The Annals of Applied Probability
2017, Vol. 27, No. 4, 2515–2538
DOI: 10.1214/16-AAP1266
© Institute of Mathematical Statistics, 2017

MINIMAX OPTIMALITY IN ROBUST DETECTION OF A DISORDER TIME IN DOUBLY-STOCHASTIC POISSON PROCESSES

BY NICOLE EL KAROUI*, STÉPHANE LOISEL† AND YAHIA SALHI†

Université Pierre et Marie Currie and Université Claude Bernard Lyon 1†*

We consider the minimax quickest detection problem of an unobservable time of proportional change in the intensity of a doubly-stochastic Poisson process. We seek a stopping rule that minimizes the robust Lorden criterion, formulated in terms of the number of events until detection, both for the worst-case delay and the false alarm constraint. This problem, introduced by Page [*Biometrika* **41** (1954) 100–115], has received more attention in the continuous path framework (for Wiener processes) than for point processes, where optimality results only concern the Bayesian framework [In *Advances in Finance and Stochastics* (2002) 295–312, Springer, Berlin]. We prove the CUSUM optimality conjectured but not solved for the Poisson case of the CUSUM strategy in the general setting of the stochastic intensity framework. We use finite variation calculus and elementary martingale properties to characterize the performance functions of the CUSUM stopping rule in terms of scale functions. These are solutions of some delayed differential equations that can be solved simply. The case of detecting a decline in intensity is easier to study, because the performance functions are continuous. In the case of a rise where the performance functions are not continuous, differential calculus requires using a discontinuous local time at the discontinuity level, difficult to estimate. The conjecture was considered proven by the community, but the proof was still lacking for this reason. Some numerical considerations are provided at the end of the article.

1. Introduction. In the quickest detection problem for point processes, one observes the jumps of an inhomogeneous counting process whose intensity suddenly changes at some unobservable disorder time, but remains "stable" beforehand and after, in some sense. The process being sequentially observed, the problem is then to detect the change-point as quickly as possible after it happens. As pointed out in the introduction of Basseville and Nikiforov (1993) [4], *"It should be clear that abrupt changes by no means imply changes with large magnitude. Many change detection problems are concerned with the detection of small changes"*.

The classical fields of applications of such detection problems include, among others, queueing theory, survival analysis, reliability and denial-of-service attacks

Received April 2016; revised July 2016.

MSC2010 subject classifications. Primary 62L15; secondary 97M30.

Key words and phrases. Change-point, robust sequential detection, doubly stochastic Poisson process.

on internet services [4]. Our main practical motivation is the fast detection of the onset of mortality shifts, where a proportional relationship between two mortality intensities, for example, the insured mortality intensity and a reference, is assumed. A sudden change in the proportional relationship can induce serious financial consequences, and it is necessary to react as soon as data would suggest it; see Barrieu et al. (2012) [3].

A similar problem also arises in prompt detection of shifts in insurance claims arrival. Note that in these examples, in general, *no information* is known about the distribution of the date of change. Therefore, in this paper, we consider a non-Bayesian setting in which the change-point is unknown but deterministic, in the spirit of the early papers of Page (1954) [19] and Lorden (1971) [13]. This framework, based on the so-called Lorden procedure formulated as a minimax problem, belongs to the family of *robust* optimization, popular in statistical learning (Hastie et al. (2009) [10]).

The minimax robust detection problem in continuous time has gained renewed interest since the 1990s. For example, the problem of detecting an abrupt change in the drift of a Wiener process is well understood. In particular, the so-called cumulative sums strategy [19] (CUSUM for short) has been shown to be optimal with regards to the Lorden procedure; see Shiryaev (1996) [26], Beibel (1996) [5] and Moustakides (2004) [17]. Shiryaev (2009) [27] sheds an interesting light on the history of this problem, and different developments depending on the field of application.

In the classical sequential test analysis for continuous time processes, the statistic is the usual Log Sequential Probability Ratio process (LSPR-process) between the reference probability $\mathbb{P}$ (null assumption) and the alternative assumption $\widetilde{\mathbb{P}}$ (H1 assumption); see Dvoretzky, Kiefer and Wolfowitz (1953) [8] for the Poisson process. In the minimax detection problem, the test is based on the CUSUM strategy, consisting in sounding an alarm as soon as the LSPR-process related to $\mathbb{P}$ (no change) and $\widetilde{\mathbb{P}}$ (immediate change), *reflected at its maximum or at its minimum*, hits a barrier $m > 0$. The optimality relies on the characterization of the optimal detection time through its performance, that is, the time until detection and the false alarm frequency, which are of critical importance. In the Wiener case, the performance functions have a very simple and universal form, which is a key element in the derivation of the optimality [17]. In this paper, we are interested in the extension of this optimality result to the CUSUM strategies in the doubly stochastic Poisson process, by overcoming difficulties due to jumps and to the complex form of the performance functions, strongly related to the scale functions introduced in Lévy processes theory, for example, Bertoin (1998) [6], Pistorius (2004) [22], and reference therein.

Note that, in discrete time, only asymptotic optimality results were shown by Mei et al. (2011) [14].

The conjecture that the CUSUM strategy is optimal in the Poisson case for the minimax problem was considered by the community to be solved. However, due

to the discontinuity of some performance functions at the detection level, this is not the case for a rise in intensity ($\rho > 1$); see Section 5. The sketches of the proofs in Moustakides (2007) [18] and in Poor and Hadjiliadis (2009) [24], Section 6.4.4, pages 154–157, can be adapted in the case of a decline in intensity ($\rho < 1$) to show optimality, but do not work in the increasing case. In this paper, we directly prove the CUSUM optimality conjectured in the Poisson case in the stochastic intensity setting. To this end, we consider a modified version of Lorden's criterion that indeed corresponds to its natural extension to the stochastic intensity framework (and is equivalent to the classical Lorden criterion in the Poisson case).

The article is organized as follows. In Section 2, we introduce the robust optimization problem, with a discussion about the criterion to be associated with this minimax problem. This discussion is followed by a rigorous presentation of the basic tools in the counting process framework used throughout the paper. In Section 3, we are concerned with CUSUM processes, which are the log-sequential probability ratio processes reflected at their running maximum or minimum, and write them as solutions of some differential equations driven by the counting process. We also define the performance functions associated with the hitting time by the CUSUM processes of a barrier m, making the distinction between a rise and decline in the intensity. Stochastic differential calculus is used to extend Itô's formula to a.e. differentiable monotonic functions with one jump, thanks to a discontinuous local time. We also prepare for the proof of optimality, by the introduction of new and suitable martingales. In Section 4, we solve very elementary classical and less classical delayed equations with the help of the scale functions and their primitives from Lévy processes theory. Then we obtain closed formulas for the performance functions associated with a positive CUSUM barrier. Section 5 is dedicated to the proof of optimality of the CUSUM procedure with a given false alarm constraint. The proof relies on a modified criterion following ideas from Shiryaev (1996) [26] and Moustakides (1986, 2002) [15, 16]. To the best of our knowledge, it is the first optimality result in the rising Poisson case, for the minimax problem. Lastly, in Section 6 we discuss some numerical methods suited to solve the delayed equation associated with the scale and performance functions, and overcome numerical instability problems.

2. Problem formulation.

2.1. *Counting processes.*

Counting processes and intensity. We consider a counting process N defined on a probability space $(\Omega, \mathcal{F}_\infty, \mathbb{P})$, counting the number of arrival times $(T_i)_{i\geq 1}$ of some event family. The $(T_i)_{i\geq 1}$ form a strictly increasing sequence of times describing the process as $N_t = \sum_{n=1}^{\infty} \mathbf{1}_{\{T_n \leq t\}}$. (a) A typical example is the standard Poisson process with intensity λ whose arrival times are generated by a family of independent exponential variables τ_i with expectation $1/\lambda$, such that $T_n = \sum_{i=1}^{n} \tau_i$.

Then N_t has a Poisson distribution with parameter $\Lambda_t = \lambda t$ whose moment generating function is $\mathbb{E}[\exp(\alpha N_t)] = \exp(\lambda t(e^\alpha - 1))$, for $\alpha > 0$. (b) Here, we are concerned with a *doubly stochastic* point process [9], where the uncertainty concerns both the arrival times and intensity. Therefore, the probability space $(\Omega, \mathcal{F}_\infty, \mathbb{P})$ is equipped with some filtration $\mathbb{F} = (\mathcal{F}_t)$ satisfying the usual conditions. The $(\mathbb{P}, \mathbb{F})$-adapted counting process N is characterized by its positive (nonnegative) $(\mathcal{F}_t)$-adapted (predictable) intensity λ_t.

The increasing (nondecreasing) process $\Lambda_t = \int_0^t \lambda_s \, ds$ is said to be the *cumulative intensity* of N if and only if the compensated process $M_t = N_t - \Lambda_t$ is a $\mathbb{P}$-martingale (if Λ_t is $\mathbb{P}$-integrable for any $t \geq 0$) and a local martingale (if Λ_t is only finite a.s.). In the following, we assume $\Lambda_t \to \infty$ when t goes to ∞. When Λ_t is deterministic, the process N is an inhomogeneous Poisson process whose increments $(N_{t+h} - N_t)_{\{h>0\}}$ are independent of $\mathcal{F}_t$.

Log-likelihood ratio. The notion of *exponential martingale*, $\mathcal{L}_t^\alpha = \exp(\alpha N_t - \Lambda_t(e^\alpha - 1))$ plays the same role as that of the Laplace transform for the Poisson process. $(\mathcal{L}_t^\alpha)_t$ is sometimes called a Wald martingale [2], pages 21–23. A more convenient form for our purpose is based on a representation in terms of $\eta = \exp(\alpha)$, such that the exponential martingale $\mathcal{L}_t^\alpha$ becomes the process $\mathcal{E}_t^\eta = \eta^{U_t^\eta}$, where

$$\ln_\eta(\mathcal{E}_t^\eta) = U_t^\eta = N_t - \beta(\eta)\Lambda_t, \tag{1}$$

$$\text{with } \beta(\eta) = (\eta - 1)/\log(\eta) = \int_0^1 \eta^x \, dx, \qquad \eta \neq 1 \text{ and } \beta(1) = 1. \tag{2}$$

The function $\beta(\eta)$ is the Laplace transform of the uniform distribution on $[0, 1]$. In particular, for $\eta > 1$, β is increasing and $\beta(\eta) > 1$, and for $\eta < 1$, β is decreasing and $\beta(\eta) < 1$.

A natural candidate for the likelihood process associated with an intensity process $\rho\Lambda_t$, with $\rho \neq 1$, is the exponential process $(\mathcal{E}_t^\rho = \rho^{U_t^\rho})$, assumed to be a uniformly integrable martingale such that $\mathbb{E}(\mathcal{E}_T^\rho) = \mathcal{E}_0^\rho = 1$. We denote by $\widetilde{\mathbb{P}}$ the probability measure equivalent to $\mathbb{P}$ with density process $(\rho^{U_t^\rho})$. The $\widetilde{\mathbb{P}}$-likelihood of $\mathbb{P}$ with respect to $\widetilde{\mathbb{P}}$ is the $\widetilde{\mathbb{P}}$-exponential martingale $\rho^{-U_t^\rho} = (1/\rho)^{U_t^\rho} = (1/\rho)^{N_t - \beta(1/\rho)\rho\Lambda_t}$, since $\rho\beta(1/\rho) = \beta(\rho)$. This $\widetilde{\mathbb{P}}$-martingale property of $(1/\rho)^{U_t^\rho}$ is true only if $\tilde{M}_t = N_t - \rho\Lambda_t$ is a $\widetilde{\mathbb{P}}$-martingale.

2.2. *Detection problem.*

Detection problem. We are interested in detecting any disorder that has affected the intensity of a counting process, where changes may occur at unobservable, nonrandom times $\theta \in [0, \infty)$. Formally, we consider an intensity process such that

$$\check{\lambda}_t^\theta = \lambda_t \mathbf{1}_{\{t<\theta\}} + \rho\lambda_t \mathbf{1}_{\{t\geq\theta\}}, \qquad \rho \neq 1. \tag{3}$$

The best-suited and most frequently used framework for the detection problem is to keep the observed process N under the initial probability measure $\mathbb{P}$, and to change the occurrence likelihood using an exponential martingale density $\mathcal{L}_t^\theta$, equal to 1 between 0 and θ, and coinciding with $(\mathcal{E}_t^\rho/\mathcal{E}_\theta^\rho)$ on (θ,∞). The likelihood process $\mathcal{L}_t^\theta$ defines a new probability measure $\mathbb{P}^\theta$.

When a change does not appear, that is, $\theta=\infty$ and $\check{\lambda}_t^\infty=\lambda_t$, we keep the notation $(\lambda_t,\mathbb{P})$ instead of $(\lambda_t^\infty,\mathbb{P}_\infty)$. When a change appears *instantaneously*, $\check{\lambda}_t^0=\rho\lambda_t$ is sometimes denoted $\tilde{\lambda}_t$, and associated with a probability measure $\widetilde{\mathbb{P}}$ over the frequently-used notation $\mathbb{P}_0$ [12]. This notation from Kyprianou (2014) [12] is used to limit the proliferation of indexes.

The sequential conditional probability ratio (SCPR in short) between t and T is given by $\mathcal{E}_{t,T}^\rho=\mathcal{E}_T^\rho/\mathcal{E}_t^\rho$, and plays a key role both in sequential testing theory and quickest detection problems. Here, we are more concerned with quickest detection procedures [4]. Moreover, as our initial motivation is related to detection of changes in longevity patterns associated with nondeterministic intensities, we are more interested in robust methods than Bayesian ones.

Robust detection problem. Our main goal is to find a stopping rule T based on the filtration $\mathbb{F}$ to optimally raise the alarm for the breakpoint, with no *a priori* information on θ. The first step is to define a measure that quantifies the performance of the detection procedure.

(a) For robust estimation, the Lorden [13] procedure advocates penalizing the detection delay via its worst-case value

$$\widetilde{C}_{\mathrm{Lor}}(T)=\sup_{\theta\in[0,\infty]}\operatorname*{ess\,sup}_{\omega}\mathbb{E}^\theta\big[(T-\theta)^+|\mathcal{F}_\theta\big], \tag{4}$$

where the "ess sup" takes the "worst possible observed date before the change" in the sense of providing no information on the true change, as explained in [14].[1] In equation (4), the expectation is taken with respect to the probability measure $\mathbb{P}^\theta$, under which the counting process N has the $\mathbb{P}^\theta$-intensity given in (3). In fact, $\mathbb{P}^\theta$ can be replaced by the probability measure $\widetilde{\mathbb{P}}$ $(\theta=0)$, since the criterion only depends on what happens after the change. In addition, *false alarms* are monitored using the average run length to false alarm given by $\mathbb{E}[T]$, since the nominal probability $\mathbb{P}$ is associated with no change. With these metrics, the robust detection procedure reduces to solving the following minimax optimization problem for a given threshold π:

$$\inf_T\widetilde{C}_{\mathrm{Lor}}(T),\qquad \text{with constraint } \mathbb{E}(T)\geq\pi. \tag{5}$$

[1]Mathematically, the ess sup it is the smallest constant number y_θ such that $y_\theta\geq\mathbb{E}^\theta[(T-\theta)^+|\mathcal{F}_\theta]$, $\mathbb{P}^\theta$-a.s.

This framework serves to design the optimal stopping rule T^* with the desired characteristics based on a linear delay criterion, well adapted to constant intensities. Nevertheless, another interpretation may be given, better suited to our problem, where only the multiplicative ratio ρ is known, and not the stochastic intensity itself.

(b) Given that a doubly stochastic point process may be reduced to a Poisson process by *time rescaling* [9], we are looking for a criterion which is robust with respect to this procedure, that is, both in the delay criterion and the false alarm constraint. Observe that for Poisson processes, $\mathbb{E}^\theta[(N_T - N_\theta)^+|\mathcal{F}_\theta] = \widetilde{\mathbb{E}}[N_T - N_{T\wedge\theta}|\mathcal{F}_\theta] = \lambda\rho\widetilde{\mathbb{E}}[T - T\wedge\theta|\mathcal{F}_\theta]$, and $\mathbb{E}(N_T) = \lambda\mathbb{E}(T)$. Thus, the number of events until detection is a good candidate to take into account in the optimization criterion. Formally, we let

$$\widetilde{\Gamma}_\theta^T = \mathbb{E}^\theta\big[(N_T - N_\theta)^+|\mathcal{F}_\theta\big] = \rho\widetilde{\mathbb{E}}\bigg[\int_{\theta\wedge T}^T \lambda_s\,ds\bigg|\mathcal{F}_\theta\bigg], \qquad \mathbb{E}(N_T) = \mathbb{E}(\Lambda_T). \tag{6}$$

The minimax optimization problem is now to minimize with respect to the stopping times T, with the constraint $\mathbb{E}(N_T) \geq \pi$, the generalized Lorden criterion:

$$\widetilde{C}(T) = \sup_{\theta\in[0,\infty]} \operatorname*{ess\,sup}_{\omega} \widetilde{\mathbb{E}}\big[(N_T - N_\theta)^+|\mathcal{F}_\theta\big]. \tag{7}$$

In the same vein, Moustakides (2004) [17] has initiated a modification of the Lorden criterion that replaces expected delay with Kullback–Leibler divergence.

2.3. CUSUM *processes*. In this section, we set the value of the parameter $\rho \neq 1$, and frequently drop out the indexation on ρ from notation when it is not ambiguous. We also use indifferently the parameter $\beta = \beta(\rho)$, given the one-to-one correspondence between β and ρ.

Moreover, for ease of use, for any càdlàg (right-continuous with left limit) process Z, the running supremum is the nondecreasing càdlàg process $\bar{Z}_t = \sup_{s\leq t} Z_s$.

CUSUM *processes*. In estimation theory, the process of interest is usually the maximum likelihood process. In detection problems, since the parameter is a time, the maximum likelihood must be taken over the dates θ before t, so that the process of interest is $\sup_{\theta\leq t} \rho^{U_t - U_\theta}$.

(a) Consider the case $\rho > 1$. The maximum likelihood process is given by ρ^{V_t}, where $V_t = \sup_{s\leq t}(U_t - U_s) = U_t + \sup_{s\leq t}(-U_s)$. Here, V_t is the process U_t reflected at 0. By introducing the reversed process $X = -U$ (also called *dual process* in insurance theory [2]) which has negative jumps and nondecreasing continuous part $\rho\Lambda_t$, we see that $\sup_{s\leq t}(-U_s) = \bar{X}_t$ is a continuous process with support $\{X_t = 0\}$. The processes V and U have jumps of size 1 at the same times. The

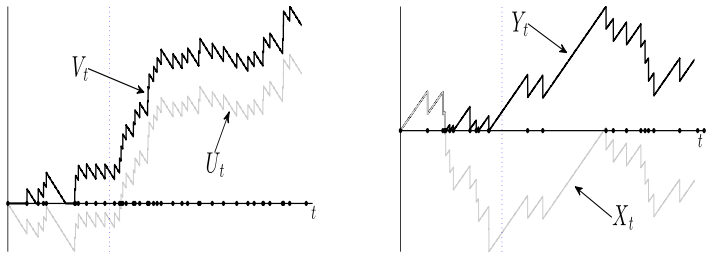

FIG. 1. *Simulated* CUSUM *processes* V, U *and* Y, X *with constant intensity* 3. *The post-change parameter* ρ *is set to* 1.5 (*left*) *and* 0.5 (*right*), *respectively.*

CUSUM process V may also be viewed as the process X reflected at its continuous maximum, using the terminology of Lévy processes. A typical path of the process V is given in Figures 1–2.

(b) When $\rho < 1$, the function ρ^z is decreasing and the maximum likelihood process is $(1/\rho)^{Y_t}$, where $Y_t = \sup_{\theta \le t} U_\theta - U_t = \bar{U}_t - U_t = X_t + \bar{U}_t$. The process Y is the process U reflected at its maximum, or the process X reflected at 0. The process $\bar{U}$ is a nondecreasing step process, and the epochs when a new supremum of $\bar{U}$ is reached, that is, $\bar{U}_t = U_t$, are arrival times of N, for which $Y_t = 0$. The jumps of Y are negative and *of absolute value less than* 1. Typical paths of the processes V and Y are given in Figures 1–2.

(c) The CUSUM stopping rule is defined as the first time where the CUSUM process V or Y exceeds a given level $m > 0$, that is, for $\rho > 1$, $\tau_m^V = \inf\{t, V_t \ge m\}$ and for $\rho < 1$, $\tau_m^Y = \inf\{t, Y_t \ge m\}$. The next section is dedicated to developing tools and self-contained arguments to calculate the CUSUM rule's performance given our particular framework.

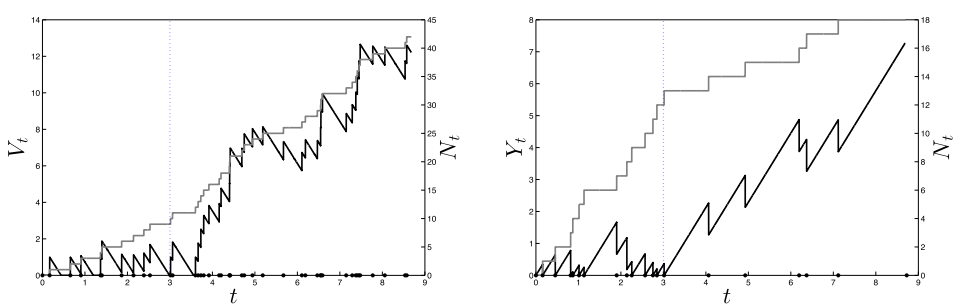

FIG. 2. *Simulated* CUSUM *processes* V *and* Y *with a change-point at time* $\theta = 3$ *and constant intensity* 3. *The post-change parameter* ρ *is set to* 1.5 (*left*) *and* 0.5 (*right*), *respectively. In each plot, the* y*-axis scale on the left is associated with the* CUSUM *process, and on the right with the nondecreasing process* N. *Vertical dotted lines indicate jump times* T_i.

3. Reflected processes and finite variation stochastic calculus. We use the name *reflected process* in place of CUSUM process when no restriction is made on the value of the parameter ρ, that is V is also defined for $\rho < 1$ and Y is also defined for $\rho > 1$. When the intensity is constant, the vocabulary is that of Lévy process theory as in Bertoin (1998) [6] and Kyprianou (2014) [12] for instance.

To use Markov type conditional properties, we need to define the processes involved for an initial condition different from 0. For the processes U and X, we add an initial linear condition Z_0, so that $U_t(Z_0) = Z_0 + N_t - \beta(\rho)\Lambda_t$, and $X_t(Z_0) = Z_0 - N_t + \beta(\rho)\Lambda_t$. For the reflected processes, we first define the running maximum processes starting from 0, $\bar{X}_t^{ad}(Z_0) = (\bar{X}_t - Z_0)^+$, and $\bar{U}_t^{ad} = (\bar{U}_t - Z_0)^+$. Then the reflected processes starting from Z_0 are defined as the sum of the two processes, $V_t(Z_0) = U_t(Z_0) + \bar{X}_t^{ad}(Z_0)$ and $Y_t(Z_0) = X_t(Z_0) + \bar{U}_t^{ad}(Z_0)$; the dependence in the initial condition is now increasing and convex. Hereafter, as a matter of notation, we frequently use $\mathbb{P}_x$ when referring to the conditional probability measure $\mathbb{P}_x = \mathbb{P}(\cdot|Z_0 = x)$.

3.1. *Reflected processes as solutions of stochastic differential equations.* Since all processes are of finite variation with a finite number of jumps, and almost everywhere (a.e.) differentiable, it is easy to develop a pathwise differential calculus. However, for ease of reading, we recall the main differential formula. For this, let ϕ be a right continuous finite variation function defined on a real interval $[z_{\min}, \infty)$, ($z_{\min}$ possibly infinite) with a finite number of jumps $\phi(\alpha) - \phi(\alpha-) = \delta\phi(\alpha)$, and almost surely differentiable with derivative ϕ'. Such functions are hereafter called "*regular finite variation* functions". Then we have the integral representation:

$$\phi(z) = \phi(z_{\min}) + \int_{z_{\min}}^{z} \phi'(u)\,du + \sum_{z_{\min}<\alpha\le z} \phi(\alpha) - \phi(\alpha-). \tag{8}$$

In the sense of Schwartz distributions, the function ϕ is differentiable, with derivative $\phi(du) = \phi'(u)\,du + \sum_{\{\alpha,\delta\phi(\alpha)\neq 0\}} \delta\phi(\alpha)\delta_\alpha(du)$, where $\delta_\alpha(du)$ is the Dirac measure.

3.1.1. *Reflected processes as solutions of stochastic differential equations.* The study of the differential properties of the reflected processes $V_t(Z_0)$ and $Y_t(Z_0)$ is based on the properties of the supremum processes $\bar{X}_t^{ad}(Z_0)$ and $\bar{U}_t^{ad}(Z_0)$. Recall that $\bar{X}^{ad}$ is continuous and only increases when $\{V = 0\}$, whereas the process $\bar{U}^{ad}(Z_0)$ is a pure jump process such that $\bar{U}_t^{ad} - \bar{U}_{t-}^{ad} = 1 - Y_{t-}$, since $Y_t = 0$. These observations suggest the following differential properties of the reflected processes.

PROPOSITION 1.

(a) *The* CUSUM *process V_t is the unique nonnegative solution of the pathwise differential equation*:

$$dV_t = dN_t - \beta\mathbf{1}_{(0,\infty)}(V_s)\,d\Lambda_s, \qquad \textit{implying that } d\bar{X}_t^{ad} = \beta\mathbf{1}_{\{V_t=0\}}\,d\Lambda_t. \tag{9}$$

(b) *The reflected process Y_t is the unique nonnegative solution of the differential equation, with jump function $-j(y)$ with $j(y) = y \wedge 1$:*

$$dY_t = -j(Y_{t-})\,dN_t + \beta\,d\Lambda_t, \quad \text{implying that } d\bar{U}_t^{ad} = \mathbf{1}_{\{Y_t=0\}}(1 - Y_{t-})\,dN_t. \tag{10}$$

(c) *In the Poisson case, both reflected processes V and Y are Markov processes.*

PROOF. See the Appendix. □

3.1.2. *Some remarkable functions related to martingales.* As a corollary, we obtain the differential decomposition of the processes $g_m(Y_t)$ and $h_m(V_t)$, when the functions g_m and h_m are regular finite variations solutions of delayed differential equations (DDE), denoted DDE(β), restricted to the interval $[0, m]$:

$$\beta g_m'(x) = g_m(x) - g_m\big((x-1)^+\big) - 1, \quad \text{with } g_m(0) = 0 \text{ (Cauchy problem)}, \tag{11}$$

$$\beta h_m'(x) = h_m(x+1) - h_m(x) + 1, \qquad h_m(0) = 0, \quad \text{with } h_m'(0) = 0 \text{ (Neumann problem)}. \tag{12}$$

Both functions are extended by the value 0 for $x \geq m$. The solutions of these DDEs are studied in the next section, where it is proved that the function g_m is continuous, whereas h_m is continuous on $(0, m)$ with positive jump at m. The same properties hold true for the functions $\tilde{g}_m$ and $\tilde{h}_m$, solutions of the same system where β is replaced by $\tilde{\beta}$, with $\tilde{\rho} = 1/\rho$, and $\tilde{\beta} = \beta(\tilde{\rho}) = \beta(1/\rho) = \beta(\rho)/\rho$. We start with the simplest case of continuous solutions.

Continuous regular functions and Y processes. We start with the function g_m applied to the reflected process Y, solution of (10). The martingale formulation allows us to make no distinction between doubly stochastic point process and a Poisson process.

PROPOSITION 2. *Let g_m be the continuous solution of the* DDE(β) (11):

(i) (a) *The process $G_t^m = g_m(Y_t) - g_m(Y_0) + N_t^{m,Y}$ is a finite variation $\mathbb{P}_y$-martingale with decomposition:*

$$dG_t^m = dg_m(Y_t) + dN_t^{m,Y} = -\beta g_m'(Y_{t-})\,dM_t, \tag{13}$$

where M_t is the $\mathbb{P}_y$-martingale $dM_t = dN_t - d\Lambda_t$ and $dN_t^{m,Y} = \mathbf{1}_{[0,m]}(Y_{t-})\,dN_t$.

(b) *Let τ_m^Y be the* CUSUM *stopping time defined in Section 2 as $\tau_m^Y = \inf\{t \geq 0, Y_t \geq m\}$. Then*

$$g_m(y) = \mathbb{E}_y(N_{\tau_m^Y}). \tag{14}$$

(ii) *Let $\tilde{g}_m$ be a continuous regular solution of the* DDE($\tilde{\beta}$).
   (a) $\widetilde{G}_t^m = \tilde{g}_m(Y_t) - \tilde{g}_m(Y_0) + N_t^{m,Y}$ *is a* $\tilde{\mathbb{P}}_y$*-martingale and* $d\widetilde{G}_t^m = -\tilde{\beta}\tilde{g}_m'(Y_{t-})\,d\tilde{M}_t$.
   (b) *Set* $\phi(y) = \rho^y(\tilde{g}_m(0) - \tilde{g}_m(y))$. *The process* $\bar{G}_t^{m,\tilde{\rho}} = \rho^{Y_t}(\tilde{g}_m(0) - \tilde{g}_m(Y_t)) + \rho\int_0^t \rho^{Y_{s-}}\,dN_s^m$ *is the* $\mathbb{P}_y$*-martingale* $\bar{G}_t^{m,\tilde{\rho}} = -\beta \times \int_0^t \phi'(Y_{s-})\,dM_s$.

PROOF. (i)(a) The proof of (13) is immediate from the differential equation (10) of Y, and from the differential formula (8) since g_m is a continuous finite variation function. We have also used the fact that $g_m'(z) = 0$ for $z > m$.

(b) The interpretation of g_m as the expectation of the number of events before crossing the level m is a simple consequence of the stopping theorem for martingales, since g is bounded on the compact interval $[0, m]$.

(ii)(a) The proof is the same as earlier, because the processes (X^ρ, Λ) and $(X^{\tilde{\rho}}, \rho\Lambda)$ defined in Section 2 are the same.

(b) The second property needs some explanation. Set $\tilde{f}_m(y) = \tilde{g}_m(0) - \tilde{g}_m(y)$, so that $\phi(y) = \tilde{\rho}^y(\tilde{g}_m(0) - \tilde{g}_m(y)) = \tilde{\rho}^y\tilde{f}_m(y)$. Then $\beta\phi'(y) = -\log\rho\beta\phi(y) + \tilde{\rho}^y\beta\tilde{f}_m'(y)$. The fact that $\tilde{f}_m(0) = 0$ allows us to put aside the influence of a purely discontinuous local time at zero, by using the fact that the derivative of $\tilde{f}_m$ is equal to $\tilde{\beta}\tilde{f}_m'(y) = \tilde{f}_m(y) - \tilde{f}_m((y-1)^+) + 1 = \tilde{f}_m(y) - \mathbf{1}_{\{y>1\}}\tilde{f}_m(y-1) + 1$. Thus, from the identities $\log(\rho)\beta = \rho - 1$ and $\beta = \rho\tilde{\beta}$, we can write

$$\begin{aligned}\beta\phi'(y) &= -(\rho-1)\phi(y) + \rho\phi(y) - \phi(y-1)\mathbf{1}_{\{y>1\}} + \rho\tilde{\rho}^y \\ &= \phi(y) - \phi\big((y-1)^+\big) + \rho\tilde{\rho}^y,\end{aligned}$$

and consequently,

$$\begin{aligned}d\phi(Y_t) &= \big[\phi\big((Y_t-1)^+\big) - \phi(Y_{t-})\big]\,dN_t + \big[\phi(Y_{t-}) - \phi\big((Y_{t-}-1)^+\big)\big]\,d\Lambda_t \\ &\quad + \rho\tilde{\rho}^{Y_{t-}}\,d\Lambda_t \\ &= \big[\phi\big((Y_t-1)^+\big) - \phi(Y_{t-}) - \rho\tilde{\rho}^{Y_{t-}}\big]\,dM_t + \rho\tilde{\rho}^{Y_{t-}}\,dN_t \\ &= -\beta\phi'(Y_{t-})\,dM_t - \rho\tilde{\rho}^{Y_{t-}}\,dN_t.\end{aligned}$$

The proof is complete. □

Monotonic function with one jump at m and the V process. For the reflected process V, we are concerned with the decreasing right-continuous function h_m, null on $[m, \infty)$, with one negative jump $\delta h_m(m) = -h_m(m-)$ at m, solution of the delayed Neumann equation (12). Then the process $h_m(V_t)$ has additional discontinuities when $V_t = V_{t-} = m$.

(i) *Discontinuous local time.* Since the process V up-crosses the level m only by jumps, the number of continuous crossings of the level m by V is discrete, and

it is only given by down-crossings. Then the number $J^{m,V}$ of continuous crossings of the level m by V, sometimes called *discontinuous local time* at m, is defined by

$$J_t^{m,V} = \lim_{\varepsilon \mapsto 0} \sum_{\varepsilon \le s \le t} \mathbf{1}_{\{V_{s-\varepsilon} > m > V_{s+\varepsilon},\, V_s = V_{s-} = m\}}. \tag{15}$$

Thus, we have to add the term $h_m(m-) J_t^{m,V}$ to the classical differential formula.

PROPOSITION 3. *Let h_m be the right continuous solution of* (12):

(i) (a) *The process $H_t^m = h_m(V_t) - h_m(V_0) + N_t^{m,V} + h_m(m-) J_t^{m,V}$ is a finite variation $\mathbb{P}_x$-martingale with decomposition*

$$dH_t^m = \beta h_m'(V_{t-}) \mathbf{1}_{\{V_{t-} > 0\}}\, dM_t, \tag{16}$$

where M_t is the $\mathbb{P}_x$-martingale $dM_t = dN_t - d\Lambda_t$ and $dN_t^{m,V} = \mathbf{1}_{[0,m]}(V_{t-})\, dN_t$.

(b) *Let τ_m^V be the* CUSUM *stopping time defined in Section 2 as $\tau_m^V = \inf\{t \ge 0, V_t \ge m\}$. Then*

$$h_m(y) = \mathbb{E}_x(N_{\tau_m^V}). \tag{17}$$

(ii) *Let $\tilde{h}_m$ be a continuous regular solution of the* DDE($\tilde{\beta}$).

(a) *$\tilde{h}_t^m = \tilde{h}_m(Y_t) + N_t^{m,V} + \delta \tilde{h}_m(m) J_t^{m,V}$ is a $\tilde{\mathbb{P}}_x$-martingale and $d\tilde{H}_t^m = -\tilde{\beta} \tilde{h}_m'(V_{t-})\, d\tilde{M}_t$.*

(b) *Put $\psi(x) = \rho^x(\tilde{h}_m(0) - \tilde{h}_m(y))$. The $\mathbb{P}_x$-martingale, $\bar{H}_t^{m,\tilde{\rho}} = -\beta \times \int_0^t \phi'(Y_{s-})\, dM_s$ is equal to the process $\bar{H}_t^{m,\tilde{\rho}} = \rho^{V_t}(\tilde{h}_m(0) - \tilde{h}_m(V_t)) - \rho \int_0^t \rho^{V_{s-}}\, dN_s^{m,V} - \delta \tilde{h}_m(m) \rho^m J_t^{m,V}$.*

PROOF. (i)(a) The proof of (16) is an easy consequence of the differential formula (8) applied to $h_m(V_t)$ whose jumps are $(h_m(V_{t-} + 1) - h_m(V_{t-})\, dN_t + \delta h_m(m)\, dJ_t^{m,V})$.

(b) The Neumann DDE (12) expresses that the derivative of h_m at 0 must be equal to 0 to compensate for the fact that the continuous drift of V does not charge 0. Solving the Neumann DDEs is postponed to the next section.

(ii)(a) The same relationship is true for $\tilde{h}_m$ with respect to the probability measure $\tilde{\mathbb{P}}_x$, with the coefficients $[\tilde{\rho} = 1/\rho, \tilde{\beta} = \beta(1/\rho) = \beta(\rho)/\rho]$, the martingale $\tilde{M}_t = N_t - \rho \Lambda_t$, and the Neumann DDE($\tilde{\beta}$) defined on $[0, m)$ by $\tilde{\beta} \tilde{h}_m'(x) = \tilde{h}_m(x+1) - \tilde{h}_m(x) + 1, \tilde{h}_m'(0) = 0$.

(b) However, we are also interested in the process $\psi(V_t)$, where $\psi(x) = \rho^x(\tilde{h}_m(0) - \tilde{h}_m(x))$, as a way of studying the properties of $\tilde{h}_m(V_t)$ under $\mathbb{P}$ in terms of properties under $\tilde{\mathbb{P}}$. Set $\tilde{k}_m(x) = \tilde{h}_m(0) - \tilde{h}_m(x)$ and $\psi(x) = \rho^x \tilde{k}_m(x)$. Then, using similar arguments as in the proof of Proposition 3, we first write $\beta \psi'(x) = (\rho - 1)\psi(x) + \rho \rho^x \tilde{\beta} \tilde{k}_m'(x) = (\rho - 1)\psi(x) + \rho \rho^x (\tilde{k}_m(x+1) - \tilde{k}_m(x) - 1)$, which

can be expressed by using only ψ as $\beta\psi'(x) = \psi(x+1) - \psi(x) - \rho\rho^x$. Moreover, the jump of $\psi(x)$ at m is $\delta\psi(m) = \rho^m \tilde{h}_m(m-)$. Therefore, the differential of $\psi(V_t)$ is

$$\begin{aligned} d\psi(V_t) &= \big(\psi(V_t+1) - \psi(V_{t-})\big)\, dN_t - \beta\psi'(V_t)\, d\Lambda_t + \rho^m \tilde{h}_m(m-)\, dJ_t^{d,m} \\ &= \big(\psi(V_t+1) - \psi(V_{t-}) - \beta\psi'(V_{t-})\big)\, dN_t + \beta\psi'(V_{t-})\, dM_t \\ &\quad + \rho^m \tilde{h}_m(m-)\, dJ_t^{d,m}. \end{aligned}$$

By the previous calculation on $\beta\psi'(x)$, the coefficient of dN_t may be simplified as $\rho\rho^{V_{t-}} = \rho^{V_t}$ and $d\psi(V_t) = \rho\rho^{V_{t-}}\, dN_t + \rho^m \tilde{h}_m(m-)\, dJ_t^{m,d} + \beta\psi'(V_{t-})\, dM_t$, which completes the proof. $\square$

(ii) *Comparison of performance criteria.* All of these properties allow us to be able to compare the criteria studied under the different probability measures.

COROLLARY 4.

(a) *On* $(0, m)$, *if* $\rho < 1$, $\tilde{h}_m(x) \geq h_m(x)$, *and if* $\rho > 1$, $\tilde{h}_m(x) \leq h_m(x)$.
(b) *On* $(0, m)$, *if* $\rho < 1$, $\tilde{g}_m(y) \leq g_m(y)$, *and if* $\rho > 1$, $\tilde{g}_m(y) \geq g_m(y)$.

PROOF. (i) 0n $[0, \tau_m^V]$, the process $J^{m,V}$ is the null process since V is below m until τ_m^V. Then $\widetilde{H}_t^m = \tilde{h}_m(V_t) - \tilde{h}_m(x) + N_t^{m,V} = \int_0^t \tilde{\beta}\tilde{h}'_m(V_{s-})\, d\widetilde{M}_s = \int_0^t \tilde{\beta}\tilde{h}'_m(V_{s-})\, dM_s - (\rho - 1)\, d\Lambda_s)$, is a $\widetilde{\mathbb{P}}_x$-martingale; since $\tilde{h}_m$ is decreasing and since the sign of $\tilde{\beta}(\rho - 1) = \rho \ln \rho$ is the same as the sign of $\rho - 1$, $\widetilde{H}_t^m$ is a $\mathbb{P}_x$-submartingale if $\rho > 1$, and a $\mathbb{P}_x$-supermartingale if $\rho < 1$.

(ii) Then we have that $\widetilde{\mathbb{E}}_x[N_{\tau_m^V}] - \tilde{h}_m(x) = 0$. Moreover, $\mathbb{E}_x[N_{\tau_m^V}] - \tilde{h}_m(x)$ is positive or negative depending on the sign of $\rho - 1$. For the optimality application, the more important inequality is $\tilde{h}_m(x) \leq h_m(x)$ when $\rho > 1$.

(iii) The proof is the same for the inequalities concerning the functions $\tilde{g}_m(y)$ and $g_m(y)$. $\square$

4. Solving and interpreting the DDEs. In this section, we show that the solutions of the above-mentioned DDEs can be obtained in closed form from the well-known solution u of the fundamental $\mathrm{DDE_F}(\beta)$ without drift, $\beta u'(x) = u(x) - u(x-1)$, defined on $\mathbb{R}^+$, and extended by the value zero for $x < 0$.

4.1. *The fundamental delay equation and the Pollaczek–Khintchine formula.* Let us consider the fundamental delayed equation $\mathrm{DDE_F}(\beta)$ on $[0, \infty)$:

$$\beta u'(x) = u(x) - u(x-1), \qquad \beta > 0, \tag{18}$$

where only the continuous regular finite variation functions solutions u on $(0, \infty)$ are considered; $u(0)$ is defined as $u(0) = \lim_{\varepsilon \mapsto 0} u(\varepsilon)$. The standard extension to

the whore real line is given by putting $u(x) = 0, \forall x < 0$. Then a jump at 0 equal to $\delta u(0) = u(0) - u(0-) = u(0)$ is introduced when working on the real line.

The solution with initial condition $\beta u(0) = 1$ is called the *scale function* in Lévy processes theory and denoted W in [12]. We shall come back to this interpretation at the end of the section.

(i) The right-continuous function u' has two jumps. The first is at $x = 1$ with size $-u(0)/\beta$, since $\beta u'(1) = u(1) - u(0)$ and $\beta u'(1-) = u(1)$. The second is at $x = 0$ with size $u(0)/\beta$, since $\beta u'(0) = u(0)$ and $\beta u'(0-) = 0$. On the interval $[0, 1)$, $u(x)$ is an exponential function, that is, $u(0)e^{x/\beta}$.

(ii) In the case $\beta > 1$, the solution corresponds to the well-known Pollaczek–Khintchine formula [1], page 281. One can interpret u as the c.d.f. of the sum S_ν of i.i.d. r.v. uniformly distributed on $[0, 1]$ stopped at an independent geometric r.v. ν. Note that $\mathbb{P}(S_\nu = 0) = \mathbb{P}(\nu = 0) = 1 - 1/\beta$. When $\beta < 1$, there is no such easy interpretation.

(iii) However, our aim is to examine a universal representation under the probability measures $\mathbb{P}$ and $\widetilde{\mathbb{P}}$, which does not distinguish between cases $\beta > 1$ and $\beta < 1$. Natural tools are the so-called scale functions W (mentioned above) and $\widetilde{W}$. They are (resp.) the solutions of $\mathrm{DDE_F}(\beta)$ and $\mathrm{DDE_F}(\tilde{\beta})$, whose values at 0 are respectively $W(0) = \frac{1}{\beta(\rho)}$ and $\widetilde{W}(0) = \frac{1}{\tilde{\beta}(\rho)} = \frac{\rho}{\beta(\rho)} = \rho W(0)$. The advantage of choosing these functions comes from their alternate definition as functions whose ($\mathbb{P}$, resp., $\widetilde{\mathbb{P}}$)-Laplace transform is the inverse of the ($\mathbb{P}$, resp., $\widetilde{\mathbb{P}}$)-Laplace exponent of the Lévy process U^ρ in the Poisson case [6, 12].

4.2. *Computation of performance functions from scale functions.*

Conventional notation for primitive functions. The notation $\mathrm{F}^\diamond$ refers to the primitive of a function F, null at 0, $\mathrm{F}^\diamond(z) = \int_0^z \mathrm{F}(z)\,dz$.

The fundamental $\mathrm{DDE_F}(\beta)$ given in (18) and studied in the previous section, yields a large family of useful new DDEs by simple transformation. We adopt the same kind of notation $\mathrm{DDE}(\beta, k)$, for a $\mathrm{DDE}(\beta)$ with an additional constant term k in the right-hand side of the equation, so that $\mathrm{DDE_F}(\beta) = \mathrm{DDE}(\beta, 0)$.

THEOREM 5. *The* $\mathrm{DDE_F}(\beta)$ *defined by* $\beta u'(x) = u(x) - u(x-1)$, *is stable under certain simple transformations.*

(i) (a) *If u is solution of the* $\mathrm{DDE_F}(\beta)$, *with* $\beta = \beta(\rho)$, *then* $\rho^x u(x)$ *is solution of* $\mathrm{DDE_F}(\tilde{\beta})$ *with* $\tilde{\beta}(\rho) = \beta(\tilde{\rho}) := \beta(1/\rho) = \beta(\rho)/\rho$.

(b) *The function* $\widetilde{W}(x) = \rho\rho^x W(x)$ *is the scale function under* $\widetilde{\mathbb{P}}$ ($\widetilde{W}(0) = 1/\tilde{\beta}$), *solution of the* $\mathrm{DDE_F}(\tilde{\beta})$.

(ii) *If u is solution of* $\mathrm{DDE_F}(\beta)$, *then the primitive* $u^\diamond(x) = \int_0^x u(z)\,dz$, *equal to* 0 *for* $x \le 0$, *is solution of the* $\mathrm{DDE}(\beta, k)$ $\beta u^{\diamond,\prime}(x) = u^\diamond(x) - u^\diamond(x-1) + k$, *with* $k = \beta u(0)$, *satisfying the derivative constraint at* 0, $\beta u^{\diamond,\prime}(0) = \beta u(0) = k$. *Any other derivative constraint at* 0 *may be satisfied by the addition of a well-chosen function* $\theta u(x)$.

(iii) *Same results hold for the equation* $\mathrm{DDE_F}(\tilde{\beta})$ *and its elementary transformation as in part* (ii), *by using the scale function* $\widetilde{W}(x) = \rho\rho^x W(x)$.

PROOF. (i) The derivative of $\rho^x u(x) = \tilde{u}(x)$ is $\tilde{u}'(x) = \log(\rho)\rho^x u(x) + \rho^x u'(x)$. The relationship $\log(\rho) = \beta(\rho)(\rho - 1)$, and some algebra gives the required result:

$$\big(\beta(\rho)/\rho\big)\tilde{u}'(x) = \beta(\tilde{\rho})\tilde{u}'(x) = \tilde{u}(x) - \tilde{u}(x-1).$$

(ii) The integral equation is based on formula (8) since $\int_0^x u'(z)\,dz = u(x) - u(0)$. Then the DDE of the primitive function $u^\diamond$ differs from the previous one by the addition of the constant $u(0)\beta$.

(iii) Any linear combination of the scale function and its primitive is a solution of $\mathrm{DDE}(\beta, k)$ whose parameters have an easy interpretation, the drift term $k = \hat{\theta}$ and an additional constraint on the value at 0, or on the derivative at 0 or at another level, determines the value of θ. $\square$

As a consequence, we can derive closed-form formulas for the performance functions $g_m(y)$ and $h_m(x)$, solutions of the $\mathrm{DDE}(\beta, -1)$ with different constraints. Indeed, $g_m(y)$ is solution of the $\mathrm{DDE}(\beta, -1)$ with terminal condition $g_m(m) = 0$. So $g_m(y)$ is a primitive null at m of $-W(x)$. On the other hand, the transformed performance $\mathsf{k}_m(z) = h_m(m - z)$ is a Neumann's solution of the $\mathrm{DDE}(\beta, -1)$, with the Neumann constraint $\mathsf{k}_m'(m-) = 0$. Thus, thanks to Theorem 5, these functions are linear combinations of the scale functions and their primitive (without restriction on the value of the parameter β):

$$g_m(y) = \int_y^m W(z)\,dz, \qquad \text{with } g_m(0) = \int_0^m W(z)\,dz, \tag{19}$$

$$h_m(x) = W(m-x)\frac{W(m)}{W'(m)} - \int_0^{m-x} W(y)\,dy, \qquad h_m(m-) = \frac{W(m)}{\beta W'(m)}, \tag{20}$$

$$\tilde{g}_m(y) = \rho\int_y^m \rho^z W(z)\,dz, \tag{21}$$

$$\tilde{h}_m(x) = \widetilde{W}(m-x)\frac{\widetilde{W}(m)}{\widetilde{W}'(m)} - \int_0^{m-x} \widetilde{W}(z)\,dz, \qquad \tilde{h}_m(m-) = \frac{\widetilde{W}(m)}{\tilde{\beta}\widetilde{W}'(m)}. \tag{22}$$

The results can be deduced indirectly from similar formulas from Pistorius (2004) [22] or Kyprianou (2014) [12]. Nevertheless, in the Poisson case, our proofs are very elementary and do not require excursion theory as in [12, 22], only elementary differential calculus.

5. Optimality of the CUSUM stopping rule. Since the CUSUM process depends on the value of ρ, we reintroduce the distinction between $\rho > 1$ associated

with the process V and the performance functions h_m and $\tilde{h}_m$, and $\rho < 1$ associated with the process Y and the performance functions g_m and $\tilde{g}_m$. To solve the optimality problem with false alarm constraint $\mathbb{E}(N_T) \geq \pi$, we can limit ourselves to stopping times that satisfy the false alarm constraint with equality. Therefore, the aim is to prove the minimal optimality of the CUSUM stopping times τ_m^V and τ_m^Y with false alarm constraint $\mathbb{E}(N_{\tau_m^V}) = h_m(0)$ [resp., $\mathbb{E}(N_{\tau_m^V}) = g_m(0)$], that is, to show that for any stopping time T satisfying the false alarm constraint, the criterion

$$\widetilde{C}(T) = \sup_{\theta \in [0,\infty]} \operatorname*{ess\,sup}_{\omega} \tilde{\Gamma}_\theta^T, \qquad \text{with } \tilde{\Gamma}_\theta^T = \widetilde{\mathbb{E}}\big((N_T - N_\theta)^+ | \mathcal{F}_\theta\big) \tag{23}$$

is lower-bounded by $\tilde{h}_m(0)$ when $\rho > 1$ and by $\tilde{g}_m(0)$ when $\rho < 1$.

5.1. *Verification of the false alarm constraint for the* CUSUM *strategies*. The first step is to verify that the previous inequalities are satisfied by the stopping times τ_m^V or τ_m^Y. This is easily done by the martingale properties established in Propositions 2 and 3 and the fact that the functions $\tilde{h}_m$ and $\tilde{g}_m$ are decreasing. The main consequence is that the conditional criteria are deterministic functions of the reflected processes for $t \leq \tau_m^V$ or $t \leq \tau_m^Y$,

$$\tilde{\Gamma}_t^{\tau_m^V} = \widetilde{\mathbb{E}}\left(\int_t^{\tau_m^V} dN_s \middle| \mathcal{F}_t\right) = \tilde{h}_m(V_t) \leq \tilde{h}_m(0), \quad \text{and}$$

$$\tilde{\Gamma}_t^{\tau_m^Y} = \widetilde{\mathbb{E}}\left(\int_t^{\tau_m^Y} dN_s \middle| \mathcal{F}_t\right) = \tilde{g}_m(Y_t) \leq \tilde{g}_m(0).$$

The ess sup in t of the two criteria is attained at $t = 0$ (worst case) and equal to $\tilde{h}_m(0)$ [resp., $\tilde{g}_m(0)$]. That is, starting from 0,

$$\operatorname*{ess\,sup}_t \tilde{\Gamma}_t^{\tau_m^V} = \tilde{h}_m(0), \quad \text{and} \quad \operatorname*{ess\,sup}_t \tilde{\Gamma}_t^{\tau_m^Y} = \tilde{g}_m(0). \tag{24}$$

Here, $\tilde{h}_m(0)$ or $\tilde{g}_m(0)$ are indeed the CUSUM bounds of these strategies with false alarm constraints $h_m(0)$ or $g_m(0)$.

5.2. *Modification of the* CUSUM *criterion*. Showing the optimality of the CUSUM strategies is a more difficult task. The first step is to define a new criterion allowing us to work under the same probability measure $\mathbb{P}$. This idea is in the same vein as Shiryaev [26] and Moustakides [17]. But given that our criterion is based on an integration with respect to the pure jump process dN_t in place of the Lebesgue measure as in the cited works, we have to justify the form of the new criteria. A useful tool is a formula for integration by parts under the probability $\widetilde{\mathbb{P}}$.

PROPOSITION 6. *Consider the conditional criterion* $\widetilde{\Gamma}_t^T := \widetilde{\mathbb{E}}(\int_t^T dN_s | \mathcal{F}_t)$ *assumed to be bounded by a constant* $\widetilde{C}(T)$ *as in the Lorden criterion* (5). *The fol-*

lowing lower-bounds hold:

$$\text{for } \rho > 1, \qquad \rho\mathbb{E}\left[\int_t^T \rho^{V_{s-}}\, dN_s \middle| \mathcal{F}_t\right] \le \widetilde{C}(T)\mathbb{E}\big(\rho^{V_T}|\mathcal{F}_t\big),$$

$$\text{for } \rho < 1, \qquad \rho\mathbb{E}\left[\int_t^T \tilde{\rho}^{Y_{s-}}\, dN_s \middle| \mathcal{F}_t\right] \le \widetilde{C}(T)\mathbb{E}\big(\tilde{\rho}^{Y_T}|\mathcal{F}_t\big).$$

PROOF. We start with calculations under the probability $\widetilde{\mathbb{P}}$ and take the primitive of the conditional criterion $\widetilde{\Gamma}_t^T = \widetilde{\mathbb{E}}(\int_t^T dN_u|\mathcal{F}_t)$ with respect to the nondecreasing process $\rho^{\bar{Z}}$, equal to $\rho^{\bar{X}}$ if $\rho > 1$ or $\rho^{-\bar{U}} = \tilde{\rho}^{\bar{U}}$ if $\rho < 1$. The first one is continuous, the second one is the exponential of a pure jump process. In what follows, we use the notation $\rho^{\bar{Z}}$ to refer to the nondecreasing power to one of these max processes. Then $\widetilde{\mathbb{E}}[\int_t^T \widetilde{\Gamma}_\alpha^T\, d\rho^{\bar{Z}_\alpha}|\mathcal{F}_t] = \widetilde{\mathbb{E}}[\int_t^T (\int_{(\alpha,T]} dN_u)\, d\rho^{\bar{Z}_\alpha}|\mathcal{F}_t]$, and by integration by parts, $\widetilde{\mathbb{E}}[\int_t^T \widetilde{\Gamma}_\alpha^T\, d\rho^{\bar{Z}_\alpha}|\mathcal{F}_t] = \widetilde{\mathbb{E}}[\int_t^T dN_u(\rho^{\bar{Z}_{u-}} - \rho^{\bar{Z}_t})|\mathcal{F}_t]$.

Using the fact that $\widetilde{C}(T)$ is an upper bound for $\widetilde{\Gamma}_t^T$, we obtain the series of inequalities:

$$\begin{aligned}\widetilde{C}(T)\widetilde{\mathbb{E}}[\rho^{\bar{Z}_T}|\mathcal{F}_t] &\ge \widetilde{\mathbb{E}}\left[\int_t^T dN_u \rho^{\bar{Z}_{u-}} \middle| \mathcal{F}_t\right] + \left(\widetilde{C}(T) - \widetilde{\mathbb{E}}\left[\int_t^T dN_u \middle| \mathcal{F}_t\right]\right)\rho^{\bar{Z}_t} \\ &\ge \widetilde{\mathbb{E}}\left[\int_t^T dN_u \rho^{\bar{Z}_{u-}} \middle| \mathcal{F}_t\right].\end{aligned}$$

Then, since ρ_t^U is the martingale density of $\widetilde{\mathbb{P}}$ with respect to $\mathbb{P}$, we can now make the comparison under the probability measure $\mathbb{P}$, $\widetilde{C}(T)\mathbb{E}[\rho^{U_T+\bar{Z}_T}|\mathcal{F}_t] \ge \mathbb{E}[\int_t^T dN_u \rho^{U_u+\bar{Z}_{u-}}|\mathcal{F}_t]$.

We can be more explicit about these inequalities since the process $U + \bar{Z}$ is one of the reflected processes V or $-Y$. First, using the process V and the continuity of $\bar{X}$, the inequality becomes

$$\widetilde{C}(T)\mathbb{E}\big[\rho^{V_T}|\mathcal{F}_t\big] \ge \mathbb{E}\left[\int_t^T dN_u \rho^{V_u} \middle| \mathcal{F}_t\right] = \rho\mathbb{E}\left[\int_t^T dN_u \rho^{V_{u-}} \middle| \mathcal{F}_t\right].$$

For the process $Y = \bar{U} - U$, we have to be careful with respect to the noncontinuity of $\bar{U}$, so that $U_t + \bar{Z}_{t-} = U_t - \bar{U}_{t-} = -Y_{t-} + U_t - U_{t-}$. Then

$$\widetilde{C}(T)\mathbb{E}\big[\rho^{-Y_T}|\mathcal{F}_t\big] \ge \mathbb{E}\left[\int_t^T dN_u \rho^{1-Y_{u-}} \middle| \mathcal{F}_t\right] = \rho\mathbb{E}\left[\int_t^T dN_u \tilde{\rho}^{Y_{u-}} \middle| \mathcal{F}_t\right].$$

This last formulation shows the symmetry found between the cases ($\rho > 1$) and ($\rho < 1$). $\square$

5.3. *Optimality for a decline in intensity.* Let us reconsider the optimality of the CUSUM stopping time τ_m in the class of stopping times with finite CUSUM performance, satisfying the false alarm constraint, that is, $\mathbb{E}(N_{\tau_m}) \le \mathbb{E}(N_T) < +\infty$.

Given that the criterion is dependent on the jump process N, we have to consider separately the cases $\rho < 1$ and $\rho > 1$. In the case of a decline in intensity, that is, $\rho < 1$, it is sufficient to prove that, for a process Y starting from 0,

$$\rho\mathbb{E}\left(\int_0^T \tilde{\rho}^{Y_{s-}}\,dN_s\right) \geq \tilde{g}_m(0)\mathbb{E}\big(\tilde{\rho}^{Y_T}\big), \qquad \text{given } \mathbb{E}(N_T) = \mathbb{E}(N_{\tau_m^Y}) = g_m(0). \tag{25}$$

Propositions 2 and 3 provide useful tools to reformulate these inequalities in a more tractable form, based on the processes $N_t^{m,Y} = \int_0^t \mathbf{1}_{[0,m)}(Y_{s-})\,dN_s$ and $N_t^{m,Y,\tilde{\rho}} = \int_0^t \tilde{\rho}^{Y_{s-}}\,dN_s^m$. Using the continuity of functions g_m and $\tilde{g}_m$, and the martingale properties established in Propositions 2 and 3, we know that $\mathbb{E}(N_T) = \mathbb{E}(g_m(Y_T) + N_T^{m,Y}) = g_m(0)$ and $\mathbb{E}(\rho N_t^{m,Y,\tilde{\rho}} + \tilde{\rho}^{Y_t}(\tilde{g}_m(0) - \tilde{g}_m(Y_t))) = 0$. So, as in Poor and Hadjiliadis (2009) [24], page 157, the inequality (25) is equivalent to

$$\begin{aligned} &\mathbb{E}\left(\int_0^T \rho\tilde{\rho}^{Y_{s-}}\mathbf{1}_{[m,\infty)}(Y_{s-})\,dN_s - \tilde{\rho}^{Y_T}\tilde{g}_m(Y_T)\right) \geq 0, \\ &\qquad\qquad \text{given } \mathbb{E}\left(\int_0^T \mathbf{1}_{[m,\infty)}(Y_{s-})\,dN_s - g_m(Y_T)\right) = 0. \end{aligned} \tag{26}$$

Thus, the optimality of the CUSUM rule is easy to prove.

THEOREM 7 (Optimality result for a decline in intensity). *Assume $\rho < 1$, and consider the* CUSUM *process Y. Let T be a stopping time with finite* CUSUM *performance, and false alarm constraint $\mathbb{E}(N_{\tau_m^Y}) = \mathbb{E}(N_T) < +\infty$:*

(i) *$\tilde{g}_m(0)$ is a lower bound for the criterion*

$$\rho\mathbb{E}\left(\int_0^T \tilde{\rho}^{Y_{s-}}\,dN_s\right) \geq \tilde{g}_m(0)\mathbb{E}\big(\tilde{\rho}^{Y_T}\big), \qquad \textit{given } \mathbb{E}(N_T) = \mathbb{E}(N_{\tau_m^Y}) = g_m(0). \tag{27}$$

(ii) *τ_m^Y is an optimal detection rule for the problem (7) under the false alarm constraint $\mathbb{E}(N_T) = \mathbb{E}(N_{\tau_m^Y}) = g_m(0)$.*

PROOF. (i) The CUSUM process is the process Y, with CUSUM stopping time τ_m^Y (τ_m in short). We have seen that the inequality (27) is equivalent to the inequality (26). Since $\mathbb{E}(\int_0^T \mathbf{1}_{(m,\infty)}(Y_{s-})\,dN_s - g_m(Y_T)) = 0$, the idea is to control the term $\rho\tilde{\rho}^{Y_{s-}}\mathbf{1}_{[m,\infty)}(Y_{s-})$ by $\tilde{\rho}^m\mathbf{1}_{[m,\infty)}(Y_{s-})$, since the difference is still nonnegative ($\tilde{\rho} > 1$).

Thus, the inequality (26) will be proved if we show that $\rho\tilde{\rho}^m g_m(Y_T) - \tilde{\rho}^{Y_T}\tilde{g}_m(Y_T)$ is nonnegative in expectation. Let us study the function $\psi(y) = \tilde{\rho}^{m-y}g_m(y) - \tilde{\rho}\tilde{g}_m(y)$, equal to 0 when $y \geq m$, using the description of the functions g_m and $\tilde{g}_m$ given in equations (19)–(22) in terms of scale functions. Recall that $g_m(y) = \int_y^m W(z)\,dz$, $\tilde{g}_m(y) = \int_y^m \widetilde{W}(z)\,dz$ and $\widetilde{W}(z) = \rho\rho^x W(x)$.

The derivative $\psi'(y)$ is negative since $A(y) = -\log(\tilde{\rho})\tilde{\rho}^{(m-y)}g_m(y)$ is negative, $\tilde{\rho}^m > 1$, and

$$\begin{aligned}\psi'(y) - A(y) &= -\tilde{\rho}^{(m-y)}W(y) + \tilde{\rho}\widetilde{W}(y) = -\tilde{\rho}^{(m-y)}W(y) + \tilde{\rho}^{-y}W(y)\\ &= \tilde{\rho}^{-y}W(y)\big(1-\tilde{\rho}^m\big).\end{aligned}$$

Therefore, $\psi'(y)$ is negative and $\psi(y)$ is decreasing, null at $y = m$ and then non-negative. Thus, the lower bound is proved.

(ii) We have seen in Proposition 6 that the lower bound is an equality for τ_m^Y. Thus, optimality is proved. □

5.4. *Resolution of the conjecture on* CUSUM *rule optimality for a rising intensity.* In the case of rising intensity ($\rho > 1$), it is sufficient to prove that, for V starting from 0,

$$\mathbb{E}\left(\int_0^T \rho^{V_s}\,dN_s\right) \geq \tilde{h}_m(0)\mathbb{E}\big(\rho^{V_T}\big), \qquad \text{when } \mathbb{E}(N_T) = \mathbb{E}(N_{\tau_m^V}) = h_m(0). \tag{28}$$

Unlike the case $\rho < 1$, the functions h_m and $\tilde{h}_m$ are not continuous at $x = m$. Therefore, from Propositions 2 and 3, one must take into account additional terms as follows:

$$\begin{cases}\mathbb{E}(N_T) = h_m(0) = \mathbb{E}\big(N_T^m + h_m(V_T) - h_m(m-)J_T^{d,m}\big),\\ \mathbb{E}\left(\int_0^T \rho^{V_s}\,dN_s^{m,V} - \tilde{h}_m(0)\rho^{V_T}\right) = \mathbb{E}\big(\tilde{h}_m(m-)\rho^m J_T^{d,m} - \rho^{V_T}\tilde{h}_m(V_T)\big).\end{cases}$$

Thus, we have to show that, for $\rho > 1$,

$$\begin{cases}\mathbb{E}\left(\int_0^T \rho^{V_s}\mathbf{1}_{[m,\infty)}(V_{s-})\,dN_s + \tilde{h}_m(m-)\rho^m J_T^{d,m} - \rho^{V_T}\tilde{h}_m(V_T)\right) \geq 0,\\ \qquad \text{when } \mathbb{E}\left(\int_0^T \mathbf{1}_{(m,\infty)}(V_{s-})\,dN_s + h_m(m-)J_T^{d,m} - h_m(V_T)\right) = 0.\end{cases} \tag{29}$$

These equations differ from equation (26) by the terms in $J^{d,m}$ on which we have little information. The idea is to drop these additional terms by using an appropriate linear combination of these two equations, and to reduce the problem to showing that for $\rho > 1$:

$$\begin{aligned}\mathbb{E}\Bigg(\int_0^T &\left(\rho^{V_s} - \frac{\tilde{h}_m(m-)}{h_m(m-)}\rho^m\right)\mathbf{1}_{[m,\infty)}(V_{s-})\,dN_s\\ &+ \frac{\tilde{h}_m(m-)}{h_m(m-)}\rho^m h_m(V_T) - \rho^{V_T}\tilde{h}_m(V_T)\Bigg) \geq 0.\end{aligned} \tag{30}$$

Condition (30) is the corrected version of equation (6.135) in Poor and Hadjiliadis (2009) [23], page 157, in which the additional jump term has not been taken into account. The fact that the correction factor $\omega_m = \tilde{h}_m(m-)/h_m(m-)$ is smaller than 1 (see Corollary 4) makes it more complicated to obtain the desired inequality. As a by-product, we solve the conjecture concerning the Poisson case.

THEOREM 8 (Optimality result for a proportional rise in stochastic intensity). *Assume $\rho > 1$, and consider the* CUSUM *process V. Let T be a stopping time with finite* CUSUM *performance, and false alarm constraint $\mathbb{E}(N_{\tau_m^V}) = \mathbb{E}(N_T) = h_m(0)$:*

(i) *$\tilde{h}_m(0)$ is a lower-bound for the criterion*

$$\mathbb{E}\left(\int_0^T \rho^{V_s}\, dN_s\right) \geq \tilde{h}_m(0)\mathbb{E}\big(\rho^{V_T}\big), \qquad \text{given } \mathbb{E}(N_T) = \mathbb{E}(N_{\tau_m^V}) = h_m(0). \tag{31}$$

(ii) *τ_m^V is an optimal detection time for the false alarm constraint $\mathbb{E}(N_T) = h_m(0)$.*

PROOF. For simplicity, we set $\omega_m = \tilde{h}_m(m-)/h_m(m-)$. Then we have to show that

$$\mathbb{E}\left(\int_0^T (\rho^{V_s} - \omega_m\rho^m)\mathbf{1}_{[m,\infty)}(V_{s-})\, dN_s + \omega_m\rho^m h_m(V_T) - \rho^{V_T}\tilde{h}_m(V_T)\right) \geq 0. \tag{32}$$

(i) The first term of the inequality is nonnegative, since from Corollary 4, $\omega_m \leq 1$. Recall that $\mathsf{k}_m(x) = h_m(m-x)$ and $\tilde{\mathsf{k}}_m(x) = \tilde{h}_m(m-x)$. The sign of second term, null when $V_T \geq m$, and rewritten as $\rho^{V_T}(\omega_m\rho^{m-V_T}\mathsf{k}_m(m-V_T) - \tilde{\mathsf{k}}_m(m-V_T))$, depends on the properties of the continuous function ϕ_m defined as $\phi_m(z) = \omega_m\rho^z\mathsf{k}_m(z) - \tilde{\mathsf{k}}_m(z)$. The proof relies on the identity $\tilde{\mathsf{k}}_m'(x) = \omega_m\rho^x\mathsf{k}_m'(x)$, that we shall now prove.

(a) We start by showing that $\omega_m = \tilde{h}_m(m-)/h_m(m-) = \rho - (\rho-1)\tilde{h}_m(m-)$. This property is a simple consequence of the sequence of equalities

$$\tilde{\beta} = \beta/\rho, \qquad \tilde{\beta}\tilde{h}_m(m-) = \frac{\widetilde{W}(m)}{\widetilde{W}'(m)}, \qquad \widetilde{W}(m) = \rho\rho^m W(m).$$

Next, using the logarithmic derivative of $\widetilde{W}(x)$, we obtain

$$\frac{1}{\tilde{\beta}\tilde{h}_m(m-)} = \frac{\widetilde{W}'(m)}{\widetilde{W}(m)} = \ln\rho + \frac{W'(m)}{W(m)} = \ln\rho + \frac{1}{\beta h_m(m-)} = \frac{\rho-1}{\beta} + \frac{1}{\beta h_m(m-)}.$$

(b) There is no difficulty in studying the derivative of $\mathsf{k}_m(x) = \beta h_m(m-) \times W(x) - W^\diamond(x)$ and to show that $W(x) + \mathsf{k}_m'(x) = \beta h_m(m-)W'(x)$. Similarly, the same relationship holds true for the *tilded* functions and their derivatives $\widetilde{W}(x) + \tilde{\mathsf{k}}_m'(x) = \tilde{\beta}\tilde{h}_m(m-)\widetilde{W}'(x)$. The link with h_m is deduced from the relationship $\widetilde{W}(x) = \rho\rho^x W(x)$ as follows:

$$\begin{aligned}
\tilde{\mathsf{k}}_m'(x) &= \beta\tilde{h}_m(m-)\rho^x W'(x) + \rho^x W(x)\big(\beta\log(\rho)\tilde{h}_m(m-) - \rho\big)\\
&= \omega_m\rho^x\beta h_m(m-)W'(x) + \rho^x W(x)\big((\rho-1)\tilde{h}_m(m-) - \rho\big)\\
&= \omega_m\rho^x\mathsf{k}_m'(x) + \rho^x W(x)\big(\omega_m + (\rho-1)\tilde{h}_m(m-) - \rho\big) = \omega_m\rho^x\mathsf{k}_m'(x).
\end{aligned}$$

(c) We are now able to study the sign of $\phi_m(x) = \omega_m\rho^x\mathsf{k}_m(x) - \tilde{\mathsf{k}}_m(x)$, since its derivative is

$$\phi_m'(x) = \omega_m\log(\rho)\rho^x\mathsf{k}_m(x) + \rho^x\omega_m\mathsf{k}_m'(x) - \tilde{\mathsf{k}}_m'(x) = \omega_m\log(\rho)\rho^x\mathsf{k}_m(x).$$

Consequently, the function $\phi_m(m-z)$ is positive, decreasing on $[0,m]$ and still convex. The lower bound is established.

(ii) We have seen in Proposition 6 that the lower bound is an equality for τ_m^V. The optimality is proved. $\square$

REMARK 9. These theorems give the infimum of the worst-case mean number of jumps until detection, that is, $\sup_{\theta\in[0,\infty]} \operatorname{ess\,sup}_\omega \widetilde{\mathbb{E}}[(N_T-N_\theta)^+|\mathcal{F}_\theta]$ for a class of stopping times with preassigned false alarm rate $\mathbb{E}(N_T)\geq\pi$. The bounds depend on whether an increase or decrease in intensity is being investigated. As observed by Basseville and Nikiforov [4], this result is important not only for the CUSUM algorithm, but in general, since in some sense, *these lower bounds play the same role in the change detection theory as the Cramer–Rao lower bound in estimation theory*.

6. Numerical illustrations. In this section, we provide an illustration of the performance of the CUSUM procedure. Recall that this only depends on the performance functions $h_m, \tilde{h}_m, g_m$ and $\tilde{g}_m$, given in equations (19)–(22) in terms of the scale functions, and their primitive. Therefore, we can give a closed formula for the scale function and its primitive. For $\beta>1$, recall from Section 4.1 that the scale function is given as $W(x)=\mathbb{P}(S_\nu\leq x)/(\beta-1)$:

(i) The r.v. ν has a geometric distribution with parameter $1/\beta$, that is, $\mathbb{P}(\nu=n)=(1-(1/\beta))\beta^{-n}$, and $S_n=\sum_{i=1}^n U_i$ is the sum of n r.v. U_i being i.i.d. and uniformly distributed on $(0,1)$. The density p_u^{*n} of S_n is known as the Irwin–Hall density [11], equal to 0 when $x\geq j$, and to

$$p_u^{*n}(x)=\frac{1}{(n-1)!}\sum_{k=0}^{\lfloor x\rfloor}(-1)^k\binom{n}{k}(x-k)^{n-1},\qquad \text{when } 0<x<j. \tag{33}$$

Therefore, the r.v. $S_\nu=\sum_{i=0}^\nu U_i$ has a Dirac mass at 0 with probability $1-(1/\beta)$, and a density distribution $p_u^{*\nu}(x)=(1-(1/\beta))\sum_{j=1}^\infty \beta^{-j}p_u^{*j}(x)$.

(ii) Integrating $p_u^{*\nu}$ allows us to derive the cumulative distribution, and thus the scale function W (for $\beta>1$) as follows:

$$W(x)=\frac{1}{\beta}\sum_{k=0}^{\lfloor x\rfloor}\frac{(-1)^k}{k!}\big((x-k)/\beta\big)^k\exp\big((x-k)/\beta\big). \tag{34}$$

Similarly, using elementary calculations and noting that $\frac{(-1)^k}{k!}\int_0^x y^k e^y\,dy = e^x\sum_{j=0}^k\frac{(-x)^j}{j!}-1$, we can derive the closed form of the primitive $\mathrm{W}^\diamond(x)=\int_0^x W(y)\,dy$ as follows:

$$\mathrm{W}^\diamond(x)=\sum_{k=0}^{\lfloor x\rfloor}\left(e^{(x-k)/\beta}\left(\sum_{i=0}^k\frac{(-1)^j}{j!}\big((x-k)/\beta\big)^j\right)-1\right). \tag{35}$$

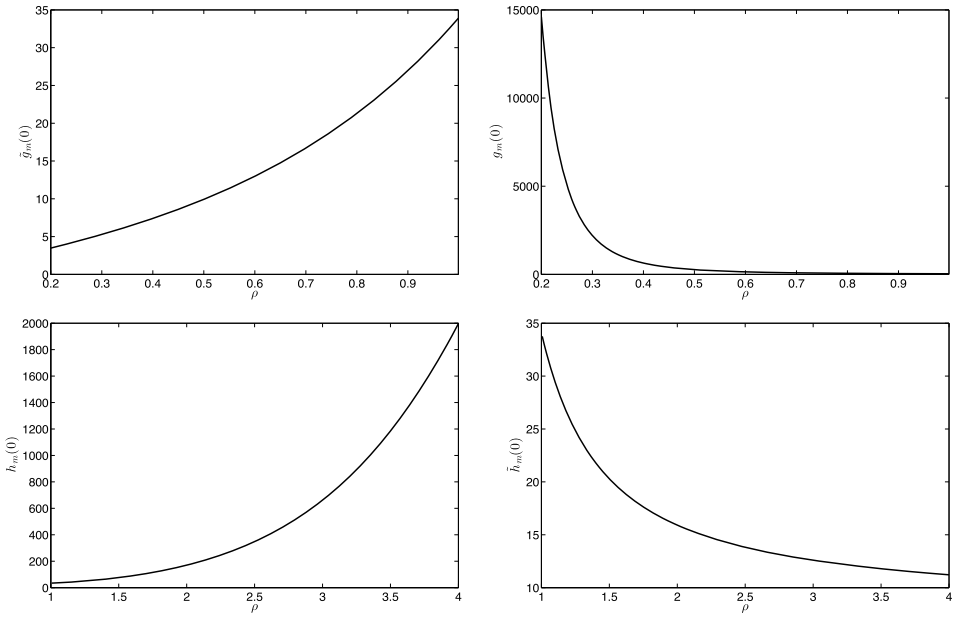

FIG. 3. *Performance of the* CUSUM *as a function of the parameter* ρ *with* $m = 5.5$.

In Figure 3, we depict the scale functions in terms of ρ for $\beta > 1$ and $\beta < 1$. We set the parameters β respectively equal to 1.5 and 0.5. When $\beta < 1$, one should permute the role of W and $\widetilde{W}$ in order to use equations (34) and (35), as W is no longer a cumulative distribution function. Thus, to compute the scale function W, we first compute $\widetilde{W}$ using (35) to characterize $\widetilde{\mathbb{P}}(S_{\tilde{\nu}} \leq x)$ with $\tilde{\beta} = \beta(\rho)/\rho$, and then write $W(x) = \tilde{\rho}\tilde{\rho}^x \widetilde{W}(x)$.

(iii) In Figure 3, for a fixed threshold level $m = 5.5$, we show the performance of the CUSUM procedure for different values of ρ, both for a rise and decline of the intensity. This plot was produced using the series representation of W and $\widetilde{W}$ and the performance functions' closed forms in equations (19)–(22). Since intensities are chosen to be constant (Poisson case), this represents the average delay until detection as well as the false alarm constraint. We observe that the detection is quicker as ρ moves off the critical value 1. Moreover, as noted in Section 5, we can see that the functions h_m and $\tilde{g}_m$ have similar behavior. This is also the case for $\tilde{h}_m$ and g_m. For $\rho > 1$, when either m or ρ becomes too large, some numerical instability may arise in the numerical calculation of $\tilde{h}_m(0)$. This is mainly due to the particular series representation of the scale function W, where we are summing alternating individual terms that increase quickly in absolute value, in the case of $\tilde{h}_m(0)$. This phenomenon has also been observed in sequential hypotheses tests by DeLucia and Poor [7] and in ruin theory by Picard and Lefèvre [21] and Rullière and Loisel [25], where the scale function W also plays a central role. This insta-

bility does not occur for g_m. This is one more difference between the cases $\rho < 1$ and $\rho > 1$.

7. Conclusion. Initially motivated by a problem of quickest detection of a change in certain longevity patterns, we have considered and solved the exact optimality of the minimax robust detection of a disorder time in the rate of doubly stochastic Poisson processes, with a self-contained presentation. In this Lorden-type context, the CUSUM stopping rule is shown to be optimal, both for an increase or decrease in intensity after the change. Given the abundant literature on sequential testing and quickest detection, it may be surprising that this classical problem has not been solved earlier. We believe that this is due to the difference between the increasing and decreasing cases, the former featuring noncontinuous performance functions, and requiring the use of a discontinuous local time. As scale functions appear in the proof, one may wonder if it is possible to extend this result to a broader class of Lévy processes. This is left for further research, along with a detailed analysis of the use of this detection strategy for insurance data sets, and comparison with other detection strategies.

APPENDIX: PROOF OF PROPOSITION 1

(i) *Properties of the process V.*

(a) We have seen that the process $\bar{X}_t^{ad}(Z_0)$ is continuous in time, and only increases when $V_t = 0$. To show (9), we apply Itô's formula to the nonnegative process V_t and the continuous function $\varepsilon \mapsto x^{1+\varepsilon}$, so that $dV_t^{1+\varepsilon} = ((V_{t-}+1)^{1+\varepsilon} - V_{t-}^{1+\varepsilon})\,dN_t - (1+\varepsilon)V_t^{\varepsilon}(\beta\,d\Lambda_t - d\bar{X}_t^{ad})$. The continuous part reduces to $-(1+\varepsilon)\times V_t^{\varepsilon}(\beta d\Lambda_t)$, since $d\bar{X}_t^{ad}$ only increases when $V = 0$. Let ε go to 0, and observe that V_t^{ε} goes to $\mathbf{1}_{\{V_t>0\}}$. The equation then becomes $dV_t = dN_t - \mathbf{1}_{\{V_t>0\}}\beta\,d\Lambda_t$. Comparing this with the definition of V, we obtain that $d\bar{X}_t^{ad} = \beta\mathbf{1}_{\{V_t=0\}}\,d\Lambda_t$.

(b) If V and V' are two nonnegative solutions starting from x, the difference is a differentiable process, and we have $d(V_t - V_t')^+ = \mathbf{1}_{\{V_t>V_t'\}}(\mathbf{1}_{\{V_t>0\}} - \mathbf{1}_{\{V_t'>0\}})(-\beta)\,d\Lambda_t = \mathbf{1}_{\{V_t>0,\,V_t'=0\}}(-\beta)\,d\Lambda_t$. The process $(V_t - V_t')^+$ starting from 0 is nonnegative and nonincreasing, so the null process is also. Inverting the roles of V and V', we obtain that $V_t \equiv V_t'$, and thus uniqueness.

(ii) *Properties of the process Y.*

(a) The previous study of the jumps of $\bar{U}_t^{ad}$ immediately provides the differential form $d\bar{U}_t^{ad} = \mathbf{1}_{\{Y_t=0\}}(1 - Y_{t-})\,dN_t$ in (10). Given the definition of $Y_t(Z_0) = X_t(Z_0) + \bar{U}_t^{ad}(Z_0)$, we only have to justify that the jumps of Y are function of Y_{t-} only, and not of Y_t. Observe that at the time of a jump, $\delta Y_{t-} = -1$ except when in addition $Y_t = Y_{t-} - 1 = 0$. Thus, in terms of the function $j(y) = y \wedge 1$, we have that $j(Y_{t-}) = 1$ if and only if $Y_t \neq 0$, which is the property required to complete the proof of (10).

(b) If Y and Y' are two nonnegative solutions starting from x, the difference $Y_t - Y_t'$ is a pure jump process, starting from 0, such that $d(Y_t - Y_t') = -(j(Y_{t-}) -$

$j(Y'_{t-}))\,dN_t$. Let T_1 be the time of the first jump of N. On the interval $[0, T_1]$, $Y_{t-} - Y'_{t-} = 0$, and $j(Y_t) - j(Y'_t) = 0$. Then, at T_1, the jump of the difference is 0, that is, $Y_{T_1} - Y'_{T_1} = 0$. Using the same argument between successive jump times of N, we see that the difference process $Y_t - Y'_t$ is still equal to 0.

Acknowledgements. The authors acknowledge support from the ANR Project LoLitA (Dynamic Population Models for Human Longevity with Lifestyle Adjustments), the research chair *Actuariat Durable* sponsored by Milliman Paris, the research chair DAMI (Data Analytics and Models for Insurance) sponsored by BNP Paribas Cardif and the research chair *Risques Financiers* sponsored by Société Générale.

REFERENCES

[1] ASMUSSEN, S. (2003). *Applied Probability and Queues*: *Stochastic Modelling and Applied Probability*, 2nd ed. *Applications of Mathematics* (*New York*) **51**. Springer, New York. MR1978607

[2] ASMUSSEN, S. and ALBRECHER, H. (2010). *Ruin Probabilities*, 2nd ed. *Advanced Series on Statistical Science & Applied Probability* **14**. World Scientific, Hackensack, NJ. MR2766220

[3] BARRIEU, P., BENSUSAN, H., EL KAROUI, N., HILLAIRET, C., LOISEL, S., RAVANELLI, C. and SALHI, Y. (2012). Understanding, modelling and managing longevity risk: Key issues and main challenges. *Scand. Actuar. J.* **2012** 203–231. MR2971988

[4] BASSEVILLE, M. and NIKIFOROV, I. V. (1986). *Detection of Abrupt Changes in Signals and Dynamics Systems*. Springer, Berlin.

[5] BEIBEL, M. (1996). A note on Ritov's Bayes approach to the minimax property of the cusum procedure. *Ann. Statist.* **24** 1804–1812. MR1416661

[6] BERTOIN, J. (1998). *Lévy Processes*. *Cambridge Tracts in Mathematics* **121**. Cambridge Univ. Press, Cambridge. MR1406564

[7] DELUCIA, J. and POOR, H. V. (1997). Performance analysis of sequential tests between Poisson processes. *IEEE Trans. Inform. Theory* **43** 221–238. MR1426247

[8] DVORETZKY, A., KIEFER, J. and WOLFOWITZ, J. (1953). Sequential decision problems for processes with continuous time parameter. Testing hypotheses. *Ann. Math. Stat.* **24** 254–264. MR0054911

[9] GRANDELL, J. (1976). *Doubly Stochastic Poisson Processes*. *Lecture Notes in Mathematics* **529**. Springer, Berlin. MR0433591

[10] HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2009). *The Elements of Statistical Learning*: *Data Mining*, *Inference*, *and Prediction*, 2nd ed. Springer, New York. MR2722294

[11] IRWIN, J. O. (1927). On the frequency distribution of the means of samples from a population having any law of frequency with finite moments, with special reference to Pearson's Type II. *Biometrika* 225–239.

[12] KYPRIANOU, A. (2014). *Fluctuations of Lévy Processes with Applications*: *Introductory Lectures*, 2nd ed. Springer, Berlin.

[13] LORDEN, G. (1971). Procedures for reacting to a change in distribution. *Ann. Math. Stat.* **42** 1897–1908. MR0309251

[14] MEI, Y., HAN, S. W. and TSUI, K. L. (2011). Early detection of a change in Poisson rate after accounting for population size effects. *Statist. Sinica* **21** 597–624.

[15] MOUSTAKIDES, G. V. (1986). Optimal stopping times for detecting changes in distributions. *Ann. Statist.* **14** 1379–1387.

[16] MOUSTAKIDES, G. V. (2002). Performance of CUSUM tests for detecting changes in continuous time processes. In *Proceedings of the* 2002 *IEEE International Symposium on Information Theory* 186.

[17] MOUSTAKIDES, G. V. (2004). Optimality of the CUSUM procedure in continuous time. *Ann. Statist.* **32** 302–315.

[18] MOUSTAKIDES, G. V. (2007). CUSUM techniques for sequential change detection. Lecture Notes in Probability, Course 8201, Columbia University, New York.

[19] PAGE, E. S. (1954). Continuous inspection schemes. *Biometrika* **41** 100–115. MR0088850

[20] PESKIR, G. and SHIRYAEV, A. N. (2002). Solving the Poisson disorder problem. In *Advances in Finance and Stochastics* 295–312. Springer, Berlin.

[21] PICARD, P. and LEFÈVRE, C. (1997). The probability of ruin in finite time with discrete claim size distribution. *Scand. Actuar. J.* **1997** 58–69. MR1929384

[22] PISTORIUS, M. R. (2004). On exit and ergodicity of the spectrally one-sided Lévy process reflected at its infimum. *J. Theoret. Probab.* **17** 183–220. MR2054585

[23] POOR, H. V. (1998). Quickest detection with exponential penalty for delay. *Ann. Statist.* **26** 2179–2205.

[24] POOR, H. V. and HADJILIADIS, O. (2009). *Quickest Detection*. Cambridge Univ. Press, Cambridge.

[25] RULLIÈRE, D. and LOISEL, S. (2004). Another look at the Picard–Lefèvre formula for finite-time ruin probabilities. *Insurance Math. Econom.* **35** 187–203.

[26] SHIRYAEV, A. N. (1996). Minimax optimality of the method of cumulative sums (CUSUM) in the case of continuous time. *Russian Math. Surveys* **51** 750–751.

[27] SHIRYAEV, A. N. (2009). On stochastic models and optimal methods in the quickest detection problems. *Theory Probab. Appl.* **53** 385–401.

N. EL KAROUI
LPMA
UNIVERSITÉ PIERRE ET MARIE CURRIE
CASE COURRIER 188
75252 PARIS CEDEX 05
FRANCE
E-MAIL: elkaroui@gmail.com

S. LOISEL
Y. SALHI
LSAF EA 2429, ISFA
UNIVERSITÉ CLAUDE BERNARD LYON 1
50 AVENUE TONY GARNIER
69007 LYON
FRANCE
E-MAIL: stephane.loisel@univ-lyon1.fr
yahia.salhi@univ-lyon1.fr