The matrix sequential probability ratio test and multivariate ruin theory

Hansjoerg Albrecher and Oscar Peralta

Abstract The matrix sequential probability ratio test (MSPRT) is a statistical method to decide which law governs a collection of independent and identically distributed data amongst a finite set of possibilities. By focusing on the case where the possible laws are exponentially tilted versions of each other, in this note we exhibit novel links between the MSPRT and multivariate risk processes with common shocks, as well as with one-dimensional renewal theory.

1 Introduction

The celebrated *sequential probability ratio test* (SPRT) developed in [8] is an efficient tool to identify the nature of a collection of i.i.d. observations employing only a finite subset of data points whose cardinality is random. More specifically, for an independent and identically distributed (i.i.d.) sequence $\{X_\ell\}_{\ell\geq 1}$ with density f, one tests the null hypothesis $H_0 = \{f = f_0\}$ against the alternative hypothesis $H_1 = \{f = f_1\}$, where f_0 and f_1 are given density functions. The decision rule associated to the SPRT is based on employing the likelihood ratio of the first T observations, say $\Lambda^{(T)}$ for $\Lambda^{(k)} := \prod_{\ell=1}^{k} f_0(X_\ell)/f_1(X_\ell)$, where T corresponds to the number of observations it takes $\{\Lambda^{(k)}\}_{k\geq 1}$ to exit the strip (a,b), $0 < a < 1 < b < \infty$. In particular, the null hypothesis H_0 is rejected if and only if $\Lambda^{(T)} \leq a$, with corresponding errors

$$\alpha_0 := \mathbb{P}(\text{reject } H_0 \mid H_0) \quad \text{and} \quad \alpha_1 := \mathbb{P}(\text{do not reject } H_0 \mid H_1)$$

Hansjoerg Albrecher
University of Lausanne and Swiss Finance Institute, Faculty of Business and Economics, Quartier de Chambronne, 1015 Lausanne, Switzerland, e-mail: hansjoerg.albrecher@unil.ch

Oscar Peralta
Cornell University, School of Operations Research and Information Engineering, Rhodes Hall, Ithaca, New York 14853, United States of America, e-mail: op65@cornell.edu

that depend on a and b.

Although the SPRT is a well-established method which in practice produces efficient results in terms of the number T of observations needed, in the general case it is difficult to obtain constants a, b that precisely yield prescribed errors α_0 and α_1; in most cases only bounds exist (see e.g. [9]). In [1], the authors established exact boundaries (a, b) for the prescribed errors for the case f_0 follows a phase-type distribution and f_1 is an exponentially-tilted transform of f_0. To do so, the authors follow a probabilistic approach based on establishing a link between the SPRT and certain Markov additive risk processes and their exit probabilities. The theory of scale functions developed in [7] was then used to compute α_0 and α_1, as well as the expected amount of observations needed, $\mathbb{E}(T)$.

In this paper, we are interested in establishing novel geometric considerations for the matrix sequential probability ratio test (MSPRT) as defined in [5], an extension of the SPRT that allows for multiple hypotheses. Similarly to [1], we focus on the case where the candidate densities are exponentially-tilted transforms of some fixed function. This allows us to redefine the decision rule associated to the MSPRT in terms of exit times of a multivariate risk process with common shocks. An additional transformation resets the problem in terms of a one-dimensional renewal process and its exit time from a region that varies in time. While both problems are, at the moment, not tractable in the ruin theory and applied probability literature, our geometric approach elucidates avenues of research within these disciplines that can yield powerful results for the MSPRT.

The structure of this paper is as follows. In Section 2, we provide a brief introduction of the MSPRT and its properties. In Section 3, we specialize onto the case of multiple hypothesis testing for exponentially-tilted versions of each other, which will allow to draw novel risk- and renewal-theoretical perpectives on the MSPRT. In Section 4 we provide some synthetic examples that illustrate the performance of the MSPRT in the case of exponentially-tilted hypotheses. We finalize in Section 5 by summarizing our findings and pointing out some potential avenues for further resarch.

2 The Matrix Sequential Probability Ratio Test

In a probability space $(\Omega, \mathbb{P}, \mathscr{F})$, suppose that we have a sequence of identically distributed random variables $\{X_\ell\}_{\ell\geq 1}$ and a collection of events that partition Ω, say $\{H_i\}_{i\in\Theta}$ with $i \in \Theta = \{1, \dots, M\}$, where

$$H_i = \{X_\ell \sim F_i \text{ for all } \ell \geq 1\},$$

and each F_i corresponding to a pre-specified distribution function. Furthermore, suppose that conditional on each event H_i, the random variables $\{X_\ell\}_{\ell\geq 1}$ are independent. The previous corresponds to the classical Bayesian setting for *multiple hypothesis testing*. Given a realization of the random variables $\{X_\ell\}_{\ell\geq 1}$, the main goal of multiple hypothesis testing is to choose the event H_i which is the most likely, given the observed sequence.

Let $\mathbb{P}_i$ be the probability measure conditioned on the hypothesis H_i. Here we assume that $\mathbb{P}_i$ and $\mathbb{P}_j$ are equivalent measures for all $i, j \in \Theta$, in the sense that the Radon-Nikodym derivative $\mathrm{d}\mathbb{P}_i/\mathrm{d}\mathbb{P}$ exists and is $\mathbb{P}$-a.e. positive. This is equivalent to assuming the existence of a reference measure ν on $(-\infty,\infty)$ and ν-integrable positive functions $f_1, f_2, \ldots, f_M$, which share the same support, such that

$$F_i(x) = \int_{(-\infty,x]} f_i(y)\mathrm{d}\nu(y), \qquad x \in (-\infty,\infty), i \in \Theta.$$

For the set of observations $\{X_\ell\}$, here we are interested in a method that, within some prespecified error bounds, allows us to reach a consensus w.r.t. the possible events $\{H_i\}_{i\in\Theta}$, using a finite amount of observations *only*. The *matrix sequential probability ratio test* (MSPRT), constructed in [5], produces a pair (d,T) where $T \geq 1$ is an $\mathscr{F}_n$-stopping time associated to the number of observations needed, and d is an $\mathscr{F}_T$-measurable function taking values in Θ associated to the hypothesis that is the most likely. In the previous, $\mathscr{F}_n$ denotes the σ-algebra generated by $X_1, X_2, \ldots, X_n$, while $\mathscr{F}_T$ is the σ-algebra generated by $X_1, \ldots, X_T$.

For the MSPRT, one of the aims is to get small error probabilities

$$\alpha_{ij} = \mathbb{P}_i(d = j), \quad i \neq j, \tag{1}$$

as well as a "small" stopping time T. To properly specify what we mean by this, define the *n-th step likelihood function*

$$\Lambda_i^{(n)} := \left.\frac{\mathrm{d}\mathbb{P}_i}{\mathrm{d}\nu}\right|_{\mathscr{F}_n} (\{X_\ell\}_{\ell\geq 1}) = \prod_{\ell=1}^{n} f_i(X_\ell). \tag{2}$$

Then (d,T) corresponding to the MSPRT in [5] takes the form

$$T = \min_{i\in\Theta}\{T_i\} \quad \text{where} \quad T_i = \inf\left\{n \geq 1 : \Lambda_i^{(n)} \geq A_i\left(\Lambda_1^{(n)}, \cdots, \Lambda_{i-1}^{(n)}, \Lambda_{i+1}^{(n)}, \ldots, \Lambda_M^{(n)}\right)\right\}, \tag{3}$$

$$d = \arg\min_{i\in\Theta}\{T_i\}, \tag{4}$$

for some collection of deterministic functions $A_i : \mathbb{R}^{M-1} \mapsto \mathbb{R}$, $i \in \Theta$. The choice for the collection of functions A_i will impact the *risk* associated to making the decision i, here defined by

$$R_i(d,T) = \sum_{j\neq i} \pi_j \alpha_{ji} \quad \text{with} \quad \pi_j = \mathbb{P}(H_j),$$

as well as on the expected number of observations until a decision is met, $\mathbb{E}(T)$. In fact, in [5] it is shown that $A_i(\cdot)$ can be chosen to be of the form

$$A_i(x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_M) = \frac{M}{\rho_i}\max_{j\neq i}\pi_j x_j \text{ for some fixed } \rho_i > 0, \tag{5}$$

and such a choice yields a risk for the decision i which is lower than ρ_i, i.e. $R_i(d,T) \leq \rho_i$.

While the choice (5) does not yield the lowest expected number of observations for a predetermined level of risk ρ_i, it is a tractable option which still produces an optimal decision in an asymptotic sense. More specifically, let $\boldsymbol{\rho} = (\rho_1,\rho_2,\ldots,\rho_M)$ be a vector with positive entries. We then consider the *class of tests* $\boldsymbol{\Delta}(\boldsymbol{\rho})$ with

$$\boldsymbol{\Delta}(\boldsymbol{\rho}) = \{(c,S) : R_i(c,S) \leq \rho_i \text{ for all } i \in \Theta\},$$

where the set on the r.h.s. spans over all pairs (c,S) defined analogously as (d,T) was defined in (3) and (4) for arbitrary measurable functions A_i.

Theorem 1 ([5]). *Let $\boldsymbol{\rho} \downarrow 0$ denote the instance when the elements of $\boldsymbol{\rho}$ uniformly tend to 0 while mantaining the asymptotic relationship*

$$\frac{\log\rho_i}{\log\rho_j} \to c_{ij}, \quad 0 < c_{ij} < \infty.$$

Then,

$$\inf_{(c,S)\in\boldsymbol{\Delta}(\boldsymbol{\rho})} \mathbb{E}_i(S^m) = \mathbb{E}_i(T^m)(1+o(1)),$$

where $o(1)$ denotes some generic real function which converges to 0 as $\boldsymbol{\rho} \downarrow 0$.

Theorem 1 essentially implies that, amongst all the tests which yield risks which are below a prespecified vector $\boldsymbol{\rho}$, (d,T) with A_i defined as in (5) enjoys asymptotic optimality w.r.t. the moments of T as the entries of $\boldsymbol{\rho}$ tend to 0. Due to this and its simplicity, in practice it is common to employ the MSRPT with the test (d,T), even if it is not optimal in a non-asymptotic sense. From now on, whenever we refer to the MSPRT test (d,T), it will be understood that it is the one defined by (5).

3 MSPRT for exponential tilting

In this section, we look at a special type of MSPRT, where the collection of hypotheses corresponds to distributions which are all exponentially tilted versions of some reference probability measure. More specifically, suppose that there exists a distribution function taking the form

$$F(x) = \int_{(-\infty,x]} f(y)\mathrm{d}\nu(y), \quad x \in (-\infty,\infty) \tag{6}$$

for some nonnegative measurable function f, and a collection of real numbers $\{\gamma_i\}_{i\in\Theta}$ such that $f_i(x) = \frac{e^{\gamma_i x}f(x)}{L(\gamma_i)}$, where $L(\gamma_i) = \int e^{\gamma_i s}f(s)\mathrm{d}\nu(s)$ is assumed to be finite for all $i \in \Theta$. Under these particular circumstances, the n-th step likelihood function takes the form

$$\Lambda_i^{(n)} := \prod_{\ell=1}^{n} \frac{e^{\gamma_i X_\ell}}{L(\gamma_i)}. \tag{7}$$

Transforming (7) by taking logarithms leads us to an equivalent definition to (3),

$$T_i = \inf\left\{ n \geq 1 : \lambda_i^{(n)} \geq \log M/\rho_i + \max_{j \neq i} \left\{ \log \pi_j + \lambda_j^{(n)} \right\} \right\}$$

where

$$\lambda_i^{(n)} := \log \Lambda_i^{(n)} = \gamma_i X_\ell - \log L(\gamma_i).$$

3.1 Links to multivariate risk processes

From now on, let us assume that f has support in $(0, \infty)$, i.e. $X_\ell > 0$ a.s. for all $\ell \geq 1$, and w.l.o.g. suppose that $\gamma_1 < \gamma_2 < \cdots < \gamma_M$. Let $\{N(t)\}_{t \geq 0}$ be the *renewal process* whose interarrival times are $\{X_\ell\}_{\ell \geq 1}$. A key aspect of this paper is to notice that the stopping time T_i corresponds to the minimum number of arrivals of $\{N(t)\}_{t \geq 0}$ needed so that the i-th row of the multidimensional array process $\boldsymbol{Y}(t) = \{Y_{ij}(t)\}_{i,j \in \Theta}$ of the form

$$Y_{ij}(t) = (\gamma_i - \gamma_j)t - \sum_{\ell=1}^{N(t)} (\log L(\gamma_i) - \log L(\gamma_j))$$

is simultaneously above certain boundaries. More precisely, T_i can be expressed as

$$T_i = \inf\left\{ \ell \geq 1 : Y_{ij}(W_\ell) \geq b_{ij} \text{ for all } j \neq i \right\},$$

where $W_n = \sum_{\ell=1}^{n} X_\ell$ and $b_{ij} = \log M/\rho_i + \log \pi_j$. Examining each of the processes $\{Y_{ij}(t)\}_{t \geq 0}$ for all $i, j \in \Theta$, we get:

- In the case $i > j$, $\{Y_{ij}(t)\}_{t \geq 0}$ corresponds to a process that increases linearly with slope $\gamma_i - \gamma_j$ and has negative jumps of size $(\log L(\gamma_i) - \log L(\gamma_j))$ at the arrival times of $\{N(t)\}$; this can be regarded as a Sparre-Andersen model with deterministic jump sizes (cf. [1]).
- In the case $i < j$, $\{Y_{ij}(t)\}_{t \geq 0}$ decreases linearly with slope $\gamma_i - \gamma_j$ and has positive jumps of size $(\log L(\gamma_i) - \log L(\gamma_j))$ instead; this can be also be regarded as a Sparre-Andersen model, but flipped along the x-axis.
- The case $i = j$ simply yields $\{Y_{ij}(t)\}_{t \geq 0}$ to be the zero process.

This elucidates a novel risk-theoretic understanding of the MSPRT, for the exponentially tilted case, in terms of risk processes (classical and flipped) and their upcrossing probabilities at the epochs $\{S_\ell\}_{\ell \geq 0}$. In case we want to link the MSPRT to the continuous upcrossing probabilities (as opposed to the discretely observed ones), we can define

$$b_{ij}^* = \begin{cases} b_{ij} & \text{if } i < j, \\ b_{ij} + (\log L(\gamma_i) - \log L(\gamma_j)) & \text{if } i > j, \\ 0 & \text{if } i = j. \end{cases}$$

In this case, it can be readily verified that

$$T_i = \inf\left\{\ell \geq 1 : \text{for some } s \in [0, W_\ell], Z_{ij}(s) \geq (\gamma_i - \gamma_j)^{-1} b_{ij}^* \text{ for all } j \in \theta\right\}, \quad i \in \Theta.$$

For geometric considerations, it is convenient to normalize the processes Y_{ij} to move linearly at a positive unit rate. Thus, we define the Sparre-Andersen processes

$$Z_{ij}(t) = \frac{Y_{ij}}{\gamma_i - \gamma_j} = t - \sum_{\ell}^{N(t)} \frac{\log L(\gamma_i) - \log L(\gamma_j)}{\gamma_i - \gamma_j}, \quad i \neq j. \tag{8}$$

Given that $\gamma_i - \gamma_j > 0$ if and only if $i > j$, then the stopping rule for the i-th hypothesis becomes

$$\begin{aligned} T_i &= \inf\left\{\ell \geq 1 : Z_{ij}(W_\ell) \geq c_{ij} \text{ for all } j < i \text{ and } Z_{ij}(W_\ell) \leq c_{ij} \text{ for all } j > i\right\} \qquad (9) \\ &= \inf\left\{\ell \geq 1 : \begin{array}{c} \text{for some } s \in [0, S_\ell], Z_{ij}(s) \geq c_{ij}^* \text{ for all } j < i \\ \text{and } Z_{ij}(s) \leq c_{ij}^* \text{ for all } j > i \end{array}\right\}, \end{aligned}$$

where $c_{ij} = (\gamma_i - \gamma_j)^{-1} b_{ij}$ and $c_{ij}^* = (\gamma_i - \gamma_j)^{-1} b_{ij}^*$. Then T_i can be seen to correspond to the simultaneous hitting time of a multivariate Sparre-Andersen process with common shocks.

Remark 1. In the case that $f_1, \ldots, f_M$ are all exponential density functions, the processes will be linked to Cramér-Lundberg processes with common shocks, and moreover, with jumps that are constant. In fact, the resulting multivariate process is a subclass of the risk models considered in [4], where the authors present a multidimensional Cramér-Lundberg process with simultaneous arrivals and ordered claim sizes. Unfortunately, their results yield ruin probabilities only, while in our case, simultaneous downcrossing and upcrossing probabilities are needed. Nevertheless, this presents a promising avenue for further research.

3.2 Links to one-dimensional renewal theory

Similar to the transformation of the bivariate problem in [2] and [3], an additional understanding of T_i can be obtained in terms of the *one-dimensional* process $\{N(t)\}_{t \geq 0}$ and certain hitting times of barriers that are time-varying. More specifically, dividing (8) by $(\log L(\gamma_i) - \log L(\gamma_j))/(\gamma_i - \gamma_j)$, we get

$$N(t) = (t - Z(t)) \frac{\gamma_i - \gamma_j}{\log L(\gamma_i) - \log L(\gamma_j)}.$$

By plugging the previous in (9),

$$T_i = \inf\left\{\ell \geq 1 : N(W_\ell) \leq s_{ij}W_\ell - d_{ij} \text{ for all } j < i \text{ and } N(W_\ell) \geq s_{ij}W_\ell - d_{ij} \text{ for all } j > i\right\} \tag{10}$$

$$= \inf\left\{\ell \geq 1 : \begin{array}{c} \text{for some } s \in [0, S_\ell],\, N(s) \leq s_{ij}t - d^*_{ij} \text{ for all } j < i \\ \text{and } N(s) \geq s_{ij}t - d^*_{ij} \text{ for all } j > i \end{array}\right\},$$

where $s_{ij} = \frac{\gamma_i - \gamma_j}{\log L(\gamma_i) - \log L(\gamma_j)}$, $d_{ij} = (\log L(\gamma_i) - \log L(\gamma_j))^{-1} b_{ij}$ and $d^*_{ij} = (\log L(\gamma_i) - \log L(\gamma_j))^{-1} b^*_{ij}$.

Now, note that due to the monotonicity of $L(\cdot)$, $s_{ij} = s_{ji} > 0$, $d^*_{ij} > 0$ for all $j < i$, and $d^*_{ij} < 0$ for all $j > i$. The values $\{s_{ij}\}_{i \neq j}$ are, in general, not ordered. Nevertheless, it is possible to order a subset of it. To showcase this, let us first prove the following technical result.

Lemma 1. *Define $\Sigma(\lambda) = \log(L(\gamma))$ for all $\gamma \in (-\infty, c)$, where $c = \sup\{z \in \mathbb{R} : L(\gamma) < \infty\}$. Then, the function $\Sigma(\cdot)$ is convex in $(-\infty, c)$.*

Proof. Fix $-\infty < \gamma < \gamma_* < c$ and $s \in (0,1)$. Then, for a random variable $Z \sim f$ and constants $p, q \geq 0$ such that $1/p + 1/q = 1$, by Hölder's inequality we get

$$\begin{aligned} \Sigma(s\gamma + (1-s)\gamma_*) &= \log\left(\mathbb{E}\left(e^{s\gamma Z} e^{(1-s)\gamma_* Z}\right)\right) \\ &\leq \log\left(\mathbb{E}\left(e^{ps\gamma Z}\right)^{1/p} \mathbb{E}\left(e^{q(1-s)\gamma_* Z}\right)^{1/q}\right) \\ &= \frac{1}{p}\Sigma(ps\gamma) + \frac{1}{q}\Sigma(q(1-s)\gamma_*) \\ &= s\Sigma(\gamma) + (1-s)\Sigma(\gamma_*), \end{aligned}$$

where we subsitute $p = 1/s$ and $q = 1/(1-s)$ in the last equality, verifying the convexity of $\Sigma(\cdot)$.

Corollary 1. *Define $s_{M,M+1} = 0$. Then, for all $0 \leq i \leq j \leq M$, we have $s_{i,i+1} \geq s_{j,j+1}$.*

Proof. If $j = M$, then the inequality $s_{i,i+1} \geq s_{j,j+1}$ trivially holds. If $j < M$, employing the convexity of Σ we get

$$\frac{1}{s_{i,i+1}} = \frac{\Sigma(\gamma_{i+1}) - \Sigma(\gamma_i)}{\gamma_{i+1} - \gamma_i} \leq \frac{\Sigma(\gamma_{j+1}) - \Sigma(\gamma_j)}{\gamma_{j+1} - \gamma_j} = \frac{1}{s_{j,j+1}},$$

which in turn implies that $s_{i,i+1} \geq s_{j,j+1}$.

Given the order of the collection $\{s_{i,i+1}\}$ that Corollary 1 yields, the decision rule takes the following simple geometric form:

1. In the Euclidian space $\mathbb{R}^2_+$, draw lines that cross 0 with slopes $\{s_{ij}\}_{i<j}$.

2. For the line corresponding to the slope s_{ij} draw a barrier d_{ij}-units above and another barrier $|d_{ji}|$-units below. Let $B_{ij} \subset \mathbb{R}^2_+$ be the set of points lying strictly between those barriers (see Figure 1).
3. Let C_i be the set of points lying strictly between the lines with slope $s_{(i-1),i}$ and $s_{i,(i+1)}$ (defining $s_{0,1} = +\infty$ and $s_{M,M+1} = 0$).
4. Accept H_i as the correct hypothesis for the i for which $\{(W_\ell, N(W_\ell))\}_{\ell \geq 1}$ enters $C_i \setminus ((\cup_{\ell<i} B_{\ell i}) \cup (\cup_{\ell>i} B_{i\ell}))$ for the first time (see Figure 1).

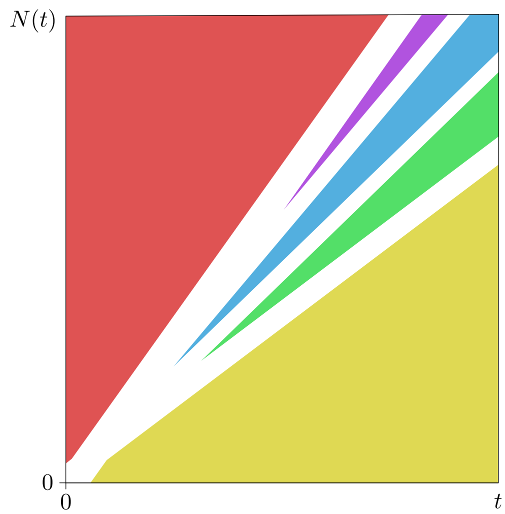

Fig. 1: Acceptance areas associated to $C_i \setminus ((\cup_{\ell<i} B_{\ell i}) \cup (\cup_{\ell>i} B_{i\ell}))$ for $i \in \Theta = \{1,2,3,4,5\}$: Acceptance area for H_1 (red), for H_2 (violet), for H_3 (blue), for H_4 (green), and for H_5 (mustard).

Remark 2. Analogously, we can also define the process in terms of the hitting events of the continuously observed process $\{(t, N(t))\}_{t \geq 0}$. More specifically, we declare i as the correct hypothesis if $\{(t, N(t))\}_{t \geq 0}$ enters $C_i \setminus \left(\left(\cup_{\ell<i} B^*_{\ell i}\right) \cup \left(\cup_{\ell>i} B^*_{i\ell}\right)\right)$ first, where the definition of B^*_{ij} is akin to that of B_{ij} with d_{ij} and d_{ji} being replaced by d^*_{ij} and d^*_{ji}.

Within this framework, there are some non-trivial quantities of interest to compute. First, note that

$$\mathbb{P}_i(d \neq i) \leq \mathbb{P}_i(\sigma_i < \infty), \quad \text{where} \tag{11}$$

$$\sigma_i = \inf\{\ell \geq 1 : (W_\ell, N(W_\ell)) \notin C_i \cup B_{i-1,i} \cup B_{i,i+1}\} \tag{12}$$

The previous simply states that no *wrong* decision can be taken if the set of points $\{(W_\ell, N(W_\ell))\}_{\ell\geq 1}$ never leaves an area which contains the entire acceptance area for H_i. Interestingly, the r.h.s in (11) corresponds to the exit probability of a two-dimensional random walk on a cone. This is investigated in [6]. However, the authors there investigate the exponential decay of the non-exit probability. Their work relies on identifying the dual cone K_* of C_i, defined by

$$K_* = \{z \in \mathbb{R}^2 : \langle x, z\rangle \geq 0 \;\forall\, x \in C_i\}. \tag{13}$$

Then, they show that

$$\limsup_{n\to\infty} \mathbb{P}\left(\sigma_i > n\right)^{1/n} = \inf_{z\in K_*} \tilde{L}(z), \quad \text{where} \tag{14}$$

$$\tilde{L}(z) = \int_{\mathbb{R}^2} e^{\langle z,y\rangle} \left(\delta_1(y_1) \times f_i(y_2)\mathrm{d}\nu(y_2)\right). \tag{15}$$

In our case, K_* is the cone between the lines $z_2 = (-1/s_{i,i+1})z_1$ and $z_2 = (-1/s_{i-1,i})z_1$; it is readily verified that the infimum of $\tilde{L}(z)$ over K^* is actually reached at these lines. Thus, we need to find the infimum of the functions

$$g_1(z_2) = \int_{\mathbb{R}^2} e^{-s_{i,i+1}z_2y_1 + z_2y_2} \left(\delta_1(y_1) \times f_i(y_2)\mathrm{d}\nu(y_2)\right)$$
$$g_2(z_2) = \int_{\mathbb{R}^2} e^{-s_{i-1,i}z_2y_1 + z_2y_2} \left(\delta_1(y_1) \times f_i(y_2)\mathrm{d}\nu(y_2)\right).$$

In the case of g_1,

$$\begin{aligned}\log(g_1(z_1)) &= -s_{i,i+1}z_2 + \log\left(\int_{\mathbb{R}^2} e^{z_2y_2} f_i(y_2)\mathrm{d}\nu(y_2)\right)\\ &= -s_{i,i+1}z_2 + \log\left(L(z_2+\gamma_i)\right).\end{aligned}$$

Given that the function Σ is convex and increasing, then

$$\frac{\log\left(L(z_2+\gamma_i)\right)}{\mathrm{d}z_2} = \Sigma'(\gamma_i) \geq \frac{\Sigma(\gamma_i) - \Sigma(\gamma_{i-1})}{\gamma_i - \gamma_{i-1}} = s_{i-1,i},$$

meaning that the minumum of g_1 is reached at $z_2 = 0$. Analogous arguments follow for g_2, which in turn implies that the infimum of $\tilde{L}$ over K^* is reached at $(z_1, z_2) = (0,0)$. In short, the results of [6] confirm that exiting from the cone C_i is not a certain event.

4 Numerical examples

In the following, we perform a simulation study of the MSPRT in order to measure its efficiency from an empirical perspective; we do this following the implementation laid out in Subsection 3.2, though we stress that the one in [5] or Subsection 3.1 produces the same results by alternative means. Suppose that ν in (6) is the Lebesgue measure and $f(x) = e^{-x}$, from which we sample an i.i.d. sequence $\{X_\ell\}_{\ell\geq 1}$ with $X_\ell \sim f$. In this analysis, we desire to test if the data $\{X_\ell\}_{\ell\geq 1}$ identifies via the MSPRT the correct hypohesis or not, with alternative hypotheses being "close" to the original one. More specifically, we take $\Theta = \{1,2,3,4,5\}$, and let the tilting values be $\gamma_1 = -2\beta$, $\gamma_2 = -\beta$, $\gamma_3 = 0$, $\gamma_4 = \beta$ and $\gamma_5 = 2\beta$ for some $\beta \in (0,1/2)$. Furthermore, let $\boldsymbol{\rho}$ take the form

$$\boldsymbol{\rho} = (\rho/2, \rho, \rho, \rho, \rho/2), \quad \rho \in (0,1);$$

given the geometric considerations of Figure 1, it makes sense to consider a stricter level of risk for H_1 and H_5, as they have only one neighbouring region each.

Now, take $\beta = 0.2$ and generate 200 samples of sequences $\{X_\ell\}_{\ell\geq 1}$ which follow the density function f; note that the MSPRT choosing H_3 for a given sample would correspond to a correct labeling, and any other case is a mislabel; see Figure 2 for a sample which was labelled correctly. Out of the 200 samples, we

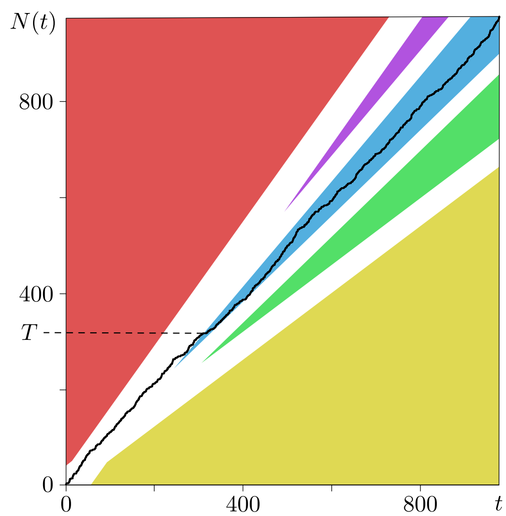

Fig. 2: Renewal process $N(t)$, shown in black, associated to one sample of simulated data $\{X_\ell\}_{\ell\geq 1}$. The process enters the acceptance region of H_3 at $T = 321$ observations.

then count for how many of them the MSPRT chooses H_i, $i \in \Theta$, for different $\rho = 0.5, 0.25, 0.1, 0.05, 0.025$, and provide the empirical average of samples needed to reach a conclusion, T. These results are shown in Table 1 below.

ρ	H_1	H_2	H_3	H_4	H_5	Empirical mean of T
0.5	0	23	159	18	0	123.97
0.25	0	1	199	0	0	373.15
0.1	0	0	200	0	0	981.975
0.05	0	0	200	0	0	1904.245
0.025	0	0	200	0	0	3720.985

Table 1: Counts of 200 sample runs resulting in H_i, $i \in \Theta$, for $\beta = 0.2$ and $\rho = 0.5, 0.25, 0.1, 0.05, 0.025$, along with empirical average of T.

As we can observe, the MSPRT is remarkably accurate at identifying the correct hypothesis, even when ρ is moderately large. In fact, the discrepancy between the risk ρ and the proportion of instances the algorithm chooses the wrong hypothesis is considerable: this is explained by the fact that ρ is proven to be an upper bound to the actual risk, which heuristically seems to be much lower. As expected, the mean number of samples until a decision is taken increases as ρ decreases.

Repeating the aforementioned procedure for $\beta = 0.1$ yields the results presented in Table 2 below.

ρ	H_1	H_2	H_3	H_4	H_5	Empirical mean of T.
0.5	0	23	151	26	0	467.95
0.25	0	0	199	1	0	1469.19
0.1	0	0	200	0	0	3741.455
0.05	0	0	200	0	0	7433.485
0.025	0	0	200	0	0	14445.49

Table 2: Counts of 200 sample runs resulting in H_i, $i \in \Theta$, for $\beta = 0.1$ and $\rho = 0.5, 0.25, 0.1, 0.05, 0.025$, along with empirical average of T

We can notice that in this case, the proportion of times the MSPRT chooses the right hypothesis is comparable to the case $\beta = 0.2$, however, the mean number of samples until a decision is taken is considerably higher. This is consistent with the fact that we are comparing hypotheses that are "closer" than those in the case $\beta = 0.2$, and the algorithm needs more data to reach a proper conclusion.

To stress the MSPRT further, we consider the case where $\beta = 0.05$ and collect the results in Table 3 below.

Once again, the frequency with which the MSPRT chooses the right hypothesis in this instance is comparable to the cases $\beta = 0.2, 0.1$, while the mean number of needed samples until a decision is taken is the largest amongst all cases.

ρ	H_1	H_2	H_3	H_4	H_5	Empirical mean of T
0.5	0	24	156	20	0	1857.165
0.25	0	1	197	2	0	5738.82
0.1	0	0	200	0	0	15084.21
0.05	0	0	200	0	0	29105.39
0.025	0	0	200	0	0	56722.98

Table 3: Counts of 200 sample runs resulting in H_i, $i \in \Theta$, for $\beta = 0.05$ and $\rho = 0.5, 0.25, 0.1, 0.05, 0.025$, along with empirical average of T.

5 Conclusion

We provided a novel understanding of the MSPRT for exponential tilting in terms of the first passage times of a multivariate risk process and a renewal process with time-varying barriers. Both proposed alternative approaches indicate interesting future avenues of research: the former is connected to recent research on queues with simultaneous arrivals [4], and the latter with first passage probabilities of multidimensional random walks in cones [6]. Furthermore, our simulated numerical experiments suggest the need to pursue tighter bounds for the risks guaranteed for a given MSPRT, via both proposed alternative interpretations provided in this paper.

Acknowledgements The authors would like to acknowledge financial support from the Swiss National Science Foundation Project 200021_191984.
Declaration HA and OP declare no conflict of interest related to the current manuscript.

References

1. Albrecher, H., Asadi, P., Ivanovs, J.: Exact boundaries in sequential testing for phase-type distributions. Journal of Applied Probability **51**(A), 347–358 (2014)
2. Avram, F., Palmowski, Z., Pistorius, M.: A two-dimensional ruin problem on the positive quadrant. Insurance: Mathematics and Economics **42**(1), 227–234 (2008)
3. Badescu, A.L., Cheung, E.C., Rabehasaina, L.: A two-dimensional risk model with proportional reinsurance. Journal of Applied Probability **48**(3), 749–765 (2011)
4. Badila, E., Boxma, O., Resing, J., Winands, E.: Queues and risk models with simultaneous arrivals. Advances in Applied Probability **46**(3), 812–831 (2014)
5. Draglia, V., Tartakovsky, A.G., Veeravalli, V.V.: Multihypothesis sequential probability ratio tests. I. Asymptotic optimality. IEEE Transactions on Information Theory **45**(7), 2448–2461 (1999)
6. Garbit, R., Raschel, K.: On the exit time from a cone for random walks with drift. Revista Matematica Iberoamericana **32**(2), 511–532 (2016)
7. Ivanovs, J., Palmowski, Z.: Occupation densities in solving exit problems for Markov additive processes and their reflections. Stochastic Processes and their Applications **122**(9), 3342–3360 (2012)
8. Wald, A.: Sequential Tests of Statistical Hypotheses. The Annals of Mathematical Statistics **16**(2), 117 – 186 (1945)
9. Wald, A., Wolfowitz, J.: Optimum character of the sequential probability ratio test. The Annals of Mathematical Statistics pp. 326–339 (1948)