Exact boundaries in sequential testing for phase-type distributions

Hansjörg Albrecher, Peiman Asadi and Jevgenijs Ivanovs

Department of Actuarial Science, University of Lausanne, CH-1015 Lausanne, Switzerland

Abstract: We consider Wald's sequential probability ratio test for deciding whether a sequence of independent and identically distributed observations comes from a specified phase-type distribution or from an exponentially tilted alternative distribution. In this setting, we derive exact decision boundaries for given Type I and Type II errors by establishing a link with ruin theory. Information on the mean sample size of the test can be retrieved as well. The approach relies on the use of matrix-valued scale functions associated to a certain one-sided Markov additive process. By suitable transformations the results also apply to other types of distributions including some distributions with regularly varying tail.

Keywords and phrases: sequential probability ratio test, Markov additive process, scale function, two-sided exit problems.

1. Introduction

Consider Wald's sequential probability ratio test [12] of a simple hypothesis against a simple alternative. Let $\zeta_1, \zeta_2, \dots$ be a sequence of independent and identically distributed random variables (observations) with density f, where either $f = f_0$ (hypothesis H_0) or $f = f_1$ (hypothesis H_1). The log-likelihood ratio Λ_k for the first k observations is then given by

$$\Lambda_k = \sum_{i=1}^{k} \log \frac{f_0(\zeta_i)}{f_1(\zeta_i)}, \qquad \Lambda_0 = 0,$$

and its first exit time N from the interval (a, b) by

$$N = \inf\{k \geq 0 : \Lambda_k \notin (a, b)\}, \tag{1}$$

where $a < 0 < b$. At time N the sampling is stopped and a decision is made: accept H_0 if $\Lambda_N \geq b$, and accept H_1 if $\Lambda_N \leq a$. The corresponding errors are given by $\alpha_0 = \mathbb{P}_0(\text{reject } H_0)$ and $\alpha_1 = \mathbb{P}_1(\text{reject } H_1)$, where $\mathbb{P}_i$ indicates that hypothesis H_i is valid.

One now wants to choose decision boundaries a and b so that the errors are below prespecified thresholds. If it is possible to find a and b, such that the errors coincide with their respective thresholds, then Wald's test with such boundaries is known to be optimal (i.e. the expected number of observations (under both hypotheses) is minimal) among all tests respecting these thresholds, see [13] and [10, Thm. IV.4]. Such a pair (a, b) of boundaries is unique under very weak assumptions [14], which do hold in our setting below. Usually, a pair (a, b) resulting in the prespecified errors exists, unless the problem is 'too easy', in which case an optimal test will use zero observations with positive probability, cf. [16] for an analysis of a more general test.

The following simple bounds on the decision boundaries are known, see [12]:

$$a \geq \log \frac{\alpha_1}{1 - \alpha_0}, \quad b \leq \log \frac{1 - \alpha_1}{\alpha_0}. \tag{2}$$

In practice these bounds are often used as actual decision boundaries. As a result, N increases and one of the errors may surpass its threshold, however usually not by a large amount for small errors, see [12].

If the errors α_0 and α_1 can be determined for any fixed pair (a, b), then the optimal decision boundaries can be found by a numerical search for any given pair of errors α_0, α_1 of interest. This inverse problem is however hard even for simple cases. Some tractable examples can be found in Wald [12] and Teugels & Van Assche [11], where the latter assume f_0 and f_1 to be densities of exponential distributions. Some strong asymptotic results were obtained in [8], but they still require identification of the Wiener-Hopf factors corresponding to the random walk Λ_k, which can be done explicitly only in some cases.

In the present work we assume that f_0 and f_1 are densities of phase-type distributions where one can be obtained by exponential tilting of the other. This includes the case of two exponential densities, as well as two Erlang densities with identical shape parameter. After translating the inverse problem of Wald's test into a two-sided exit problem embedded in classical ruin theory (Section 2), we use techniques for Markov additive processes (Section 3) to establish a surprisingly simple identity, which leads to explicit formulas in Section 4. The approach simplifies the proof for the exponential case developed in [11] and extends the results to phase-type densities (taking monotone transformations of the original observations, the results are also applicable for other distributions, such as distributions with regularly varying tails obtained from exponentiating phase-type random variables). In Section 5 we discuss the Erlang case in more detail, for which a very explicit treatment is possible. Section 6 provides a general formula for the expected number of observations in Wald's test. Section 7 studies the uniqueness issue further and considers an extension to a Bayesian version, where an a priori probability for the correctness of H_0 is available. Finally Section 8 provides some numerical illustrations.

2. Wald's test and ruin theory

Let f_0 be a probability density function of some positive random variable ζ, and let $\mathbb{P}_0$ be the corresponding probability measure. Consider the Laplace-Stieltjes transform $G_0(\theta) = \mathbb{E}_0 e^{-\theta\zeta}, \theta \geq 0$ of ζ and define a new tilted measure $\mathbb{P}_1$ according to $\frac{\mathrm{d}\mathbb{P}_1}{\mathrm{d}\mathbb{P}_0} = \frac{1}{G_0(\theta)} e^{-\theta\zeta}$. Then, under $\mathbb{P}_1$, ζ has a probability density function f_1 given by

$$f_1(x) = \frac{1}{G_0(\theta)} e^{-\theta x} f_0(x). \tag{3}$$

Consider Wald's test for densities f_0 and f_1, where $\theta > 0$, and observe that

$$\log \frac{f_0(x)}{f_1(x)} = \theta x + \log G_0(\theta).$$

Hence the log-likelihood ratio Λ_k is a random walk with increments distributed as $\theta\zeta - d$, where $d = -\log G_0(\theta) > 0$ and ζ has density f_0 (under H_0) or f_1 (under H_1). Define the closely related continuous-time stochastic process

$$X_t = \theta t - \sum_{i=1}^{N_t} d, \quad t \geq 0, \tag{4}$$

where N_t is a renewal process with inter-arrival times distributed as ζ, see Figure 1. One can interpret X_t as a surplus process of an insurance portfolio under a Sparre Andersen risk model with initial capital 0, where premiums are collected at constant rate θ, and claims of (deterministic) size d arrive according to the renewal process N_t (see e.g. [2]). Importantly, one can recover the random walk Λ_k from the continuous-time process X_t by considering it at the epochs of jumps. Letting

$$\tau_a^- = \inf\{t \geq 0 : X_t \leq a\}, \quad \tau_b^+ = \inf\{t \geq 0 : X_t \geq b\}$$

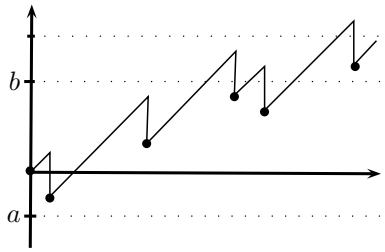


Fig 1: Sparre Andersen risk process

for $a < 0 < b$, we observe that

$$\alpha_0 = \mathbb{P}_0(\tau_a^- < \tau_{b+d}^+), \quad \alpha_1 = \mathbb{P}_1(\tau_{b+d}^+ < \tau_a^-), \tag{5}$$

which is an artifact of the deterministic jumps. Thus we have arrived at a pair of two-sided exit problems for the risk process X_t – one under H_0 and the other one under H_1.

3. Phase-type distributions and Markov additive processes

In this section we present a solution of the two-sided exit problem for the process X_t under the assumption that the generic interarrival time ζ has a phase-type (PH) distribution, i.e. the distribution of the life time of a transient continuous-time Markov Chain (MC) on finitely many states $1, \dots, n$, see e.g. [2]. A PH distribution is parametrized by the transition rate matrix T of the corresponding MC and the row vector $\boldsymbol{\nu}$ representing the initial distribution. Denoting by $\boldsymbol{t} = -T\mathbf{1}$ the column vector of killing (absorption) rates, one can express the density of ζ as

$$f(x) = \boldsymbol{\nu} e^{Tx} \boldsymbol{t}. \tag{6}$$

The Erlang distribution of rate λ is retrieved for $\boldsymbol{\nu} = (1, 0, \dots, 0)$ and choosing T as a square matrix with $-\lambda$ on the main diagonal, λ on the upper diagonal, and 0 elsewhere. Note that the class of PH distributions is dense in the class of all distributions on $(0, \infty)$.

Consider now a bivariate process (X_t, J_t), where X_t is the risk process defined in (4), and J_t tracks the phase of the current interarrival time, which has PH distribution. It is not hard to see that J_t is a MC with transition rate matrix $T + \boldsymbol{t}\boldsymbol{\nu}$, i.e. the transitions can happen due to phase change or due to arrival of a claim (kill and restart). Furthermore, (X_t, J_t) is a simple example of a Markov Additive Process (MAP) without positive jumps, see [2] for a definition. Such a process is characterized by a matrix-valued function $F(s), s \geq 0$, which satisfies $\mathbb{E}(e^{sX_t} 1_{\{J_t = j\}} | J_0 = i) = (e^{F(s)t})_{ij}$ for all $t \geq 0$ and $i, j \in \{1, \dots, n\}$. In our case we have the identity

$$F(s) = T + \theta s \mathbb{I} + \boldsymbol{t}\boldsymbol{\nu} e^{-ds}, \tag{7}$$

where $\mathbb{I}$ is the identity matrix. The diagonal elenments θs represent the linear evolution of X_t with slope θ (the same value in every phase) and $\boldsymbol{t}\boldsymbol{\nu}$ is a matrix of transition rates of J_t causing the jump in X_t with transform e^{-ds}, see [2, Prop. 4.2].

The two-sided exit problem for MAPs without positive jumps was solved in [6], and the solution resembles the one for a Lévy process without positive jumps [7, Thm. 8.1]. According to [6], the matrix of probabilities with

ijth element $\mathbb{P}(\tau_b^+ < \tau_a^-, J_{\tau_b^+} = j | J_0 = i)$ is given by $W(-a)W(-a+b)^{-1}$, where $W(x), x \geq 0$ is a continuous matrix-valued function (called *scale function*) characterized by the transform

$$\int_0^\infty e^{-sx} W(x) \mathrm{d}x = F(s)^{-1} \tag{8}$$

for s large enough. It is known that $W(x)$ is non-singular for $x > 0$ and so is $F(s)$ in the domain of interest. Since J_0 has distribution $\boldsymbol{\nu}$, we write

$$\mathbb{P}(\tau_b^+ < \tau_a^-) = \boldsymbol{\nu} W(-a) W(-a+b)^{-1} \mathbf{1}, \tag{9}$$

with $\mathbf{1} = (1, \ldots, 1)$. Note that the scale function is given in terms of its transform, and the only known explicit examples assume that all jumps of X_t have PH distributions. In the present setting the jumps are not PH but deterministic, which nevertheless gives some hope for the inversion problem. Indeed, in the case of an Erlang distribution for ζ we obtain an explicit representation of $W(x)$, see Section 5.

4. Identification of the errors

In the following we assume that f_0 is a density of a PH distribution with parameters $T_0, \boldsymbol{\nu}_0$ and $\boldsymbol{t}_0 = -T_0\mathbf{1}$, see (6). Its transform is known to be $G_0(\theta) = \boldsymbol{\nu}_0(\theta\mathbb{I} - T_0)^{-1}\boldsymbol{t}_0$. Consider the density f_1, defined in (3), of the corresponding exponentially tilted distribution with the tilt parameter $\theta > 0$. In [1] it is shown that this tilted distribution is again PH, and the parameters are given by

$$T_1 = \Delta^{-1} T_0 \Delta - \theta\mathbb{I}, \quad \boldsymbol{\nu}_1 = \boldsymbol{\nu}_0 \Delta / G_0(\theta), \quad \boldsymbol{t}_1 = \Delta^{-1}\boldsymbol{t}_0, \tag{10}$$

where Δ is a diagonal matrix with $(\theta\mathbb{I} - T_0)^{-1}\boldsymbol{t}_0$ on the diagonal. These diagonal elements are all in $(0,1)$, which can be seen from the representation of $G(\theta) \in (0,1)$ for different initial distributions $\boldsymbol{\nu}_0$.

Since both f_0 and f_1 correspond to PH distributions, we can combine (5) and (9) to obtain

$$\begin{aligned} \alpha_0 &= \mathbb{P}_0(\tau_a^- < \tau_{b+d}^+) = 1 - \boldsymbol{\nu}_0 W_0(-a) W_0(-a+b+d)^{-1}\mathbf{1}, \\ \alpha_1 &= \mathbb{P}_1(\tau_{b+d}^+ < \tau_a^-) = \boldsymbol{\nu}_1 W_1(-a) W_1(-a+b+d)^{-1}\mathbf{1}, \end{aligned} \tag{11}$$

where $W_0(x)$ and $W_1(x)$ are the (matrix-valued) scale functions corresponding to the MAP (X_t, J_t) for $f = f_0$ and $f = f_1$, respectively. Interestingly, W_0 and W_1 are intimately related:

Proposition 1. *The scale functions $W_0(x)$ and $W_1(x)$ satisfy*

$$W_1(x) = e^x \Delta^{-1} W_0(x) \Delta$$

for all $x \geq 0$.

Proof. We establish that

$$\Delta^{-1} F_0(s-1)\Delta = \Delta^{-1}(T_0 - \theta\mathbb{I} + \theta s\mathbb{I} + \boldsymbol{t}_0\boldsymbol{\nu}_0/G_0(\theta)e^{-ds})\Delta = T_1 + \theta s\mathbb{I} + \boldsymbol{t}_1\boldsymbol{\nu}_1 e^{-ds} = F_1(s),$$

by using (7), (10) and recalling that $d = -\log G_0(\theta)$. Now we can check that the proposed matrix-valued function indeed gives the desired transform, see (8):

$$\int_0^\infty e^{-sx}\left(e^x\Delta^{-1}W_0(x)\Delta\right)\mathrm{d}x = \Delta^{-1}\int_0^\infty e^{-(s-1)x} W_0(x)\mathrm{d}x\Delta = \Delta^{-1}F_0(s-1)^{-1}\Delta = F_1(s)^{-1}$$

for s sufficiently large. The result follows, because the transform identifies the continuous $W_1(x), x \geq 0$ uniquely. □

Remark 1. *This curious relation – revealed by an application in sequential testing – would be hard to obtain by simple tailoring of parameters - the corresponding quantities simplify in an intriguing way. It also paves the way for further interesting relations between the two processes, which, however, are outside the scope of the present paper.*

Combining (11) and Proposition 1, we obtain the following result.

Theorem 1. *Let f_0 be a density of a PH distribution with parameters $T_0, \boldsymbol{\nu}_0, \boldsymbol{t}_0$, and f_1 be the corresponding exponentially tilted density with the tilt parameter $\theta > 0$. The errors α_0 and α_1 corresponding to the decision boundaries $a < 0 < b$ in the Wald test of f_0 against f_1 are given by*

$$\alpha_0 = 1 - \boldsymbol{\nu}_0 W_0(-a) W_0(-a+b+d)^{-1}\mathbf{1},$$
$$\alpha_1 = e^{-b}\boldsymbol{\nu}_0 W_0(-a) W_0(-a+b+d)^{-1}(\theta\mathbb{I} - T_0)^{-1}\boldsymbol{t}_0,$$

where $d = -\log G_0(\theta) > 0$, $G_0(\theta)$ is the Laplace transform of f_0, and the continuous matrix-valued function $W_0(x), x \geq 0$, is identified by

$$\int_0^\infty e^{-sx} W_0(x)\mathrm{d}x = \left(T_0 + \theta s\mathbb{I} + \boldsymbol{t}_0\boldsymbol{\nu}_0 e^{-ds}\right)^{-1}$$

for large s.

The transform of $W_0(x)$ can be inverted in certain cases. In Section 5 we provide an explicit expression of $W_0(x)$ when f_0 (and then also f_1) is the density of an Erlang distribution. In other cases one can use numerical methods.

In addition, Theorem 1 provides simple bounds for the level b. First, observe that $\boldsymbol{\nu}_0 W_0(-a) W_0(-a+b+d)^{-1}$ is a vector of probabilities, and recall that all the entries of $(\theta\mathbb{I} - T_0)^{-1}\boldsymbol{t}_0$ are in $(0,1)$. Then we can write

$$m(1-\alpha_0) \leq \alpha_1 e^b \leq (1-\alpha_0)M,$$

where m and M are the minimal and the maximal entries of $(\theta\mathbb{I} - T_0)^{-1}\boldsymbol{t}_0$. Hence also

$$\log\frac{1-\alpha_0}{\alpha_1} + \log m \leq b \leq \log\frac{1-\alpha_0}{\alpha_1} + \log M, \tag{12}$$

where both $\log m$ and $\log M$ are negative. This provides an improvement (for the PH case) of the widely used general Wald bound $b \leq \log\frac{1-\alpha_0}{\alpha_1}$.

Example 1. *If f_0 is the density of an exponential distribution with rate λ_0, then f_1 is a density of an exponential distribution with rate $\lambda_1 = \lambda_0 + \theta$. Here the matrix T_0 reduces to a scalar $-\lambda_0$, and hence $m = M = \lambda_0/(\lambda_0+\theta) = \lambda_0/\lambda_1$ leading to $b = \log\frac{1-\alpha_0}{\alpha_1} - \log\frac{\lambda_1}{\lambda_0}$. This simple identity for exponential densities was already established in [11]. Computation of the boundary $a < 0$ is more involved, and relies on the identity*

$$W_0(-a)/W_0(-a + \log\frac{1-\alpha_0}{\alpha_1}) = 1 - \alpha_0,$$

where $W_0(x)$ will be identified in Section 5.

In general, we do not have a closed form solution for b, and hence the two equations in Theorem 1 need to be solved simultaneously.

5. Erlang against Erlang

Throughout this section we consider the case when f_0 is the density of an Erlang distribution with n phases and rate λ_0, i.e. $f_0(x) = \lambda_0^n x^{n-1} e^{-\lambda_0 x}/(n-1)!$, which has Laplace transform $G_0(\theta) = \left(\frac{\lambda_0}{\lambda_0+\theta}\right)^n$. Exponential tilting of f_0 with the tilt parameter $\theta > 0$ results in f_1, which is another Erlang density on n phases, but with rate $\lambda_1 = \lambda_0 + \theta$. Hence our setup allows to consider two arbitrary Erlang distributions with the same number of phases.

Under the Erlang assumption, the jump size $d = -n\log(\lambda_0/\lambda_1)$ only depends on the ratio $\rho = \lambda_0/\lambda_1$ of the two rates, not on their absolute values. Also, since $\theta \cdot \text{Erlang}(n, \lambda_i) \sim \text{Erlang}(n, \lambda_i/\theta)$, a scaling of θ down to 1 simply stretches the process X_t of (4) in the horizontal direction by the factor θ (under both hypotheses) and the law of the random walk Λ_k is unchanged. Hence Wald's test only depends on the ratio ρ and w.l.o.g. we can choose $\theta = 1$, i.e. $\lambda_1 = \lambda_0 + 1$, leading to $\lambda_0 = \rho/(1-\rho)$ and $\lambda_1 = 1/(1-\rho)$ for the ratio $\rho = \lambda_0/\lambda_1 \in (0,1)$.

Consider the PH parameters $T_0, \boldsymbol{\nu}_0$ and $\boldsymbol{t}_0$ of the density f_0, where T_0 is an $n \times n$ matrix with $-\lambda_0$ on the diagonal, λ_0 on the upper diagonal and 0 elsewhere; $\boldsymbol{\nu}_0 = (1, 0, \dots, 0) = \boldsymbol{e}_1$ and $\boldsymbol{t}_0 = (0, \dots, 0, \lambda_0)'$. Some algebraic manipulations show that the vector $(\theta\mathbb{I} - T_0)^{-1}\boldsymbol{t}_0$ simplifies to $(\rho^n, \rho^{n-1}, \dots, \rho^1)'$, and so by Theorem 1 we have

$$\begin{aligned} \alpha_0 &= 1 - \boldsymbol{e}_1 W_0(-a) W_0(-a+b+d)^{-1}\mathbf{1}, \\ \alpha_1 &= e^{-b}\boldsymbol{e}_1 W_0(-a) W_0(-a+b+d)^{-1}(\rho^n, \rho^{n-1}, \dots, \rho^1)', \end{aligned} \tag{13}$$

where $d = -n\log\rho$, the transform of $W_0(x)$ is given by $F_0(s)^{-1}$, and according to (7)

$$F_0(s) = \begin{pmatrix} s-\lambda_0 & \lambda_0 & 0 & \dots & 0 \\ 0 & s-\lambda_0 & \lambda_0 & \dots & 0 \\ & & \dots & & \\ \lambda_0 e^{-sd} & 0 & 0 & \dots & s-\lambda_0 \end{pmatrix} \tag{14}$$

for $n \geq 2$, whereas $F_0(s) = s - \lambda_0 + \lambda_0 e^{-sd}$ for $n = 1$. The bounds (12) for b now simplify to

$$\log\frac{1-\alpha_0}{\alpha_1} - n\log\rho^{-1} \leq b \leq \log\frac{1-\alpha_0}{\alpha_1} - \log\rho^{-1}, \tag{15}$$

where $\log\rho^{-1} > 0$. It turns out that $W_0(x)$ has a relatively simple expression as a sum of $\lfloor x/d \rfloor$ terms.

Theorem 2. *Consider a MAP with n phases characterized by $F_0(s)$ given in (14) for an arbitrary $d > 0$. Then the ijth element of the scale function $W_0(x)$ for $x \geq 0$ is given by*

$$W_0(x)_{ij} = \sum_{k=1_{\{i>j\}}}^{\lfloor x/d \rfloor} g(\lambda_0(x-dk), kn+j-i), \qquad i,j = 1, \dots, n \tag{16}$$

where $g(y,m) = \frac{(-y)^m}{m!}e^y$.

Proof. In the proof we drop the subscript 0. We need to invert the transform $\int_0^\infty e^{-sx}W(x)\mathrm{d}x = F(s)^{-1}$. Application of Cramer's rule and careful computation of co-factors yields

$$\left(F(s)^{-1}\right)_{ij} = \frac{1}{(s-\lambda)^n - (-\lambda)^n e^{-sd}} \times \begin{cases} (-\lambda)^l (s-\lambda)^{n-l-1}, & i \leq j \\ (-\lambda)^l (s-\lambda)^{n-l-1} e^{-sd}, & i > j \end{cases},$$

where $l = (j-i) \mod n$. Note that the fraction in front can be written as $(s-\lambda)^{-n} \sum_{k=0}^{\infty} (-\lambda)^{kn} (s-\lambda)^{-kn} e^{-sdk}$ for sufficiently large s. Hence

$$\left(F(s)^{-1}\right)_{ij} = \sum_{k=0}^{\infty} \frac{(-\lambda)^{kn+l}}{(s-\lambda)^{kn+l+1}} e^{-sd(k+1_{\{i>j\}})}.$$

Using $\int_0^\infty e^{-sx} \frac{x^n}{n!} e^{\lambda x} \mathrm{d}x = \frac{1}{(s-\lambda)^{n+1}}$ we invert $\frac{(-\lambda)^{kn+l}}{(s-\lambda)^{kn+l+1}}$ to obtain $\frac{(-\lambda x)^{kn+l}}{(kn+l)!} e^{\lambda x} = g(\lambda x, kn+l)$. The factor $e^{-sd(k+1_{\{i>j\}})}$ amounts to shifting x to $x - d(k+1_{\{i>j\}})$. Hence for $j \geq i$

$$W(x)_{ij} = \sum_{k=0}^{\infty} g(\lambda(x-dk), kn+j-i) 1_{\{x \geq dk\}} = \sum_{k=0}^{\lfloor x/d \rfloor} g(\lambda(x-dk), kn+j-i).$$

Similarly, for $j < i$ we have

$$W(x)_{ij} = \sum_{k=0}^{\infty} g(\lambda(x-d(k+1)), kn+n+j-i) 1_{\{x \geq d(k+1)\}} = \sum_{k=1}^{\lfloor x/d \rfloor} g(\lambda(x-dk), kn+j-i),$$

which concludes the proof. □

The quantity (16) is intimately connected with the waiting time distribution in an $E(n)/D/1$ queue, see for instance [3]. In a risk theory context, for the case $n = 1$ (which refers to the compound Poisson model with deterministic jumps), formula (16) can already be found in [9, Sec.3.3.2.1], see also [4].

6. On the number of observations

In this section we determine $\mathbb{E}z^N$ under both hypotheses, where N is the number of observations leading to a decision, see (1). To that end, some further exit theory of MAPs [6] can be used (and the present context provides an interesting illustration of the applicability of the latter). We will also utilize the concept of killing, see e.g. [5].

Suppose we kill our MAP (X_t, J_t) right before every jump $-d$ with probability $1-z$, where $z \in (0,1]$ (i.e., the process is sent to an additional absorbing state). Write $\mathbb{P}^z$ for the corresponding probability measure. Then

$$\mathbb{P}^z(\tau_a^- < \tau_{b+d}^+) = \mathbb{E}(z^N 1_{\{\Lambda_N \leq a\}}),$$

because the process has to survive N independent killing instants. Similarly,

$$z\mathbb{P}^z(\tau_{b+d}^+ < \tau_a^-) = \mathbb{E}(z^N 1_{\{\Lambda_N \geq b\}}),$$

where prefactor z comes from the fact that the MAP should not be killed at the jump following its first passage time over $b+d$. Adding these two equations we obtain $\mathbb{E}z^N$.

Importantly, all exit identities still hold for the killed MAP, which is characterized by $F^z(s) = T + \theta s \mathbb{I} + \boldsymbol{t}\boldsymbol{\nu} z e^{-ds}$. In particular, $\mathbb{P}^z(\tau_{b+d}^+ < \tau_a^-) = \boldsymbol{\nu} W^z(-a) W^z(-a+b+d)^{-1} \mathbf{1}$, where the transform of the scale function $W^z(x)$ evaluates to $F^z(s)^{-1}$. Furthermore, from Corollary 3 in [6] we have

$$\mathbb{P}^z(\tau_a^- < \tau_{b+d}^+) = \boldsymbol{\nu} \left(Z^z(-a) - W^z(-a) W^z(-a+b+d)^{-1} Z^z(-a+b+d) \right) \mathbf{1},$$

where $Z^z(x) = \mathbb{I} - \int_0^x W^z(y)\mathrm{d}yF^z(0)$. Therefore,

$$\mathbb{E}z^N = \boldsymbol{\nu}\left(Z^z(-a) - W^z(-a)W^z(-a+b+d)^{-1}Z^z(-a+b+d) + zW^z(-a)W^z(-a+b+d)^{-1}\right)\mathbf{1}.$$

Noting that $F^z(0)\mathbf{1} = (z-1)\boldsymbol{t}$, differentiating with respect to z and letting $z \uparrow 1$ we get

$$\mathbb{E}N = -\int_0^{-a} \boldsymbol{\nu}W(y)\boldsymbol{t}\mathrm{d}y + \boldsymbol{\nu}W(-a)W(-a+b+d)^{-1}\left(\int_0^{-a+b+d} W(y)\boldsymbol{t}\mathrm{d}y + \mathbf{1}\right), \tag{17}$$

where $W(x)$ corresponds to the case of no killing ($z = 1$). Here we also used differentiability of $W^z(x)$ and $Z^z(x)$ in z, which can be shown using further fluctuation identities. Formula (17) provides both $\mathbb{E}_0N$ and $\mathbb{E}_1N$, where the latter can be expressed through the quantities associated to hypothesis H_0 using Proposition 1 and (10).

7. Variational and Bayesian formulation

7.1. Variational formulation: the optimality region

So far we have focused on the variational formulation of Wald's test. According to this formulation, for given errors α_0 and α_1 one needs to determine the decision boundaries $a < 0 < b$ resulting in these errors. For that purpose one can solve the two equations of Theorem 1 using numerical methods. When such boundaries exist, they are unique and they define the optimal test minimizing both $\mathbb{E}_0N$ and $\mathbb{E}_1N$. The following algorithm can be used to determine the region R of (α_0, α_1) in $[0,1] \times [0,1]$, for which the decision boundaries (resulting in the errors) exist, and hence Wald's test is optimal. This algorithm can be analyzed using monotonicity results from [15]. We omit its thorough discussion.

Algorithm 1. *Determination of the optimality region R:*

1. *Find the errors α_0^* and α_1^* corresponding to $a = b = 0$.*
2. *Fix $b = 0$; for all α_0 in $[0, \alpha_0^*)$ determine a which results in α_0 and then find the corresponding $\alpha_1 > \alpha_1^*$.*
3. *Fix $a = 0$; for all α_1 in $[0, \alpha_1^*)$ determine b which results in α_1 and then find the corresponding $\alpha_0 > \alpha_0^*$.*

These two continuous curves (α_0, α_1), the point (α_0^*, α_1^*), and the axis provide the boundary of the optimality region R.

We provide an example for the optimality region R in Section 8. It indicates that R is large enough to include most cases of practical interest. If the pair of errors lies outside of R, then the problem of testing is 'too easy', i.e. a certain test, which uses zero observations with positive probability, will perform better than any Wald's test with $a < 0 < b$.

7.2. Bayesian formulation

In the Bayesian formulation, it is assumed that H_0 has some prior probability $\pi \in [0,1]$, see e.g. [10]. For fixed constants $c, c_0, c_1 > 0$ one defines a penalty (or average loss)

$$\gamma = \pi(c\mathbb{E}_0N + c_0\alpha_0) + (1-\pi)(c\mathbb{E}_1N + c_1\alpha_1), \tag{18}$$

which is to be minimized. It turns out that there always exists a test which is optimal for all π, i.e. it minimizes the penalty among all tests. The rule is to stop when the posterior probability of H_0 exits some interval (a^*, b^*),

where $0 \leq a^* \leq b^* \leq 1$, with the obvious decision. Expressing the posterior probability through Λ_k, one observes that an equivalent rule is to stop when Λ_k exits

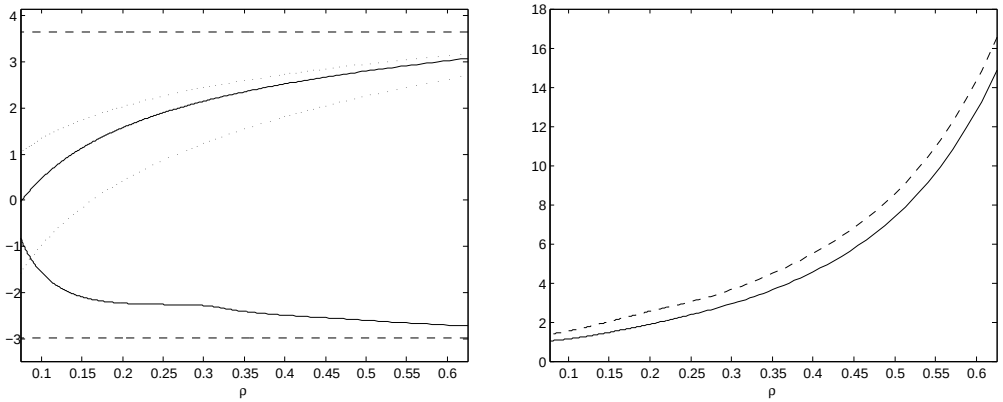
$$(a, b) = (\log \frac{a^*}{1-a^*} + \log \frac{1-\pi}{\pi}, \log \frac{b^*}{1-b^*} + \log \frac{1-\pi}{\pi}), \tag{19}$$

see [10]. Recall that for a given pair (a, b) we can find α_0, α_1 and $\mathbb{E}_0 N, \mathbb{E}_1 N$ using Theorem 1 and (17) respectively, and so we can calculate the penalty γ for a fixed prior π. Hence to find an optimal (a, b), corresponding to the minimal penalty, we only need to run a numerical optimization routine. If this (a, b) is the unique pair minimizing the penalty, then (a^*, b^*) can be recovered from the above relation.

8. Numerical illustrations

In this section we provide an illustration of the applicability of our results for both the variational and Bayesian formulation. For simplicity we choose an Erlang distribution with 2 phases and rate λ, and consider Wald's test of the simple hypothesis $\lambda = \lambda_0$ against the simple alternative $\lambda = \lambda_1$, where $\lambda_0 < \lambda_1$. In Section 5 it was shown that in such a situation Wald's test depends only on the single parameter $\rho = \lambda_0/\lambda_1 \in (0, 1)$ and the scale function $W_0(x)$ has an explicit representation.

Let us first consider the variational formulation. We choose errors $\alpha_0 = 0.05$ and $\alpha_1 = 0.025$, and find the decision boundaries $a < 0 < b$ by solving (13) numerically. Figure 2a depicts a and b as functions of ρ (solid lines), as well as their Wald bounds (2) (dashed lines) and the improved upper and lower bounds for b from (15) (dotted lines). Figure 2b depicts $\max(\mathbb{E}_0 N, \mathbb{E}_1 N)$ for the exact boundaries (solid line) and their Wald bounds (dashed line), respectively.



(a) Decision boundaries and their bounds

(b) $\max(\mathbb{E}_0 N, \mathbb{E}_1 N)$

Fig 2: Decision boundaries and the maximal expected number of observations

Let us briefly comment on the case when ρ is close to 1, i.e. the test problem is very hard. In this case the increments of the random walk Λ_k decrease in absolute value. This implies that Λ_N is very close to a or b (depending on the side of exit), which makes the Wald bounds very tight (see also a discussion in [12]). In

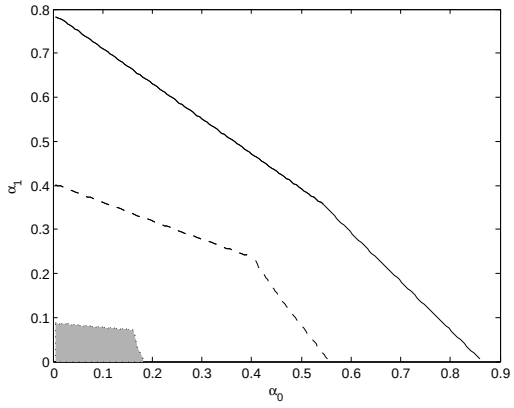


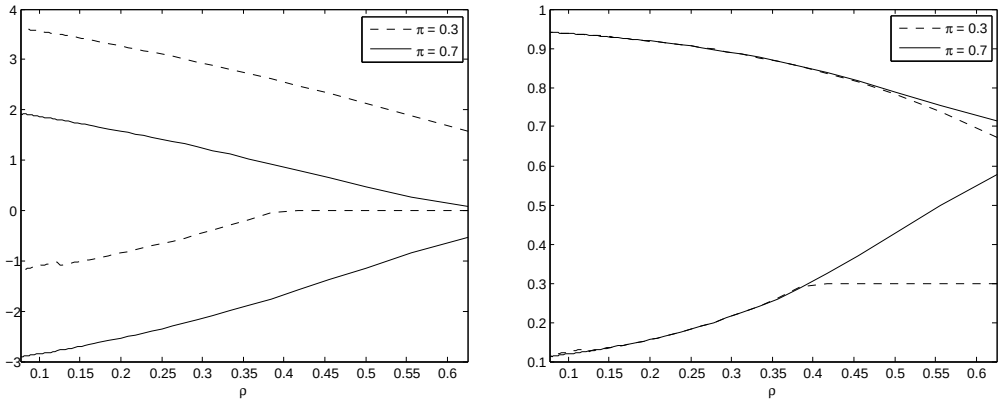
Fig 3: Optimality region R for $\rho = 1/6$ (shaded), $\rho = 1/2$ (below dashed line) and $\rho = 5/6$ (below solid line)

Figures 2a and 2b one can see that the boundaries get indeed closer to their Wald bounds and the expected number of observations increases as $\rho \to 1$. When ρ gets close to 1, also numerical problems arise, as due to small d the number of terms in the representation of $W_0(x)$ becomes large (cf. Theorem 2).

On the other hand, when ρ decreases to 0, the test problem becomes simpler. When one of the boundaries hits 0, the Wald test stops being optimal (cf. Algorithm 1). Figure 3 depicts optimality regions of the Wald test for different values of ρ for the above Erlang(2) example.

Let us turn our attention now to the Bayesian formulation, see Section 7. We choose $c = 0.1, c_0 = 1$ and $c_1 = 2$ for the penalty γ in (18) and two different values $\pi = 0.3$ and $\pi = 0.7$ for the prior. Figure 4a depicts the optimal boundaries a and b (minimizing the penalty). These boundaries are used to compute the optimal boundaries a^* and b^* for the posterior probability by virtue of (19), which can only be done if (a, b) is a unique pair achieving the minimal penalty. The result is depicted in Figure 4b. Recall that the latter boundaries do not depend on the prior π, and hence the lines for both π should coincide. This is indeed the case up to $\rho \approx 0.38$, at which point a (corresponding to $\pi = 0.3$) hits level 0 and uniqueness is lost (in this case, b can be any positive number). The correct values of a^* and b^* follow the solid lines corresponding to $\pi = 0.7$.

Note that the behavior of the boundaries a and b as functions of ρ is substantially different for the variational and the Bayesian formulation. For increasing ρ, the distance between the decision boundaries increases in the former case and decreases in the latter, where controlling the number of observations becomes the dominant issue.



(a) Decision boundaries a and b for Λ_k (b) Boundaries a^* and b^* for the posterior

Fig 4: Decision boundaries for the Bayesian formulation

9. Acknowledgments

We would like to thank Onno Boxma, Dominik Kortschak, Andreas Löpker and Jef Teugels for their valuable comments. This work was supported by the Swiss National Science Foundation Project 200020-143889 and the EU-FP7 project Impact2C.

References

[1] S. Asmussen. Exponential families generated by phase-type distributions and other Markov lifetimes. *Scand. J. Statist.*, 16(4):319–334, 1989.

[2] S. Asmussen and H. Albrecher. *Ruin probabilities*. Advanced Series on Statistical Science & Applied Probability, 14. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, second edition, 2010.

[3] G. J. Franx. A simple solution for the $M/D/c$ waiting time distribution. *Oper. Res. Lett.*, 29(5):221–229, 2001.

[4] H. U. Gerber. Mathematical fun with ruin theory. *Insurance Math. Econom.*, 7(1):15–23, 1988.

[5] J. Ivanovs. A note on killing with applications in risk theory. *Insurance: Mathematics and Economics*, 52(1):29–33, 2013.

[6] J. Ivanovs and Z. Palmowski. Occupation densities in solving exit problems for Markov additive processes and their reflections. *Stochastic Process. Appl.*, 122(9):3342–3360, 2012.

[7] A. Kyprianou. *Introductory lectures on fluctuations of Lévy processes with applications*. Universitext. Springer-Verlag, Berlin, 2006.

[8] V. I. Lotov. Asymptotic expansions for a sequential likelihood ratio test. *Teor. Veroyatnost. i Primenen.*, 32(1):62–72, 1987.

[9] C.-O. Segerdahl. A survey of results in the collective theory of risk. In *Probability and statistics: The Harald Cramér volume*, pages 276–299. Almqvist & Wiksell, Stockholm, 1959.

[10] A. N. Shiryaev. *Statistical sequential analysis*. American Mathematical Society, 1973.

[11] J. L. Teugels and W. Van Assche. Sequential testing for exponential and Pareto distributions. *Sequential Anal.*, 5(3):223–236, 1986.
[12] A. Wald. *Sequential Analysis*. John Wiley & Sons Inc., 1947.
[13] A. Wald and J. Wolfowitz. Optimum character of the sequential probability ratio test. *Ann. Math. Statistics*, 19:326–339, 1948.
[14] L. Weiss. On the uniqueness of Wald sequential tests. *Ann. Math. Statist.*, 27:1178–1181, 1956.
[15] R. Wijsman. A monotonicity property of the sequential probability ratio test. *Ann. Math. Statist.*, 31:677–684, 1960.
[16] R. A. Wijsman. Existence, uniqueness and monotonicity of sequential probability ratio tests. *Ann. Math. Statist.*, 34:1541–1548, 1963.