handle the programming of the heaters, we set $\Omega_m = (-\infty, x_{\max})$ when all the heaters are failed $m = (F, F, F)$ or when at least one is activated, otherwise we set $\Omega_m = (x_{\min}, x_{\max})$.

Due to the continuity of the temperature, the reference measure for the kernel is $\forall B \in \mathcal{B}(E), \nu_{(x,m)}(B) = \sum_{m^+ \in \mathbb{M} \setminus \{m\}} \delta_{(x,m^+)}(B)$. On the top boundary in $x_{\max}$, heaters turn off with probability 1. On the bottom boundary in $x_{\min}$, when a heater is supposed to turn on, there is a probability $\gamma = 0.01$ that the heater will
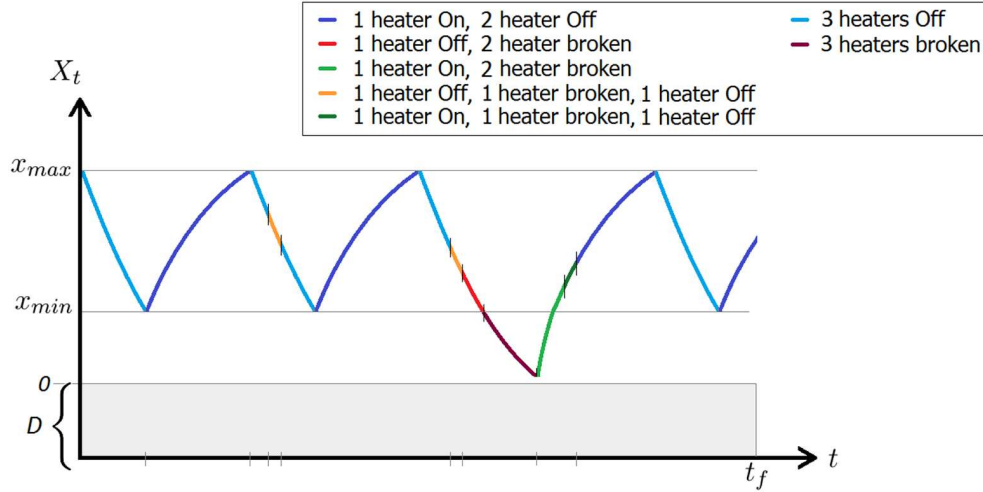
FIGURE 4. A possible trajectory of the heated-room system (the mode is represented with colors).

fail on demand. So, for instance, if $z^- = \left(x_{\min}, (OFF, F, OFF)\right)$, we have $K_{z^-}\left(x_{\min}, (ON, F, OFF)\right) = 1 - \gamma$, and $K_{z^-}\left(x_{\min}, (F, F, ON)\right) = \gamma(1 - \gamma)$, and $K_{z^-}\left(x_{\min}, (F, F, F)\right) = \gamma^2$.

For the spontaneous jumps that happen outside boundaries, we consider the position is not modified during the jumps and, if the transition j corresponds to the failure of a heater, then, for $z^- = (x^-, m^- m) \in E$, $\lambda^j(z^-) = 0.0021 + 0.00015 \times x^-$ and, if the transition j corresponds to a repair, then, for $z^- = (x^-, m^- m) \in E$, $\lambda^j(z^-) = 0.2$. A possible trajectory of the state of this system is depicted in Figure 4. Here the system failure occurs when the temperature of the room falls below zero, so $D = \{(x, m) \in E, x < 0\}$.

3. A REFERENCE MEASURE FOR TRAJECTORIES

We have seen in Section 2.3 that when the position is restricted to a bounded set in some modes, the time to the next jump can be a hybrid random variable. We have to be cautious when considering the density of a trajectory of a PDMP for several reasons: first the reference measure for the times between the jumps is a mixture of Dirac and Lebesgue measures, secondly these hybrid jumps may occur multiple times and in a nested way in the law of the trajectory of PDMP. Indeed, with these mixtures of Dirac and Lebesgue measures involved, the existence of a sigma-finite reference measure on the trajectory space is not obvious, yet it is mandatory to properly define the density of a trajectory. The existence of a reference measure is therefore crucial, because it preconditions the existence of the likelihood ratio needed to apply the importance sampling method.

We begin this Section by introducing a few notations: For a trajectory $\mathbf{Z}$ on the observation interval $[0, t_f)$, we denote by N the number of jumps before t_f, and by S_k the time of the k-th jump with the conventions $S_0 = 0$, and $S_{N+1} = t_f$. $\forall k < N$, $T_k = S_{k+1} - S_k$ is the duration between two jumps and $T_N = S_{N+1} - S_N = t_f - S_N$ is the remaining duration between the last jump and t_f. One can easily verify that the sequence of the $(Z_{S_k}, S_{k+1} - S_k)$ is a Markov chain: it is called the embedded Markov chain of the PDMP [11], see Figure 5.

3.1. The law of the trajectories

The main idea in building the law of the trajectory $\mathbf{Z}$ is to summarize the trajectory by the truncated embedded Markov chain of the process: the vector $\left(Z_{S_0}, T_0, \ldots, Z_{S_N}, T_N\right)$. This vector is also called the skeleton of the trajectory. As the trajectory is piecewise deterministic, we only need to keep the states of the arrivals of the jumps and the durations between the jumps to describe the trajectory. If we have the vector $\left(Z_{S_k}, T_k\right)_{k \leq N}$ then we have enough information to reconstruct the trajectory using (2.2) because we know the flow function
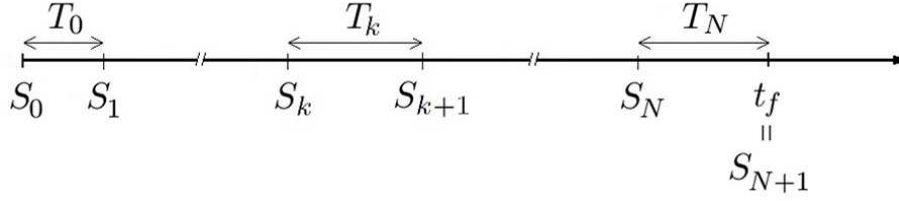
FIGURE 5. Notations.

Φ. Noting Θ the map that changes $\mathbf{Z}$ into $\left(Z_{S_k}, T_k\right)_{k \leq N}$, the law of $\mathbf{Z}$ can be defined as the image law of $\left(Z_{S_k}, T_k\right)_{k \leq N}$ through Θ. We denote by $\mathbf{E}$ the set of the trajectories defined on $[0, t_f)$. For $n \in \mathbb{N}$, let $A_n = \left\{ \left(z_{s_k}, t_k\right)_{k \leq n} \in (E \times \mathbb{R}_+^*)^n, \sum_{i=0}^{n} t_i = t_f \right\}$, so that $\Theta^{-1}(A_n)$ is the set of the trajectories including n jumps. The sets $(\Theta^{-1}(A_n))_{n \in \mathbb{N}}$ form a partition of $\mathbf{E}$. The sets $(A_n)_{n \in \mathbb{N}}$ form a partition of the set of the skeletons.

We can get the law of $\left(Z_{S_k}, T_k\right)_{k \leq N}$, by using the dependencies between its coordinates. Thanks to (2.11) and (2.4) we can get the density of T_k knowing Z_{S_k} with respect to $\mu_{Z_{S_k}}$, and the density of $Z_{S_{k+1}}$ knowing $\left(Z_{S_k}, T_k\right)$ with respect to $\nu_{Z_{S_{k+1}^-}}$, where $Z_{S_{k+1}^-} = \Phi_{Z_{S_k}}(T_k)$:

$$f_{T_k | Z_{S_k} = z}(u) = \left(\lambda_z(u)\right)^{\mathbb{1}_{u < t_z^*}} \exp\left[-\Lambda_z(u)\right], \tag{3.1}$$

$$f_{Z_{S_{k+1}} | Z_{S_k}, T_k}(z) = K_{Z_{S_{k+1}^-}}(z). \tag{3.2}$$

Using the Markov structure of the sequence $\left(Z_{S_k}, T_k\right)_{k \leq N}$, the law of $\left(Z_{S_k}, T_k\right)_{k \leq N}$ can be expressed as an integral of the product of the conditional densities given by (3.1) and (3.2).

We define the σ-algebra $\mathcal{S}$ on the set of the possible values of $\left(Z_{S_k}, T_k\right)_{k \leq N}$ as the σ-algebra generated by the sets in $\bigcup_{n \in \mathbb{N}^*} \mathcal{B}\left(\left\{ \left(z_{s_k}, t_k\right)_{k \leq n} \in (E \times \mathbb{R}_+^*)^n, \sum_{i=0}^{n} t_i = t_f \right\}\right)$.

Definition 3.1. The law of the trajectory is then defined as follows, for $B \in \mathcal{S}$

$$\mathbb{P}_{z_o}\left(\mathbf{Z} \in \Theta^{-1}(B)\right) = \int_B \prod_{k=0}^{n} \left(\lambda_{z_k}(t_k)\right)^{\mathbb{1}_{t_k < t_{z_k}^*}} \exp\left[-\Lambda_{z_k}(t_k)\right] \prod_{k=1}^{n} K_{z_k^-}(z_k)$$
$$\times\, d\delta_{t_n^*}(t_n)\, d\nu_{z_n^-}(z_n)\, d\mu_{t_{z_{n-1}}^*}(t_{n-1})\, \ldots\, d\nu_{z_1^-}(z_1)\, d\mu_{t_{z_o}^*}(t_0), \tag{3.3}$$

where $z_j^- = \Phi_{z_{j-1}}(t_{j-1})$, and $t_n^* = t_f - \sum_{i=0}^{n-1} t_i$.

Note that, depending on the set B, n can take different values in the equation (3.3). Implicitly, the equation (3.3), states that:

$$\mathbb{P}_{z_o}\left(\mathbf{Z} \in \Theta^{-1}(B)\right) = \mathbb{P}_{z_o}\left(\mathbf{Z} \in \Theta^{-1}\left(\bigcup_{n \in \mathbb{N}} B \cap A_n\right)\right)$$
$$= \sum_{n \in \mathbb{N}} \mathbb{P}_{z_o}\left(\mathbf{Z} \in \Theta^{-1}(B \cap A_n)\right)$$

$$= \sum_{n \in \mathbb{N}} \int_{B \cap A_n} \prod_{k=0}^{n} \left(\lambda_{z_k}(t_k) \right)^{\mathbb{1}_{t_k < t^*_{z_k}}} \exp \left[-\Lambda_{z_k}(t_k) \right] \prod_{k=1}^{n} K_{z_k^-}(z_k)$$

$$\times \, \mathrm{d}\delta_{t^*_n}(t_n) \, \mathrm{d}\nu_{z_n^-}(z_n) \, \mathrm{d}\mu_{t^*_{z_{n-1}}}(t_{n-1}) \, \dots \, \mathrm{d}\nu_{z_1^-}(z_1) \, \mathrm{d}\mu_{t^*_{z_o}}(t_0). \tag{3.4}$$

Also note that with our construction, this is a probability law on the space of the trajectories that satisfy (2.2), not on the set of all the trajectories with values in E.

3.2. The dominant measure and the density

Definition 3.2. We define the measure ζ so that

$$\zeta(\Theta^{-1}(B)) = \int_{(z_k, t_k)_{k \leq n} \in B} \mathrm{d}\delta_{t^*_n}(t_n) \, \mathrm{d}\nu_{z_n^-}(z_n) \, \mathrm{d}\mu_{t^*_{z_{n-1}}}(t_{n-1}) \, \dots \, \mathrm{d}\nu_{z_1^-}(z_1) \, \mathrm{d}\mu_{t^*_{z_o}}(t_0). \tag{3.5}$$

Note that, like in equation (3.3), in the equation (3.5) n can take different values depending on the set B.

Theorem 3.3. *If $\exists C > 0, \forall z \in \overline{E}, \ \nu_z(E) < C$ and $t_f < \infty$, then ζ is a σ-finite measure. By Radon-Nikodym theorem, the density of a trajectory $\mathbf{z} = \Theta\big((z_0, t_0), \dots, (z_n, t_n)\big)$ with respect to the measure ζ is*

$$f(\mathbf{z}) = \prod_{k=0}^{n} \left(\lambda_{z_k}(t_k) \right)^{\mathbb{1}_{t_k < t^*_{z_k}}} \exp \left[-\Lambda_{z_k}(t_k) \right] \prod_{k=1}^{n} K_{z_k^-}(z_k), \tag{3.6}$$

where n is the number of jumps in the trajectory $\mathbf{z}$.

The proof of Theorem 3.3 is given in Appendix A.

Note that it is always possible to choose the measures ν_{z^-} so they are all bounded by the same constant. Indeed the transition kernel is itself bounded by 1, as it is a probability measure. So, to get a measure ζ that is σ-finite, we can simply take the measures ν equal to the transition kernel, so the densities can be properly defined when the observation time t_f is finite.

3.3. Admissible importance processes

Recall that an admissible importance process is any process whose law is absolutely continuous with respect to ζ (condition C2), and which has a density g with respect to ζ satisfying $\forall \mathbf{z} \in \mathcal{D}, \ f(\mathbf{z}) \neq 0 \Rightarrow g(\mathbf{z}) \neq 0$ (condition C3). In this Section, we clarify the previous statement, and we identify to which extent we can modify the original process to obtain an admissible importance process. Throughout the rest of paper we denote the elements relative to this importance process with a $'$, except for its density that is denoted by g.

Our first remark is that condition C2 implies that the realizations of the importance process must satisfy equation (2.2). Indeed, the measure ζ involves the transformation Θ which uses the equation (2.2) to rebuild a trajectory from a skeleton. Consequently, the importance process has to piecewise follow the same flows as the original process. Similarly to the original process the importance process jumps to a new state for each change of flow. To ensure condition C2, the law of the T'_k has to be dominated by $\mu_{Z'_{S'_k}}$, and the law of $Z'_{S'_{k+1}}$ has to be dominated by $\nu_{Z'^-_{S_k}}$. This means that the boundaries of the Ω_m's and the set of the possible arrivals of a jump remain unchanged. So the modification of the original process focuses on the timing and nature of changes of modes, *i.e.* the laws of the jumps.

To generate an importance process, we keep generating trajectories by successively generating the arrival state of a jump ($Z'_{S'_k}$) and the time until the next jump (T'_k). As there is no requirement for the importance process to be Markovian, we consider that the law of a point of the trajectory Z'_t depends on the past values

of states. As the states follow the flows piecewisely, it is equivalent to say that the law of $Z'_{S'_k}$ can depend on $(Z'_{S'_i}, T'_i)_{i<k}$, and that the law of T'_k can depend on $(Z'_{S'_i}, T'_i)_{i<k}$ and $Z'_{S'_k}$. For a jump time S'_k, we denote $\underline{Z}'_{S'_k} = \big((Z'_{S'_i}, T'_i)_{i<k}, Z'_{S'_k}\big)$, and we denote by $\lambda'_{\underline{z}_k}(\cdot)$ the intensity function associated to T'_k when $\underline{Z}'_{S'_k} = \underline{z}_k$. We have:

$$\forall t \in (0, t^*_{z_k}], \quad \mathbb{P}(T'_k \leq t | \underline{Z}'_{S'_k} = \underline{z}_k) = \int_{(0,t]} \Big(\lambda'_{\underline{z}_k}(u)\Big)^{\mathbb{1}_{u < t^*_{z_k}}} \exp\Big[-\Lambda'_{\underline{z}_k}(u)\Big] \mathrm{d}\mu_{z_k}(u) \tag{3.7}$$

Noting $\underline{Z}'_{S'_k-} = \big((Z'_{S'_i}, T'_i)_{i<k-1}\big)$ and $K'_{\underline{z}^-}$ the importance kernel when $\underline{Z}'_{S'_k-} = \underline{z}_k{}^-$, we have:

$$\forall B \in \mathcal{B}(E), \quad \mathbb{P}(Z'_{S'_k} \in B | \underline{Z}'_{S'_k-} = \underline{z}_k{}^-) = \int_B K'_{\underline{z}_k^-}(z) \mathrm{d}\nu_{z_k^-}(z) \tag{3.8}$$

Notice that the intensity function $\lambda'_{\mathbf{z}_s, s}$ in equation (3.7) does not have to be of the form $\lambda' \circ \phi_{z_s}$, where λ' is a positive function on E. This means that at the time $S'_k + t$, the intensity does not depend only on the state $Z'_{S'_k+t}$ as it would be the case if $\mathbf{Z}'$ were a PDMP. So, in the importance process, we consider that the intensity can depend on the arrival state of the last jump and on previous pairs $(Z'_{S'_i}, T'_i)$. Therefore the importance process can be seen as a piecewise deterministic process (PDP) which is not necessarily Markovian.

For condition C3 to be satisfied almost everywhere we can impose that almost everywhere for any $z_k \in E$, and $z_k^- \in \overline{E}$, and $t \in (0, t^*_{z_k}]$:

$$\mathbb{E}\big[\mathbb{1}_{\mathcal{D}}(\mathbf{Z}) \big| \underline{Z}_{S_k} = \underline{z}_k\big] > 0, \text{ and } K_{z_k^-}(z_k) > 0 \Rightarrow K'_{\underline{z}_k^-}(z_k) > 0$$
$$\mathbb{E}\big[\mathbb{1}_{\mathcal{D}}(\mathbf{Z}) \big| \underline{Z}_{S_{k+1}^-} = (\underline{z}_k, t)\big] > 0, \text{ and } \lambda_{z_k}(t) > 0 \Rightarrow \lambda'_{\underline{z}_k}(t) > 0.$$

Unfortunately with complex systems, the set $\mathcal{D}$ can be very hard to manipulate, and we do not always know if $\mathbb{E}\big[\mathbb{1}_{\mathcal{D}}(\mathbf{Z}) \big| \underline{Z}_{S_k} = \underline{z}_k\big]$ or $\mathbb{E}\big[\mathbb{1}_{\mathcal{D}}(\mathbf{Z}) \big| \underline{Z}_{S_k} = \underline{z}_k, T_k = t\big]$ are positive. So in practice we often only use the following sufficient condition which states that for almost any $z_k \in E$, and $z_k^- \in \overline{E}$, and $t \in (0, t*_{z_k}]$:

$$K_{z_k^-}(z_k) > 0 \Rightarrow K'_{\underline{z}_k^-}(z_k) > 0$$
$$\lambda_{z_k}(t) > 0 \Rightarrow \lambda'_{\underline{z}_k}(t) > 0.$$

4. Optimal and practical importance process

4.1. Practical importance processes and notations

We will see in Section 4.2 that we can restrict the search of an efficient importance process within a special class of processes without any loss in efficiency, because an optimal importance process (giving an estimator with zero variance) belongs to this special class.

The processes of this class are defined through the expressions (3.7) and (3.8) but they do not use all the information contained in $\underline{z}_k$ and $\underline{z}_k^-$. The jump rates $\lambda'_{\underline{z}_k}(t)$ depend only on three variables which are: the current state $Z_{S_k+t} = \Phi_{z_k}(t)$, the time $t_f - (s_k + t)$ left before t_f, and the indicator $\mathbb{1}_{\tau_D \leq s_k+t}$ which tells if the system failure has already happened. The kernels $K'_{\underline{z}_{k+1}^-}$ depend only on three variables, which are: the current departure state $z_{k+1}^- = \Phi_{z_k}(t_k)$, the time $t_f - s_{k+1}$ left before t_f, and the indicator $\mathbb{1}_{\tau_D \leq s_{k+1}}$.

So, to ease the presentation of such jump rates and transition kernels, we slightly modify the state space by adding an active boundary at the boundary of D and we add a coordinate on the mode which indicates if the trajectory has already visited D. The state now becomes $Z = \big(X, (M, M_D)\big)$ where $M_D = 0$ if D has not been

visited, and 1 if it has. This way, for any time t we have $Z_t = (X_t, (M_t, \mathbb{1}_{\tau_D \le t}))$. For instance, with the heated-room system the set of modes becomes $\mathbb{M} = \{ON, OFF, F\}^3 \times \{0, 1\}$. The kernel K_{Z^-} is unchanged when $M_D^- = M_D^+$, and is null when $M_D^- \ne M_D^+$, except at the boundary of D where $K_{(0,(F,F,F,0))}\big(0, (F, F, F, 1)\big) = 1$.

The three variables that determine the jump rates and kernels of the processes of the special class can now be identified by the current state and the current time. Therefore, we now consider importance processes with jump rate $\lambda'_{z_k, s_k}(t)$ and transition kernel $K'_{z_k^-, s_k}$. Such processes have the following laws of jump times and jump arrivals:

$$\forall t \in (0, t_{z_k}^*], \quad \mathbb{P}(T_k' \le t | Z_{S_k'}' = z_k, S_k' = s_k)$$
$$= \int_{(0,t]} \left(\lambda'_{z_k, s_k}(u) \right)^{\mathbb{1}_{u < t_{z_k}^*}} \exp\left[-\Lambda'_{z_k, s_k}(u) \right] \mathrm{d}\mu_{z_k}(u) \tag{4.1}$$

$$\forall B \in \mathcal{B}(E), \quad \mathbb{P}(Z_{S_k'}' \in B | Z_{S_k'-}' = z_k{}^-, S_k' = s_k) = \int_B K'_{z_k^-, s_k}(z) \mathrm{d}\nu_{z_k^-}(z). \tag{4.2}$$

Note that the class of processes that can be defined by (4.1) and (4.2) is included in the class of admissible importance processes. Thanks to the new definition of the states, the conditional expectations $\mathbb{E}\big[\mathbb{1}_\mathcal{D}(\mathbf{Z}) \big| \underline{Z}_{S_k}, T_k \ge t\big]$ are equal to the conditional expectations $\mathbb{E}\big[\mathbb{1}_\mathcal{D}(\mathbf{Z}) \big| Z_{S_k+t} = \Phi_{Z_{S_k}}(t)\big]$. This makes it possible to introduce the following important definitions:

Definition 4.1. Let U^* be the function defined on $E \times \mathbb{R}^+$ by:

$$U^*(z, s) = \mathbb{E}\big[\mathbb{1}_\mathcal{D}(\mathbf{Z}) | Z_s = z\big]. \tag{4.3}$$

Definition 4.2. Let U^- be the function defined on $E \times \mathbb{R}^+$ by:

$$U^\text{-}(z^\text{-}, s) = \int_E U^*(z^+, s) K_{z^\text{-}}(z^+) \mathrm{d}\nu_{z^\text{-}}(z^+). \tag{4.4}$$

The quantity $U^*(z, s)$ measures the chances of having a system failure before t_f knowing the system is in state z at time s, and the quantity $U^-(z^-, s)$ the chances of having a system failure before t_f knowing the system is jumping from the state z^- at time s. These quantities play an important role in the latter.

4.2. A way to build an optimal importance process

In the importance process, generating the trajectories jump by jump by using (4.1) and (4.2) is not restrictive in term of efficiency, as proved by the following theorem:

Theorem 4.3. *For all $z \in E$, $z^- \in \overline{E}$, and $s \in [0, t_f)$, the jump densities with respect to μ_z such that*

$$g^*_{T_k' | Z_{S_k'}', S_k'=z, s}(u) = \frac{U^\text{-}\big(\Phi_z(u), s+u\big)}{U^*(z, s)} f_{T_k | Z_{S_k}=z}(u) \tag{4.5}$$

and the kernels $\mathcal{K}^*_{z^\text{-}, s}$ having a density with respect to ν_{z^-} which satisfies

$$K^*_{z^\text{-}, s}(z) = \frac{U^*(z, s)}{U^\text{-}(z^\text{-}, s)} K_{z^\text{-}}(z) \tag{4.6}$$

correspond to the jump densities and the transition kernels of an optimal importance process.

Note, these optimal densities do integrate to one as $U^*(z, s) = \int_0^{t_z^*} U^\text{-}\big(\Phi_z(u), s+u\big) f_{T_k | Z_{S_k}=z}(u) \mathrm{d}u$, and $U^-(z, s) = \int_E U^*(z^+, s) K_z(z^+) \mathrm{d}\nu_z(z^+)$.

Proof. Assume the trajectory $\mathbf{z} = \Theta\big((z_0, t_0), \ldots, (z_n, t_n)\big)$ has been simulated with (4.5) and (4.6). Then its density g with respect to ζ is:

$$g(\mathbf{z}) = \prod_{k=0}^{n} g^*_{T'_k | Z'_{S'_k}, S'_k = z_k, s_k}(t_k) \prod_{k=1}^{n} K^*_{z_k^-, s_k}(z_k)$$

So it verifies:

$$
\begin{aligned}
g(\mathbf{z}) &= \prod_{k=0}^{n} \frac{U^{\text{-}}\big(\Phi_{z_k}(t_k), s_k + t_k\big)}{U^*\big(z_k, s_k\big)} \prod_{k=1}^{n} \frac{U^*\big(z_k, s_k\big)}{U^{\text{-}}\big(z_k^-, s_k\big)} \prod_{k=0}^{n} f_{T_k | Z_{S_k} = z_k}(t_k) \prod_{k=1}^{n} K_{z_k^-}(z_k) \\
&= \prod_{k=0}^{n} \frac{U^{\text{-}}\big(z_{k+1}^-, s_{k+1}\big)}{U^*\big(z_k, s_k\big)} \prod_{k=0}^{n-1} \frac{U^*\big(z_{k+1}, s_{k+1}\big)}{U^{\text{-}}\big(z_{k+1}^-, s_{k+1}\big)} f(\mathbf{z}) \\
&= \frac{U^{\text{-}}\big(z_{n+1}^-, s_{n+1}\big)}{U^*\big(z_0, s_0\big)} f(\mathbf{z}) = \frac{\mathbb{1}_{\mathcal{D}}(\mathbf{z}) f(\mathbf{z})}{\mathbb{E}_{z_0}\big[\mathbb{1}_{\mathcal{D}}(\mathbf{z})\big]} = g^*(\mathbf{z}),
\end{aligned}
$$

where $g^*(\mathbf{z})$ is the density for an estimator with zero variance. $\square$

Equations (4.5) and (4.6) serve as a guide to build an importance process: one should try to specify densities as close as possible to these equations so as to get an estimator variance as close as possible to the minimal zero variance.

4.3. Observations on the optimal process

As we do not know the explicit forms of U^* and U^-, the construction of an importance process close to the optimal one is delicate. Nonetheless, the equations (4.5) and (4.6) can give us information on how to build an importance process in practice. In this Section, we investigate the properties of the optimal importance process and of the function U^* with the aim of building a good and practical importance process.

For instance, we can get the expression of the jump rate of the optimal process. For the time of the k-th jump, by definition of the jump rate and knowing that $(Z'_{S'_k}, S_k) = (z, s)$, we get:

$$\lambda^*_{z,s}(u) = \frac{g^*_{T'_k | Z'_{S'_k}, S'_k = z, s}(u)}{1 - \int_0^u g^*_{T'_k | Z'_{S'_k}, S'_k = z, s}(v) \mathrm{d}v},$$

$$\Leftrightarrow \quad \lambda^*_{z,s}(u) = \frac{U^{\text{-}}\big(\Phi_z(u), s + u\big)\big(\lambda_z(u)\big)^{\mathbb{1}_{u < t^*_z}} \exp\big[-\Lambda_z(u)\big]}{\int_{(u, t^*_z]} U^{\text{-}}\big(\Phi_z(v), s + v\big)\big(\lambda_z(v)\big)^{\mathbb{1}_{v < t^*_z}} \exp\big[-\Lambda_z(v)\big] \mathrm{d}\mu_z(v)}. \tag{4.7}$$

Using some properties of U^* and (4.7) we can prove the following theorem:

Theorem 4.4. *The jump rate of the optimal importance process defined by the densities (4.5) and (4.6) verifies:*

$$\lambda^*_{z,s}(u) = \frac{U^{\text{-}}\big(\Phi_z(u), s + u\big)}{U^*\big(\Phi_z(u), s + u\big)} \lambda_z(u). \tag{4.8}$$

The proof is provided in Appendix B.

Note that the expression (4.8) can be easily interpreted. $\lambda^*_{z,s}(u)$ corresponds to the jump rate at the state $Z_{s+u} = \Phi_z(u)$. $U^*\big(\Phi_z(u), s + u\big)$ is the probability of generating a failing trajectory if $Z_{s+u} = \Phi_z(u)$ and if there is no jump at time $s + u$. $U^-\big(\Phi_z(u), s + u\big)$ is the probability of generating a failing trajectory if there

is a jump at time $s + u$ and if the departure state is $Z_{s+u^-} = \Phi_z(u)$. So the ratio $\dfrac{U^\cdot\big(\Phi_z(u), s + u\big)}{U^*\big(\Phi_z(u), s + u\big)}$ is the factor multiplying the probability of generating a failing trajectory when there is a jump at time $s + u$. The expression indicates that, in order to reach the zero variance, one should increase the original jump rate in the same proportion as a jump would increase the probability of getting a failing trajectory.

The Theorem 4.4 is noteworthy, because in practice the law of the jump time is specified through the jump rate. Thus it can be used to specify the laws of the jump times of an importance process, as we will do in Section 4.4.

Also note that with equations (4.8) and (4.6) indicate that, once the region D has been reached, the optimal process does not differ from the original process. Indeed if τ_D is the reaching time of the critical region D, then for $s \geq \tau_D$ we have for all states z and z^-, $U^*(z, s) = U^-(z^-, s) = 1$ and so for $s \geq \tau_D$ we get $K^*_{z^-,s} = K_{z^-}$, and for $s + u \geq \tau_D$ we get $\lambda^*_{z,s}(u) = \lambda_z(u)$.

As it plays an important role in the expression of the optimal process, we look for more information about the function U^*. We first notice that: if τ is a stopping time such that $t_f > \tau > s$, then

$$
\begin{aligned}
U^*(z, s) &= \mathbb{E}\big[\mathbb{1}_{\mathcal{D}}(\mathbf{Z})\big|Z_s = z\big] \\
&= \mathbb{E}\Big[\mathbb{E}\big[\mathbb{1}_{\mathcal{D}}(\mathbf{Z})\big|Z_\tau\big]\Big|Z_s = z\Big]
\end{aligned}
$$

$$
\text{and so} \quad U^*(z, s) = \mathbb{E}\big[U^*(Z_\tau, \tau)\big|Z_s = z\big]. \tag{4.9}
$$

Using the equation (4.9) we can show the two following properties:

Theorem 4.5. U^* *is kernel invariant on boundaries:*

$$
\forall z \in E, \qquad U^\cdot\big(\Phi_z(t_z^*), s + t_z^*\big) = \lim_{t \nearrow t_z^*} U^*\big(\Phi_z(t), s + t\big). \tag{4.10}
$$

Theorem 4.6. *If* $u \to U^\cdot\big(\Phi_z(u), s + u\big)$ *and* $u \to \lambda_z(u)$ *are continuous almost everywhere on* $[0, t_z^*)$, *then almost everywhere* U^* *is differentiable along the flow, with:*

$$
\frac{\partial U^*\big(\Phi_z(v), s + v\big)}{\partial v} = U^*\big(\Phi_z(v), s + v\big)\lambda_z(v) - U^\cdot\big(\Phi_z(v), s + v\big)\lambda_z(v) \tag{4.11}
$$

Theorems 4.5 and 4.6 can in fact be seen as forward Kolmogorov equations on U^*. A complete proof for these two properties is in the Appendix B.

4.4. A parametric importance process

In order to find an importance process that gives a good variance reduction, we usually restrict the search within a parametric family of importance densities. Then we rely on optimization routines to find the parameters yielding the best variance reduction. Here, we propose to use a parametric approximation of $U^*(z, s)$, and then combine it with equations (4.8) and (4.6) to get the form of the importance kernels and of the importance intensities. If we denote $U_\alpha(z, s)$ our approximation of $U^*(z, s)$, where the parameter α belongs to the set A_{param}, and we set $U_\alpha^\cdot\big(z^-, s\big) = \int_E U_\alpha(w, s)K_{z^-}(w)\mathrm{d}\nu_{z^-}(w)$, then the corresponding importance intensities and kernels are given by:

$$
\lambda'_{z,s}(u) = \frac{U_\alpha^\cdot\big(\Phi_z(u), s + u\big)}{U_\alpha\big(\Phi_z(u), s + u\big)}\lambda_z(u), \tag{4.12}
$$

$$
K'_{z^-,s}(z^+) = \frac{U_\alpha\big(z^+, s\big)}{U_\alpha^\cdot\big(z^-, s\big)}K_{z^-}(z^+). \tag{4.13}
$$

With these settings and notations, condition (C3) can be expressed as:

$$U^*(z_k, s_k) > 0, \text{ and } K_{z_k^-}(z_k) > 0 \Rightarrow U_\alpha(z_k, s_k) > 0$$
$$U^*(z_k, s_k + t) > 0, \text{ and } \lambda_{z_k}(t) > 0 \Rightarrow U_\alpha(z_k, s_k + t) > 0,$$

for any $z_k \in E$, and $z_k^- \in \overline{E}$, and $t \in (0, t*_{z_k}]$. It is therefore satisfied if we take U_α positive everywhere for instance.

Here we switch the problem of setting a density g close to g^* by finding λ' and K', to the problem of finding a surface U_α on $E \times \mathbb{R}^+$ close to the surface U^*.

Note that this way of building a parametric family of importance processes can be applied to any kind of systems, though the shape of U_α may have to be adapted from case to case. Indeed, we expect the shape of U^* to depend on the configuration of the system and so does the shape of the U_α's.

We could also have plugged the approximations U_α and U_α^- into (4.5), rather than into (4.8), but the option we have chosen is in fact more convenient and computationally more efficient. With equation (4.8), we pass through the intensity, so the density of the T_k''s automatically integrates to 1. Conversely if we pass through equation (4.5), we have to renormalize the density so it integrates to 1 before simulating a realization of the T_k'. As this renormalization requires to compute an integral, it is less advantageous.

4.5. Remarks on the parameter optimization

As mentioned in the introduction, we propose to use the cross-entropy method presented in [13] to select the parameters of the importance density as it was done in [31]. In the case of PDMP, it is hard to use the adaptive cross-entropy method presented in [13]: the adaptive cross-entropy requires to have a function $\mathcal{O} : \mathbf{E} \to \mathbb{R}$ that orders the trajectories which is hard to specify judiciously. This function must order the states in such a way that there exists a threshold c for which

$$\mathbb{P}(\mathbf{Z} \in \mathcal{D}) = \mathbb{P}(\mathcal{O}(\mathbf{Z}) > c), \tag{4.14}$$

and there exists a sequence of thresholds $c_0 \leq c_1 \leq \cdots \leq c_k = c$ so that

$$\mathbb{P}(\mathbf{Z} \in \mathcal{D}_i) = \mathbb{P}(\mathcal{O}(\mathbf{Z}) > c_i), \tag{4.15}$$

where $\mathcal{D}_0 \subseteq \mathcal{D}_1 \subseteq \cdots \subseteq \mathcal{D}_i \subseteq \mathcal{D}_{i+1} \subseteq \cdots \subseteq \mathcal{D}_k = \mathcal{D}$. In order to run the Cross-Entropy algorithm with relatively low sample sizes at each step (from 100 to 1000), it is good to set the function $\mathcal{O}$ so that $20 \leq \frac{\mathbb{P}(\mathcal{O}(\mathbf{Z}) > c_i)}{\mathbb{P}(\mathcal{O}(\mathbf{Z}) > c_{i+1})} \leq 100$ [13]. The issue is that we find it hard to specify such a function with PDMP. For this reason we used a simplified version of the CE method considering only one threshold c.

The CE algorithm also requires to minimize an approximation of the Kullback-Leiber divergence $D(g_\alpha, g^*)$. We simulate a sample used to compute many approximations of $D(g_\alpha, g^*)$. The optimization routine uses the sample to compute some approximations of $D(g_\alpha, g^*)$ with different values of α. When the sample contains too many trajectories in $\mathcal{D}$, this approximations can be computationally heavy. Conversely, when the sample does not contain enough trajectories in $\mathcal{D}$ the approximations are not accurate enough. So we choose to increase the size of the sample gradually until it contains n_{CE} trajectories in $\mathcal{D}$, n_{CE} being a number fixed by the user. This way the objective function to minimize and its gradient are both a sum over n_{CE} terms, and thus they are not too heavy to compute. The CE algorithm we used is presented in Table 1.

However, to our knowledge, there is no guarantee that the minimization routine used in the cross entropy method converges to a global optimum. Therefore, to avoid falling in a local optimum, one should run several times the cross entropy method with different initial values for the vector of parameters. Note that the parametrization must be chosen carefully: indeed the family of the importance densities must contain densities that are close to the zero-variance density $g^*(\mathbf{z}) = \frac{\mathbb{1}_\mathcal{D}(\mathbf{z}) f(\mathbf{z})}{p}$ to obtain a good variance reduction, otherwise

TABLE 1. CE algorithm.

Initialization: choose $\alpha_0 \in A_{param}$ and $n_{CE} \in \mathbb{N}^*$ and set $t = 0$, and $\varepsilon > 0$
while $\|\alpha_t - \alpha_{t+1}\| < \varepsilon$ **do**
    Set $k = 1$, and generate $\mathbf{Z}_1' \sim g_{\alpha_t}$
    while $\sum_{i=1}^{k} \mathbb{1}_{\mathbf{Z}_i' \in \mathcal{D}} < n_{CE}$ **do**
        Generate $\mathbf{Z}_{k+1}' \sim g_{\alpha_t}$
        $k := k + 1$
    $N = k - 1$
    Compute $\alpha_{t+1} = \underset{\alpha \in A_{param}}{\operatorname{argmin}} \; \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{\mathbf{Z}_i' \in \mathcal{D}} \frac{f(\mathbf{Z}_i')}{g_{\alpha_t}(\mathbf{Z}_i')} log(g_{\alpha_t}(\mathbf{Z}_i'))$
    $t := t + 1$
End: Estimate p using the importance density $g_{\alpha_{t-1}}$

we could even obtain a variance increase. In order to avoid a variance increase, the parametric family should contain the original density f. Indeed, if we specify in the parametric family that for say $\alpha = 0$ we have $g_0 = f$ then the parameter optimization should not select a parameter worse than $\alpha = 0$ and in the worst scenario the variance remains unchanged. This is why we advise that the family of the U_α functions includes a constant function, so that the original process with jump rate λ_z and transition kernel K_z is included in the admissible importance processes.

The initial vector of parameters α_0 has a big influence on the convergence of the method. Ideally, it should be chosen to simulate n_{CE} trajectories in $\mathcal{D}$ relatively fast, but, in order to avoid an over-biasing situation with a wrong approximation of the Kullback-Leiber divergence at the first step, we recommend to choose α_0 so that g_{α_0} is as close as possible from f. Testing several values of α_0 is therefore necessary, to get a sense of what is a good α_0.

5. SIMULATION STUDY ON A TEST CASE

In this section we present how we build an importance process for the heated room system presented in Section 2.7.

5.1. A parametric family of importance processes

In the heated-room system, the three heaters are identical and are in parallel redundancy, so we expect the probability $U^*(z, s) = \mathbb{E}\big[\mathbb{1}_{\mathcal{D}}(\mathbf{z})|Z_s = z\big]$ to increase with the number of failed heaters in the state z. Therefore, noting $b(z)$ the number of failed heaters in state z, we start by setting

$$U_\alpha(z, s) = H_\alpha\big(b(z)\big) Q(x, s) \tag{5.1}$$

where Q is a function of position and time, and H_α is a function on integers. We set $H_\alpha(0) = 1$. As we want $U_\alpha(z, s)$ to increase with $b(z)$, H_α has to be an increasing function.

If T denotes the time until the next jump after a time s, using (4.9) with $\tau = s + T$ we get:

$$U^*(z, s) = \mathbb{E}\big[U^*(Z_{s+T}, s + T)\big|Z_s = z\big]. \tag{5.2}$$

As the repair rates are larger than the failure rates by one order of magnitude in practice, when there is at least one failed heater, the probability of arriving in a more degraded state Z_T is much lower than the probability of having a repair. This last remark can actually be applied to any reliable industrial system (see for instance

[10]). Ideally we would like U_α to mimic the property of U^* so we would like to have

$$U_\alpha(z,s) = \mathbb{E}\big[U_\alpha(Z_T, s+T)\big|Z_s = z\big] \tag{5.3}$$

which can be reformulated as:

$$H_\alpha\big(b(z)\big) = \sum_{m^+ \in \mathbb{M}} H_\alpha\big(b(x,m^+)\big) \int_{(0,t_z^*]} K_{\Phi_z(u)}\big((\phi_x^m(u), m^+)\big) w_z(u) \mathrm{d}\mu_z(u) \tag{5.4}$$

where $w_z(u) = \dfrac{Q(\phi_x^m(u), s+u)}{Q(x,s)} \exp\big[\text{-}\Lambda_z(u)\big]$. As a repair is much more likely than failure, if the transition from state $(\phi_x^m(u), m)$ to the state $(\phi_x^m(u), m^+)$ indexes a repair $K_{\Phi_z(u)}\big((\phi_x^m(u), m^+)\big)$ is larger than if it had indexed a failure. So, (5.4) implies that, when $b(z) > 1$, the value of $H_\alpha(b(z))$ is closer from $H_\alpha(b(z)-1)$ than from $H_\alpha(b(z)+1)$. As H_α was supposed increasing, it must be convex. So we propose that $H_\alpha(b(z)) = \exp\big[\alpha_1 b(z)^2\big]$, with $\alpha_1 > 0$. If, from a $Z_s = \Phi_z(u)$, the transition j corresponds to a failure then we have:

$$\lambda'^j_{z,s}(u) = \lambda_z^j(u) \exp\big[\alpha_1\big(2b(z)+1\big)\big], \tag{5.5}$$

and if it corresponds to a repair then we have:

$$\lambda'^j_{z,s}(u) = \lambda_z^j(u) \exp\big[-\alpha_1\big(2b(z)-1\big)\big]. \tag{5.6}$$

The jump rate satisfies:

$$\lambda'_{z,s}(u) = \sum_{i \in J} \lambda'^i_{z,s}(u) \qquad \text{and } \forall \qquad K'_{\Phi_z(u)^-,s}(z^+) = \frac{\lambda'^{z^+}_{z,s}(u)}{\int_E \lambda'^{z^+}_{z,s}(u) \mathrm{d}\nu_z(z^+)}. \tag{5.7}$$

We set the jump kernel such that its density satisfies for $u \in [0, t_z^*]$:

$$K'_{z^-}(z^+) = \frac{K_{z^-}(z^+) \exp\big[-\alpha_1\, b(z^+)^2\big]}{\int_E K_{z^-}(z) \exp\big[-\alpha_1\, b(z)^2\big] \mathrm{d}\nu_{z^-}(z)}. \tag{5.8}$$

Note that plugging U_α into the equations (4.5) and (4.6) imposes some kind of symmetry in the biasing of failure and repair rates. It is especially visible in equations (5.5) and (5.6): On the one hand the failure rate associated to the transition from a state z^- to z^+ is multiplied by a factor $\exp\big[\alpha_1\big(2b(z^-)+1\big)\big]$, and on the other hand the repair rate corresponding to the reversed transition (from state z^+ to state z^-) is divided by a factor $\exp\big[\alpha_1\big(2b(z^-)-1\big)\big]$. The equations (4.5) and (4.6) not only imply that the failures should be enhanced and the repairs inhibited, but it also states that the magnitudes of the distortion should be reciprocal.

The square in H_α's formula was introduced to strengthen the failure rates when the number of broken heaters gets larger. The idea was to shorten the duration where several heaters are simultaneously failed in the simulated trajectories. Indeed, as repair is faster than failure, the shorter are the durations with a failed heater the more likely is the trajectory. Increasing the failure rates with the number of broken heaters is a mean to simulate more trajectories in $\mathcal{D}$ while maintaining the natural proportion between the likelihoods of the trajectories, which should decrease the variance.

As the failure on demand was likely to play an important role in the system failure, we choose to separate it from spontaneous failure in our parametrisation setting $U_\alpha((x_{\min}, m), s) = \exp[-\alpha_2 b(z)^2]H_\alpha(x_{\min}, s)$. This allows to better fit U_α to U^*. Under this assumption, the equation (4.13) implies that for $z^- = (x_{\min}, m)$, the

importance kernel takes this form:

$$K'_{z^-}(z^+) = \frac{K_{z^-}(z^+) \exp\left[-\alpha_2\, b(z^+)^2\right]}{\int_E K_{z^-}(z) \exp\left[-\alpha_2\, b(z)^2\right] \mathrm{d}\nu_{z^-}(z)}. \tag{5.9}$$

5.2. Results

The Monte-Carlo simulations have been carried out using the Python library PyCATSHOO. (The flow functions ϕ_x^m were computed using a Runge-Kutta method of order 4 with a discretization step of 0.01. This discretization step is small enough so that reducing it further does not change the estimations.) As the Cross-Entropy method was not yet implemented in PyCATSHOO, we have used a specific Python code for the Cross-Entropy and the importance sampling methods. The system parameters used in the simulation were the following ones: $x_{\min} = 0.5$, $x_{\max} = 5.5$, $x_e = -1.5$, $\beta_1 = 0.1$, $\beta_2 = 5$, $t_f = 100$. Trajectories were all initiated in the state $z_0 = \left(7.5,\ (OFF, OFF, OFF)\right)$. The probability of having a system failure before t_f was estimated to $p = 1.29 \times 10^{-5}$ with an intensive Monte-Carlo estimation based on 10^8 runs.

The values of the parameters selected by the cross-entropy method were $\alpha_1 \simeq 0.915$ and $\alpha_2 \simeq 1.197$, and for the first step, the approximation of the Kullback-Leiber divergence between g^* and g_α was obtained by simulating from a biased density with parameters $(0.5, 0.5)$. The whole cross-entropy method lasted approximately 9 minutes. Most of the running time was allocated to the optimization within each step of the cross-entropy, because each evaluation of the objective function and of its gradient was costly. In order to optimize the running time of the cross-entropy method, the size of the sample used for the approximations of the Kullback-Leiber divergence were set by simulating until we would get $n_{CE} = 100$ trajectories with a system failure. The number of $n_{CE} = 100$ roughly guaranties that the two first digit of of the Kullback-Leiber divergences are identified by their approximations. For each of the three steps needed to select the parameters, samples of respectively 1970, 126, 127 trajectories were used.

A comparison between Monte-Carlo and the associated importance sampling estimates is presented in Table 2, where we display the number N_{sim} of simulations used for each method, the estimates $\hat{p}$ of the probability, the associated empirical variances $\hat{\sigma}^2/N_{\mathrm{sim}}$ and confidence intervals $\widehat{\mathrm{IC}}$, and the mean time of a simulation t_{sim} in seconds. For 10^6 simulations the results show that the Monte-Carlo estimator has not converged yet, whereas the importance sampling estimate is very accurate. To compare the two methods we estimate the efficiency of their estimators when they have converged. The efficiency is defined by the ratio of the precision and the computational time:

$$eff = \frac{1}{\sigma^2/N_{\mathrm{sim}}} \times \frac{1}{N_{\mathrm{sim}} t_{\mathrm{sim}}} = \frac{1}{\sigma^2 t_{\mathrm{sim}}}.$$

TABLE 2. Comparison between Monte-Carlo and importance sampling estimations.

|	N_{sim}	$\hat{p}$	$\hat{\sigma}^2/N_{\mathrm{sim}}$	$\widehat{\mathrm{IC}} \times 10^5$	t_{sim}	$\widehat{eff}$
	10^3	1.28×10^{-5}	4.37×10^{-13}	$[1.15, 1.41]$	0.073 s	3.1×10^{10}
IS	10^4	1.273×10^{-5}	5.07×10^{-14}	$[1.228, 1.317]$	0.073 s	2.7×10^{10}
	10^5	1.289×10^{-5}	5.01×10^{-15}	$[1.275, 1.303]$	0.077 s	2.6×10^{10}
	10^6	1.288×10^{-5}	5.05×10^{-16}	$[1.283, 1.292]$	0.079 s	2.5×10^{10}
MC	10^6	0.4×10^{-5}	4.00×10^{-12}	$[0.01, 0.79]$	0.022 s	no convergence
	10^7	1.3×10^{-5}	1.28×10^{-12}	$[1.07, 1.51]$	0.022 s	3.5×10^6
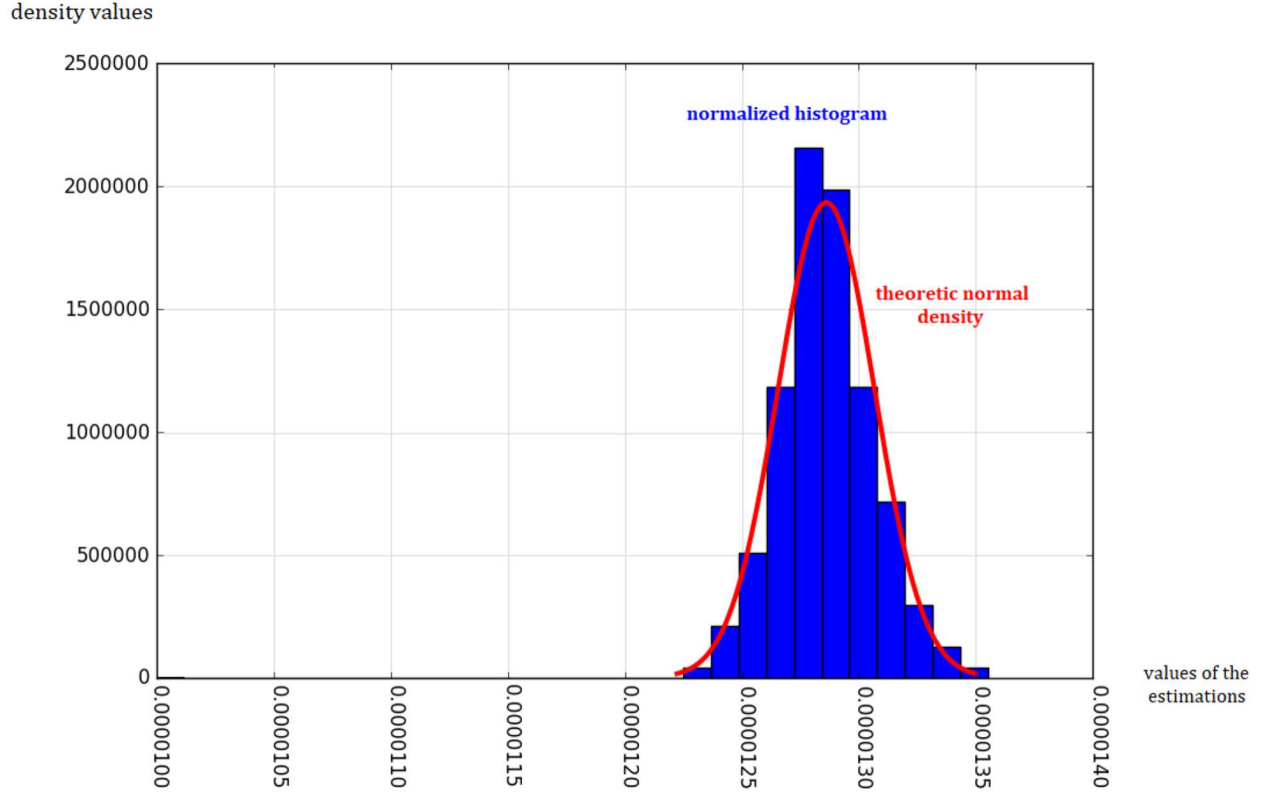
FIGURE 6. Asymptotic normality of the IS estimator (for $N = 10^4$).

The efficiency can be interpreted as the contribution of a second of computation to the precision of the estimator. We estimate it by $\widehat{eff} = \frac{1}{\hat{\sigma}^2 t_{\text{sim}}}$. The results indicate that our importance sampling strategy is approximately 7 000 times more efficient than a Monte-Carlo method.

We also verify that the importance sampling estimations are asymptotically normally distributed. The asymptotic normality was not observed for $N = 10^3$, but it was observed for larger sample sizes. For instance for $N = 10^4$, Figure 6 shows a normalized histogram on 100 estimations $\hat{p}_{\text{IS}}$ that matches the normal density with mean p and with the standard deviation of the 100 estimations. We also recorded the weights of the failing trajectories in the sample of one run of the IS method with $N = 10^4$. Figure 7 shows that the weights are close to the value p, suggesting that the importance density is close to the optimal density. Figure 8 is a zoom-in on the largest weight: It shows there is no degenerated preponderant weight such that $\frac{f(\mathbf{Z}_i')}{g_\alpha(\mathbf{Z}_i')} \gg p$, suggesting there is no sign of under-favored region of $\mathcal{D}$ in g_α. Here we do not need to check the weight degeneracy in all parts of $\mathcal{D}$ because, as we now the value of p the can simply check the estimation are unbiased and normally distributed to ensure convergence is reached. Finally, in Figures 9 and 10, we present the graphs of two trajectories obtained respectively with the original process with density f and with the importance process selected by the CE method with density $g_{(\alpha_1, \alpha_2)}$.

6. DISCUSSION

Our work shows that importance sampling is applicable to any PDMP with or without boundaries. We have given the expressions of the intensities and the kernels of the optimal importance process, and we have seen that it depends on a critical function U^*. These expressions show that the optimal importance process has a
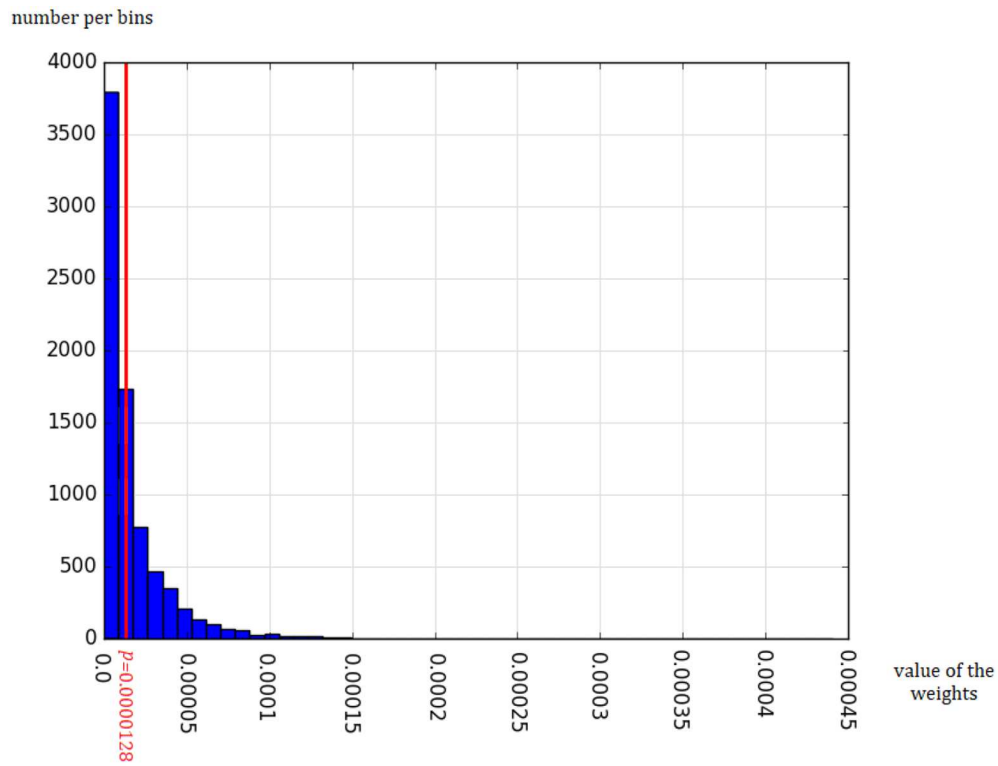
FIGURE 7. Allocation of the weights of failing trajectories (for $N = 10^4$).
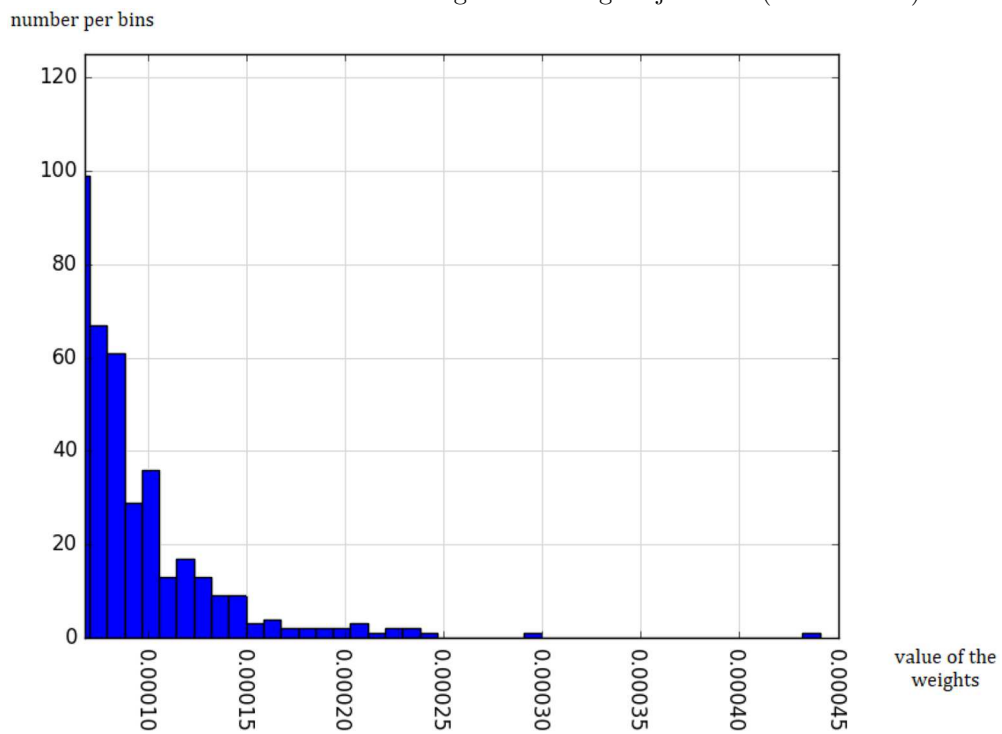
FIGURE 8. Allocation of the largest weights in the sample (for $N = 10^4$).
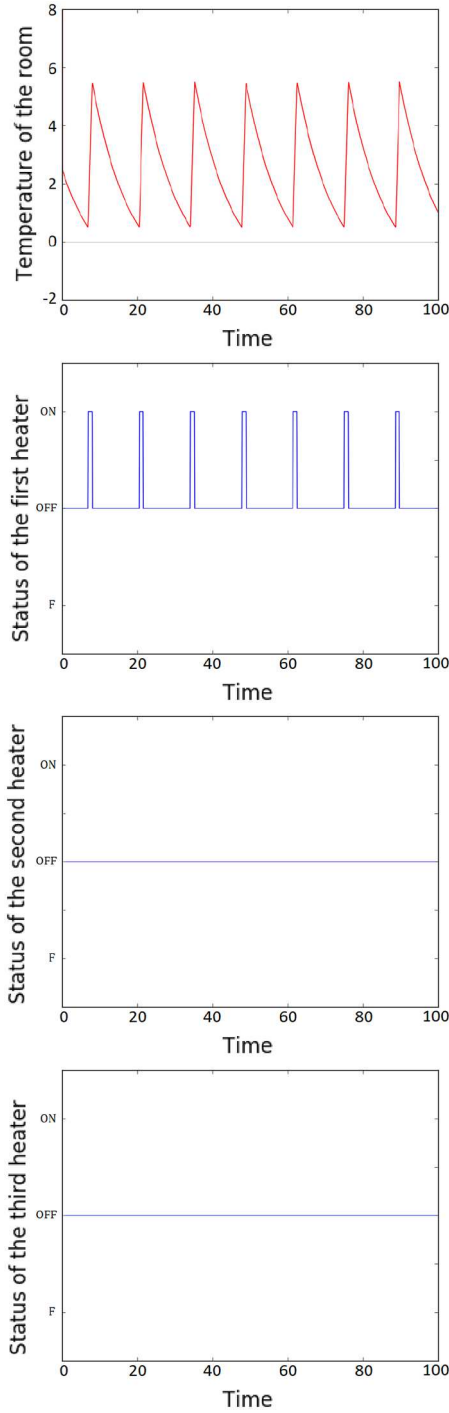
FIGURE 9. A trajectory of the coordinates of the state of the system. This trajectory was generated with the original process with density f.
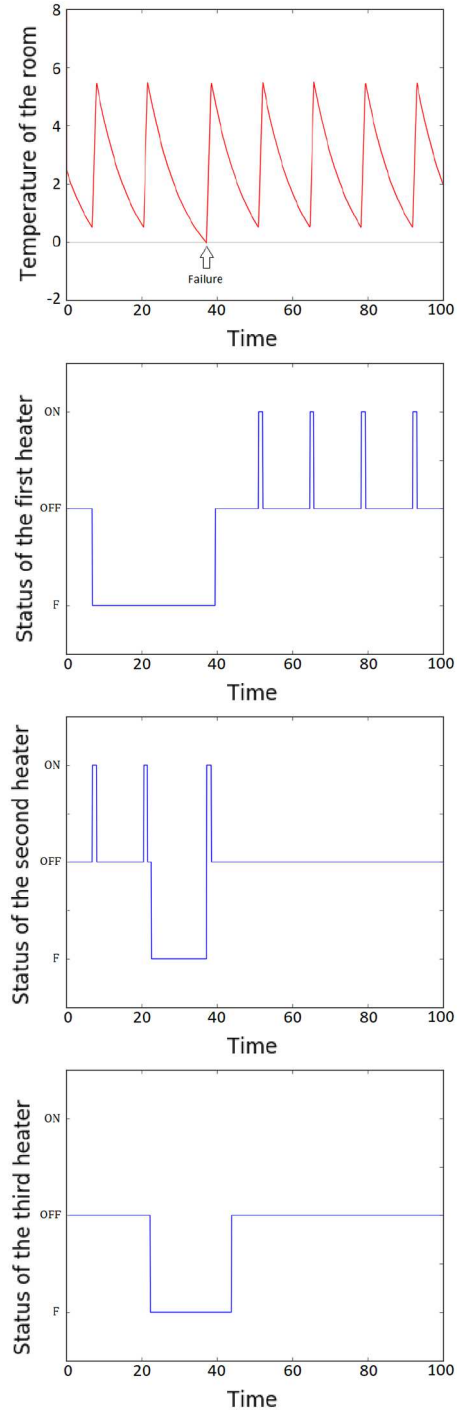
FIGURE 10. A trajectory of the coordinates of the state of the system. This trajectory was generated with the importance process with density $g_{(\alpha_1,\alpha_2)}$.

specific structure. Although we do not have a closed form expression of the function U^*, these expressions are important for two reasons: (1) they prove the existence of an optimal bias, which ensures that the importance sampling technique can be very efficient on PDMPs. (2) They can guide the practical design of an efficient and explicit importance process. Indeed, by replacing U^* by an approximation in the optimal expressions of the transition rates and kernels, we preserve the structure of the optimal importance process. The presented method therefore helps designing an importance process having the same behavior as the optimal one, and it showed good efficiency on our case study.

This biasing strategy can be applied to any system, but the parametric shape of the approximation of U^* may have to be adapted from case to case. The parametric shape presented in this article is suited to any system with similar components in terms of failure rates and repair rates and containing one minimal cut set (A minimal cut set being a group of components that need to fail so that the system can fail). For a system with a different configuration, we expect the shape of the function U^* will differ, and the method may require a different parametric approximation for the function U^*.

Our approach through the function U^* can be applied to any sub-classes of PDMP, like, for instance, Markov chains [18], or continuous time Markov Chain, or queing models. In the particular case of PDMP that is a continuous time Markov Chain, the definition of the function U^* is close to the forward committor function used in the transition path theory [23]. In the case of a general PDMP, a committor function would be a function $(z, s) \to \mathbb{E}\left[\mathbb{1}_{\mathcal{D}_A}(\mathbf{Z})|Z_s = z\right]$ where $\mathcal{D}_A$ is the set of trajectories that pass through D without passing through a set $A \subset E$ first. U^* is therefore a commitor function for which $A = \emptyset$. It is also interesting to note that, in the Adaptive Multilevel Splitting algorithm, the asymptotic variance is minimized when using the committor function as the score function [1, 6], similarly, in the interacting particles system method [14], the function U^* also plays a role in the optimal potential function method [9]. Approaching the function U^*, allows to efficiently estimate rare event for importance sampling, but also for Adaptive Multilevel Splitting algorithm and the interacting particles system method. A method that allows to approximate this function would lead to significant improvement in the reliability assessment field.

We proposed to find a good approximation of U^* searching inside a family of parametric functions $(U_\alpha)_{\alpha \in A_{param}}$. In our application on the heated room system, we used the Cross-Entropy method to select an efficient parameter α. We have noticed that the Cross-Entropy method tends to diverge quickly if it is not well initialized: The choices of α_0 and n_{CE} are critical for the convergence of the method. These two parameters must be well tuned because they impact the quality of the first approximations of the Kullback-Leiber divergence within the CE algorithm, and these approximations must be accurate enough to launch the optimization routine on a good track. To choose a high value for n_{CE} is a way to insure that these first approximations are accurate enough, but it is not worth considering in practice, as it greatly slows down the CE algorithm. The only solution is to find right away an α_0 which yields correct approximations of the Kullback-Leiber divergence. This is unfortunately difficult to do, and may be even harder with more complex systems. We believe that the CE method used in this article must be improved or substituted by an other parameter optimization method so that the initialization gets less critical.

With the IS method, depending on the importance process chosen, we can observe some weight degeneracy and therefore slow convergence. Weight degeneracy typically happens when two conditions are met: (1) there is a domain $\mathcal{D}_1 \subset \mathcal{D}$ such that the likelihood ratios within this domain are very big compared to the likelihood ratios in other domains of $\mathcal{D}$, meaning the domain $\mathcal{D}_1$ is under-favored by the importance density g compared to f and (2) though it is unlikely, few realizations of the importance process $\mathbf{Z}'_i$ are drawn in $\mathcal{D}_1$, which creates unbalanced weights. Some methods allow to reduce the risk of weight degeneracy by using resampling schemes like for instance in the interacting particle system (IPS) method [14], but, even though it is reduced, the risk of weight degeneracy still remains within the IPS method. The IPS method takes in input some potential functions G_k (also called score functions). If these functions does not favor the domain $\mathcal{D}_1$, the convergence is slowed down[9], and we can end up with the same situation. In this method the weights of the simulation outputs are the inverse of the product of resampling weights multiplied by the objective function's evaluations (see equations 2.18 in [14]). The degenerate weights would therefore appear each time some trajectories in $\mathcal{D}_1$ are selected by the resamplings, eventhough the resamplings make it unlikely. Weight degeneracy does not depend

on the method used but it rather depend on the choice of the importance process or on the choice the potential functions. Weight degeneracy is that it is the symptom of a slow convergence, therefore the sample size should be increased until the weight degeneracy fades out: the weight degeneracy is a tool to select the sample size for both methods. For a fixed sample size, if the sample contains some trajectories in $\mathcal{D}_1$, the weight degeneracy can be a criterion to reject the importance density, or the potential function, used. But this last criterion is valid only if the sample contains observations in the under-favored domains in $\mathcal{D}$, which is unlikely by definition. One important point to stress out, is that witnessing no weight degeneracy within the simulations outputs does not guarantee the convergence, we can consider we have converge if the sample size is reasonably large and that we do not witness weight degeneracy in all part of $\mathcal{D}$.

When choosing the importance process, there is a risk of over-biasing. Over-biasing corresponds to the situation where a domain $\mathcal{D}_2 \subset \mathcal{D}$ is over-favored by the importance process, resulting in an under-favoring of an other domains $\mathcal{D}_1 \subset \mathcal{D}$. In this situation a weight degeneracy exist in $\mathcal{D}_1$ but it is not witnessed because no trajectory within the sample is drawn in $\mathcal{D}_1$. This situation happens when one type of failing trajectories is over represented in the importance distribution comparatively to other types of failing trajectories. This phenomenon can result in underestimating the probability of the system failure and in underestimating the variance. To avoid it, we must satisfy two points: (1) We must design a parametric importance density that can increase the likelihoods of each type of failing trajectories separately. (2) We need to initiate the Cross-Entropy method with a sample of trajectories that contains all types of failing trajectories. It is therefore preferable to apply this method only on systems of reasonable complexity, for which it is possible to determine the different types of failing trajectories. The parametric functions $(U_\alpha)_{\alpha \in A_{param}}$ should be flexible enough to satisfy the two previous points, but one should pay attention to keep the dimension of the vector of parameter α reasonably small, so that we avoid a prohibitive computational effort during the optimization routines in the CE.

7. CONCLUSION

We have presented a model for multi-component systems based on PDMPs. In order to speed up reliability assessment on such systems, we have adapted the importance sampling method to trajectories of PDMP. We have given a dominant measure for PDMP trajectories, allowing to properly define the likelihood ratio needed to apply the importance sampling method on such processes. The possible kinds of importance processes were discussed, and the optimal biasing strategy when simulating jump by jump was exhibited. We developed and tested a biasing strategy for a three-component heated-room system. Our importance sampling method has shown good performance, increasing the efficiency of the estimator by a factor 7 000.

APPENDIX A. THE MEASURE ζ IS σ-FINITE WHEN $\mathbf{t_f} < \infty$ AND THE MEASURES $\nu_{\mathbf{z}^-}$ ARE BOUNDED

Remember that we defined the σ-algebra $\mathcal{S}$ on the set of the possible values of $\left(Z_{S_k}, T_k\right)_{k \leq N}$ as the σ-algebra generated by the sets in $\bigcup_{n \in \mathbb{N}^*} \mathcal{B}\left(\left\{\left(z_{s_k}, t_k\right)_{k \leq n} \in (E \times \mathbb{R}_+^*)^n, \sum_{i=0}^{n} t_i = t_f\right\}\right)$. The measure ζ is defined by:

$$B \in \mathcal{S}, \quad \zeta\left(\Theta^{-1}(B)\right) = \int_{(z_k, t_k)_{k \leq n} \in B} \mathrm{d}\delta_{t_n^*}(t_n) \, \mathrm{d}\nu_{z_n^-}(z_n) \, \mathrm{d}\mu_{t_{z_{n-1}}^*}(t_{n-1}) \, \ldots \, \mathrm{d}\nu_{z_1^-}(z_1) \, \mathrm{d}\mu_{t_{z_o}^*}(t_0) \qquad (A.1)$$

Proof. Let $A_n = \left\{\left(z_{s_k}, t_k\right)_{k \leq n} \in (E \times \mathbb{R}_+^*)^n, \sum_{i=0}^{n} t_i = t_f\right\}$. Then $\Theta^{-1}(A_n)$ is the set of possible trajectories with n jumps, and the sets A_n for $n \in \mathbb{N}^*$ form a partition of the set of all possible trajectories. Note that

$A_n \subseteq (E \times [0, t_f))^n$, so

$$\zeta\big(\Theta^{-1}(A_n)\big) \le \zeta\big(\Theta^{-1}\big((E \times [0, t_f))^n\big)$$

$$\le \int\limits_{(E \times [0, t_f))^n} \mathrm{d}\delta_{t_n^*}(t_n)\, \mathrm{d}\nu_{z_n^-}(z_n)\, \mathrm{d}\mu_{t_{z_{n-1}}^*}(t_{n-1})\, \ldots\, \mathrm{d}\nu_{z_1^-}(z_1)\, \mathrm{d}\mu_{t_{z_o}^*}(t_0)$$

We suppose that the ν_{z^-} are bounded, $\exists M > 0, \forall z^- \in \overline{E},\ \nu_{z^-}(E) < M$. Under this assumption, we have:

$$\zeta\big(\Theta^{-1}(A_n)\big) \le M \int\limits_{(E \times [0, t_f))^{n-1}} \mathrm{d}\mu_{t_{z_{n-1}}^*}(t_{n-1})\, \ldots\, \mathrm{d}\nu_{z_1^-}(z_1)\, \mathrm{d}\mu_{t_{z_o}^*}(t_0)$$

$$\le M \int\limits_{(E \times [0, t_f))^{n-2}} \int_E \int_{[0, t_f)} \mathrm{d}\mu_{t_{z_{n-1}}^*}(t_{n-1})\, \mathrm{d}\nu_{z_{n-1}^-}(z_{n-1})\, \ldots\, \mathrm{d}\nu_{z_1^-}(z_1)\, \mathrm{d}\mu_{t_{z_o}^*}(t_0)$$

$$\le M(t_f + 1) \int\limits_{(E \times [0, t_f))^{n-2}} \int_E \mathrm{d}\nu_{z_{n-1}^-}(z_{n-1}) \mathrm{d}\mu_{t_{z_{n-2}}^*}(t_{n-2})\, \ldots\, \mathrm{d}\nu_{z_1^-}(z_1)\, \mathrm{d}\mu_{t_{z_o}^*}(t_0)$$

$$\le M^2(t_f + 1) \int\limits_{(E \times [0, t_f))^{n-2}} \mathrm{d}\mu_{t_{z_{n-2}}^*}(t_{n-2}) \mathrm{d}\nu_{z_{n-2}^-}(z_{n-2})\, \ldots\, \mathrm{d}\nu_{z_1^-}(z_1)\, \mathrm{d}\mu_{t_{z_o}^*}(t_0)$$

By recurrence we get that $\zeta\big(\Theta^{-1}(A_n)\big) \le M^n(t_f + 1)^n$, which proves that ζ is σ-finite. $\qquad\square$

APPENDIX B. OPTIMAL INTENSITY'S EXPRESSION, AND SOME PROPERTIES OF U^*

B.1 Equality (4.10)

Let $z^- \in \delta E$ and $s \in [0, t_f)$. Remember that equality (4.10) states that

$$U^-\big(\Phi_z(t_z^*), s + t_z^*\big) = \lim_{t \nearrow t_z^*} U^*\big(\Phi_z(t), s + t\big).$$

Proof. We denote by T the time until the next jump after the trajectory has reached the state $Z_{s+t} = \phi_z(t)$. Then we have:

$$U^*\big(\Phi_z(t), s + t\big) = \mathbb{E}\big[\mathbb{1}_{\mathcal{D}}(\mathbf{z})\big|Z_{s+t} = \phi_z(t)\big]$$

$$= \mathbb{E}\Big[\mathbb{E}\big[\mathbb{1}_{\mathcal{D}}(\mathbf{z})\big|Z_{T+s+t}\big]\Big|Z_{s+t} = \phi_z(t)\Big]$$

$$= \mathbb{E}\Big[\big(\mathbb{1}_{T < t_{\Phi_z(t)}^*} + \mathbb{1}_{T = t_{\Phi_z(t)}^*}\big)U^*(Z_{T+s+t}, s + t + T)\Big|Z_{s+t} = \phi_z(t)\Big]$$

$$= \int_0^{t_{\Phi_z(t)}^*} U^-\big(\Phi_{\Phi_z(t)}(u), s + t + u\big)\lambda_{\Phi_z(t)}(u) \exp\big[-\Lambda_{\Phi_z(t)}(u)\big]\mathrm{d}u$$

$$\quad + \exp\big[-\Lambda_{\Phi_z(t)}(t_{\Phi_z(t)}^*)\big] \int_E K_{z^-}(z^+)U^*(z^+, s + t + t_{\Phi_z(t)}^*)\mathrm{d}\nu_{z^-}(z^+)$$

$$\text{where } z^- = \Phi_{\Phi_z(t)}(t_{\Phi_z(t)}^*)$$

$$U^*\big(\Phi_z(t), s + t\big) = \int_t^{t_z^*} U^-\big(\Phi_z(u), s + u\big)\lambda_z(u) \exp\big[-\Lambda_{\Phi_z(t)}(u - t)\big]\mathrm{d}u$$

$$+ \exp\big[-\Lambda_{\Phi_z(t)}(t_z^* - t)\big] \int_E K_{z^-}(z^+)U^*(z^+, s + t_z^*)\mathrm{d}\nu_{z^-}(z^+)$$

$$\text{where } z^- = \Phi_z(t_z^*)$$

so $U^*\big(\Phi_z(t), s + t\big) = o(1) + (1 + o(1))U^-\big(\Phi_z(t_z^*), s + t_z^*\big)$ as $t \to t_z^*, \ t < t_z^*.$ $\qquad\square$

B.2 Proof of Theorem 4.4

Proof. We have seen in the proof above that

$$U^*\big(\Phi_z(t), s + t\big) = \int_t^{t_z^*} U^-(\Phi_z(u), s + u)\lambda_z(u)\exp\big[-\Lambda_{\Phi_z(t)}(u - t)\big]\mathrm{d}u$$

$$+ \exp\big[-\Lambda_{\Phi_z(t)}(t_z^* - t)\big]\int_E K_{z^-}(z^+)U^*(z^+, s + t_z^*)\mathrm{d}\nu_{z^-}(z^+)$$

so

$$U^*\big(\Phi_z(t), s + t\big) = \int_t^{t_z^*} U^-(\Phi_z(u), s + u)\lambda_z(u)\exp\big[-\Lambda_z(u)\big]\exp\big[+\Lambda_z(t)\big]\mathrm{d}u$$

$$+ \exp\big[-\Lambda_z(t_z^*)\big]\exp\big[+\Lambda_z(t)\big]\int_E K_{z^-}(z^+)U^*(z^+, s + t_z^*)\mathrm{d}\nu_{z^-}(z^+)$$

$$= \frac{1}{\exp\big[-\Lambda_z(t)\big]}\int_{[t,t_z^*]} U^-(\Phi_z(u), s + u)\big(\lambda_z(u)\big)^{\mathbb{1}_{t < t_z^*}}\exp\big[-\Lambda_z(u)\big]\mathrm{d}\mu_z(t)$$

This last equality allows to transform (4.7) into (4.8). $\qquad\square$

B.3 Equality (4.11)

Let $z \in E$ and $s \in [0, t_f)$. Remember that equality (4.11) states that if the functions $u \to U^-\big(\Phi_z(u), s + u\big)$ and $u \to \lambda_z(v)$ are continuous almost everywhere on $[0, t_z^*)$, then almost everywhere

$$\frac{\partial U^*\big(\Phi_z(v), s + v\big)}{\partial v} = U^*\big(\Phi_z(v), s + v\big)\lambda_z(v) - U^-\big(\Phi_z(v), s + v\big)\lambda_z(v)$$

Proof. We denote by T the time until the next jump after the trajectory has reached $Z_s = z$. For $0 \le h < t_z^*$, we define $\tau = \min(h, T)$.

$$U^*(z, s) = \mathbb{E}\big[\mathbb{1}_{\mathcal{D}}(\mathbf{Z})\big|Z_s = z\big]$$

$$= \mathbb{E}\Big[\mathbb{E}\big[\mathbb{1}_{\mathcal{D}}(\mathbf{Z})\big|Z_{s+\tau}\big]\Big|Z_s = z\Big]$$

$$= \mathbb{E}\Big[(\mathbb{1}_{\tau=h} + \mathbb{1}_{\tau<h})\mathbb{E}\big[\mathbb{1}_{\mathcal{D}}(\mathbf{Z})\big|Z_{s+\tau}\big]\Big|Z_s = z\Big]$$

$$= \mathbb{E}\Big[\mathbb{1}_{T=h}\,\mathbb{E}\big[\mathbb{1}_{\mathcal{D}}(\mathbf{Z})\big|Z_{s+h} = \Phi_z(h)\big]\Big|Z_s = z\Big] + \mathbb{E}\Big[\mathbb{1}_{T<h}\,\mathbb{E}\big[\mathbb{1}_{\mathcal{D}}(\mathbf{Z})\big|Z_{s+T}\big]\Big|Z_s = z\Big]$$

$$= U^*(\phi_z(h), s + h)\,\mathbb{E}\big[\mathbb{1}_{T=h}\big|Z_s = z\big] + \mathbb{E}\Big[\mathbb{1}_{T<h}\,U^*(Z_{s+T}, s + T)\Big|Z_s = z\Big]$$

$$= U^*(\phi_z(h), s + h)\exp\big[-\Lambda_z(h)\big] + \int_0^h \int_E K_{\Phi_z(u)}(z^+)U^*(z^+, s + u)\mathrm{d}\nu_{\Phi_z(u)}(z^+)\lambda_z(u)\exp\big[-\Lambda_z(u)\big]\mathrm{d}u$$

As $\lambda_z(\cdot)$ is continuous almost everywhere we have that almost everywhere:

$$U^*(z,s) = U^*(\phi_z(h), s+h)\,(1 - \lambda_z(0)h + o(h)) + \int_0^h U^\cdot(\Phi_z(u), s+u)\lambda_z(u)\exp\big[-\Lambda_z(u)\big]\mathrm{d}u$$

As $u \to U^\cdot(\phi_z(u), s+u)\lambda_z(u)$ is continuous almost everywhere, and we can do a Taylor approximation of the integral, which gives:

$$U^*(z,s) - U^*(\phi_z(h), s+h) = -\lambda_z(0)\,.h\,.U^*(\phi_z(h), s+h) + h\,.U^\cdot(z,s)\lambda_z(0) + o(h)$$

So $u \to U^*(\phi_z(u), s+u)$ is right-continuous almost everywhere. Therefore $U^*(\phi_z(h), s+h) = U^*(z,s) + o(1)$, and we get:

$$\frac{U^*(z,s) - U^*(\phi_z(h), s+h)}{h} = -\lambda_z(0)\,U^*(z,s) + U^\cdot(z,s)\lambda_z(0) + o(1)$$

Making h tends to zero we get that $u \to U^*(\phi_z(u), s+u)$ has a right-derivative in zero. Applying the same kind of reasoning in state $\Phi_z(-h)$ instead of z, we would find that the left-derivative exists and is equal to the right-derivative. So for almost every state $z \in E$,

$$\left(\frac{\partial U^*\big(\Phi_z(v), s+v\big)}{\partial v}\right)_{v=0} = U^*\big(\Phi_z(0), s+0\big)\lambda_z(0) - U^\cdot\big(\Phi_z(0), s+0\big)\lambda_z(0)$$

Applying the same reasoning in a state $\Phi_{z_o}(v)$ instead of z and using the additivity of the flow, we get that almost everywhere:

$$\forall z_o \in E, v > 0, \quad \frac{\partial U^*\big(\Phi_{z_o}(v), s+v\big)}{\partial v} = U^*\big(\Phi_{z_o}(v), s+v\big)\lambda_{z_o}(v) - U^\cdot\big(\Phi_{z_o}(v), s+v\big)\lambda_{z_o}(v)$$

$\square$

References

[1] C.-E. Bréhier, T. Lelièvre and M. Rousset, Analysis of adaptive multilevel splitting algorithms in an idealized case. *ESAIM: PS* **19** (2015) 361–394.

[2] J. Bucklew, Introduction to rare event simulation. Springer Science & Business Media (2013).

[3] V. Caron, A. Guyader, M. Zuniga and B. Tuffin, Some recent results in rare event estimation. *ESAIM: Proc.* **44** (2014) 239–259.

[4] M. Čepin, Assessment of power system reliability: methods and applications. Springer Science & Business Media (2011).

[5] F. Cérou, P. Del Moral, F. Le Gland and P. Lezaud, Genetic genealogical models in rare event analy-sis. *ALEA Latin Am. J. Probab. Math. Stat.* **1** (2006) 181–203.

[6] F. Cérou, B. Delyon, A. Guyader and M. Rousset, On the asymptotic normality of adaptive multilevel splitting. *SIAM/ASA J. Uncert. Quant.* **7** (2019) 1–30.

[7] J.C. Chan, P.W. Glynn, D.P. Kroese et al., A comparison of cross-entropy and variance minimization strategies. *J. Appl. Probab.* **48** (2011) 183–194.

[8] H. Chraibi, Dynamic reliability modeling and assessment with PyCATSHOO: application to a test case. *PSAM congress* (2013).

[9] H. Chraibi, A. Dutfoy, T. Galtier and J. Garnier, Optimal input potential functions in the interacting particle system method. Preprint arXiv:1811.10450 (2018).

[10] H. Chraibi, J.-C. Houbedine and A. Sibler, Pycatshoo: Toward a new platform dedicated to dynamic reliability assessments of hybrid systems. *PSAM congress* (2016).

[11] M.H. Davis, Piecewise-deterministic Markov processes: A general class of non-diffusion stochastic models. *J. Roy. Stat. Soc. Ser. B (Methodological)* **46** (1984) 353–388.

[12] M.H. Davis, Markov Models & Optimization. In Vol. 49 of Monogaphs on Statistics and Applied Probability. CRC Press (1993).

[13] P.-T. de Boer, D.P. Kroese, S. Mannor and R.Y. Rubinstein, A tutorial on the cross-entropy method. *Ann. Oper. Res.* **134** (2005) 19–67.

[14] P. Del Moral and J. Garnier, Genealogical particle analysis of rare events. *Ann. Appl. Probab.* **15** (2005) 2496–2534.

[15] F. Dufour, H. Zhang and B. de Saporta, Numerical methods for simulation and optimization of piecewise deterministic Markov processes: application to reliability. John Wiley & Sons (2015).

[16] P. Dupuis and H. Wang, Importance sampling, large deviations, and differential games. *Stochastics* **76** (2004) 481–508.

[17] P. Heidelberger, Fast simulation of rare events in queueing and reliability models. *ACM Trans. Model. Comput. Simul.* **5** (1995) 43–85.

[18] I. Kuruganti, Importance sampling for markov chains: computing variance and determining optimal measures. In *Proceedings of the 1996 Winter Simulation Conference.* IEEE (1996) 273–280.

[19] P.-E. Labeau, A Monte-Carlo estimation of the marginal distributions in a problem of probabilistic dynamics. *Reliab. Eng. Syst. Safety* **52** (1996) 65–75.

[20] P.-E. Labeau, Probabilistic dynamics: estimation of generalized unreliability through efficient Monte-Carlo simulation. *Ann. Nucl. Energy* **23** (1996) 1355–1369.

[21] E. Lewis and F. Böhm, Monte-Carlo simulation of Markov unreliability models. *Nucl. Eng. Des.* **77** (1984) 49–62.

[22] M. Marseguerra and E. Zio, Monte-Carlo approach to psa for dynamic process systems. *Reliab. Eng. Syst. Safety* **52** (1996) 227–241.

[23] P. Metzner, C. Schütte and E. Vanden-Eijnden, Transition path theory for markov jump processes. *Multis. Model. Simul.* **7** (2009) 1192–1219.

[24] J. Morio, M. Balesdent, D. Jacquemart and C. Vergé, A survey of rare event simulation methods for static input–output models. *Simul. Model. Pract. Theory* **49** (2014) 287–304.

[25] M. Ramakrishnan, Unavailability estimation of shutdown system of a fast reactor by Monte-Carlo simulation. *Ann. Nucl. Energy* **90** (2016) 264–274.

[26] T. Rolski, H. Schmidli, V. Schmidt and J.L. Teugels, Stochastic processes for insurance and finance. In Vol. 505 of *Wiley Series in Probability and Statistics.* John Wiley & Sons (2009).

[27] D. Siegmund, Importance sampling in the Monte-Carlo study of sequential tests. *Ann. Stat.* **4** (1976) 673–684.

[28] N. Whiteley, A.M. Johansen and S. Godsill, Monte carlo filtering of piecewise deterministic processes. *J. Comput. Graph. Stat.* **20** (2011) 119–139.

[29] H. Zhang, F. Dufour, Y. Dutuit and K. Gonzalez, Piecewise deterministic Markov processes and dynamic reliability. *Proc. Inst. Mech. Eng., Part O: J. Risk Reliab.* **222** (2008) 545–551.

[30] E. Zio, The Monte-Carlo simulation method for system reliability and risk analysis. Springer (2013).

[31] P. Zuliani, C. Baier and E.M. Clarke, Rare-event verification for stochastic hybrid systems. In *Proceedings of the 15th ACM international conference on Hybrid Systems: Computation and Control.* ACM (2012) 217–226.