

# CELEBRATING 50 YEARS OF THE APPLIED PROBABILITY TRUST

Edited by  
S. ASMUSSEN, P. JAGERS, I. MOLCHANOV and L. C. G. ROGERS

Part 8. Markov processes and renewal theory

## EXACT BOUNDARIES IN SEQUENTIAL TESTING FOR PHASE-TYPE DISTRIBUTIONS

HANSJÖRG ALBRECHER, *University of Lausanne and Swiss Finance Institute*  
Department of Actuarial Science, Faculty of Business and Economics, University of Lausanne,  
Quartier UNIL-Dorigny, 1015 Lausanne, Switzerland. Email address: hansjoerg.albrecher@unil.ch

PEIMAN ASADI, *University of Lausanne*  
Department of Actuarial Science, Faculty of Business and Economics, University of Lausanne,  
Quartier UNIL-Dorigny, 1015 Lausanne, Switzerland. Email address: peiman.asadi@unil.ch

JEVGENIJS IVANOV, *University of Lausanne*  
Department of Actuarial Science, Faculty of Business and Economics, University of Lausanne,  
Quartier UNIL-Dorigny, 1015 Lausanne, Switzerland. Email address: jevgenijs.ivanovs@unil.ch



APPLIED PROBABILITY TRUST  
DECEMBER 2014

# EXACT BOUNDARIES IN SEQUENTIAL TESTING FOR PHASE-TYPE DISTRIBUTIONS

BY HANSJÖRG ALBRECHER, PEIMAN ASADI AND JEVGENIJS IVANOVŠ

---

## Abstract

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. Exact decision boundaries for given type-I and type-II errors are derived 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 with 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 tails.

*Keywords:* Sequential probability ratio test; Markov additive process; scale function; two-sided exit problem; Esscher transform

2010 Mathematics Subject Classification: Primary 62L12  
Secondary 91B30

---

## 1. Introduction

Consider Wald's sequential probability ratio test [12] of a simple hypothesis  $H_0$  against a simple alternative  $H_1$ . 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  $\Lambda_n$  for the first  $n$  observations is then given by

$$\Lambda_n = \sum_{i=1}^n \log \frac{f_0(\zeta_i)}{f_1(\zeta_i)}, \quad \Lambda_0 = 0,$$

and its first exit time  $N$  from the interval  $(a, b)$  for  $a < 0 < b$  by

$$N = \inf\{n \geq 0 : \Lambda_n \notin (a, b)\}. \tag{1}$$

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 we can 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 [10, Theorem IV.4] and [13]. Such a pair  $(a, b)$  of boundaries is unique under very weak assumptions [14] that 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; see [16] for an analysis of a more general test.

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

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

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, though usually not by a large amount for small errors; see [12].

If the errors  $\alpha_0$  and  $\alpha_1$  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  $\alpha_0, \alpha_1$  of interest. However, this inverse problem is hard even for simple cases. Some tractable examples can be found in [12] and [11], where in the latter work  $f_0$  and  $f_1$  are assumed 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  $\Lambda_n$ ; this can be done explicitly only in some cases.

In this paper we assume that  $f_0$  and  $f_1$  are densities of phase-type (PH) distributions such that  $f_1$  can be obtained by exponential tilting of  $f_0$ . 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 that leads to explicit formulae in Section 4. This approach simplifies the proof for the exponential case developed in [11] and extends the results to PH densities; by taking monotone transformations of the original observations, the results are also applicable to other distributions, such as distributions with regularly varying tails obtained by exponentiating PH random variables. In Section 5 we discuss the Erlang case in more detail; for this case, an explicit treatment is possible. In Section 6 we provide a general formula for the expected number of observations in Wald’s test. In Section 7 we study the uniqueness issue further and consider an extension to a Bayesian version for which an *a priori* probability for the validity of  $H_0$  is available. Finally, in Section 8 we provide some numerical illustrations.

## 2. Wald’s test and ruin theory

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

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

In actuarial contexts,  $f_1$  is called the Esscher transform of  $f_0$ .

For  $\theta > 0$ , consider Wald’s test for such densities  $f_0$  and  $f_1$ , and observe that

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

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

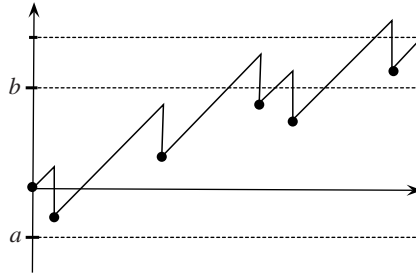


FIGURE 1: The Sparre Andersen risk process.

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 interarrival times distributed as  $\zeta$ ; see Figure 1.

We can interpret  $X_t$  as the surplus process of an insurance portfolio under a Sparre Andersen risk model for which the initial capital is 0, the premiums are collected at constant rate  $\theta$ , and claims of (deterministic) size  $d$  arrive according to the renewal process  $N_t$  (see, e.g. [2, Chapter VI]). Importantly, one can recover the random walk  $\Lambda_n$  from the continuous-time process  $X_t$  by considering it at the jump epochs. Letting

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

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. We have thus arrived at a pair of two-sided exit problems for the risk process  $X_t$ , one under  $H_0$  and the other 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  $\zeta$  has a PH distribution, i.e. the distribution of the lifetime of a transient continuous-time Markov chain (MC) on finitely many states  $1, \dots, k$ ; see, e.g. [2, Chapter IX.1]. A PH distribution is parameterized by the transition rate matrix  $T$  of the corresponding MC and the row vector  $\mathbf{v}$  representing the initial distribution. Introduce the column vector  $\mathbf{t} = -T\mathbf{1}$  to denote the killing (absorption) rates, where  $\mathbf{1} = (1, 1, \dots, 1)^\top$ . Then the density of  $\zeta$  can be expressed as

$$f(x) = \mathbf{v}e^{T x} \mathbf{t}. \tag{6}$$

The Erlang distribution of rate  $\lambda$  is recovered by setting  $\mathbf{v} = (1, 0, \dots, 0)$  and choosing  $T$  to be a square matrix with  $-\lambda$  on the main diagonal,  $\lambda$  on the upper diagonal, and 0 elsewhere. Recall 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 a PH distribution. It is not hard to see that  $J_t$  is an MC with transition rate matrix  $T + \mathbf{t} \mathbf{v}$ , i.e. transitions occur

because of either a phase change or the 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, Chapter VII]. Such a process is characterized by a matrix-valued function  $F(s)$ ,  $s \geq 0$ , satisfying  $\mathbb{E}(e^{sX_t} 1_{\{J_t=j\}} \mid J_0 = i) = (e^{F(s)t})_{ij}$  for all  $t \geq 0$  and  $i, j \in \{1, \dots, k\}$ . In our case we have the identity

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

where  $\mathbf{I}$  is the identity matrix. The diagonal elements  $\theta s$  represent the linear evolution of  $X_t$  with slope  $\theta$  (the same value in every phase) and  $t \mathbf{v}$  is a matrix of transition rates of  $J_t$  causing the jump in  $X_t$  with transform  $e^{-ds}$ ; see [2, Proposition 4.2].

The two-sided exit problem for MAPs without positive jumps was solved in [6]; the solution resembles that for a Lévy process without positive jumps [7, Theorem 8.1]. According to Ivanovs and Palmowski [6], the matrix of probabilities with  $ij$ th element  $\mathbb{P}(\tau_b^+ < \tau_a^-, J_{\tau_b^+} = j \mid J_0 = i)$  is given by  $W(-a)W(-a + b)^{-1}$ , where  $W(x)$ ,  $x \geq 0$ , is a continuous, matrix-valued function (called a *scale function*) characterized by the transform

$$\int_0^\infty e^{-sx} W(x) dx = F(s)^{-1} \quad \text{for sufficiently large } s. \tag{8}$$

It is known that  $W(x)$  is nonsingular for  $x > 0$  and so is  $F(s)$  in the domain of interest. Since  $J_0$  has distribution  $\mathbf{v}$ , we can write

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

Note that the scale function is given in terms of its transform; 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, when  $\zeta$  has an Erlang distribution as in Section 5, we obtain an explicit representation of  $W(x)$ .

#### 4. Identification of the errors

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

$$T_1 = \Delta^{-1} T_0 \Delta - \theta \mathbf{I}, \quad \mathbf{v}_1 = \frac{\mathbf{v}_0 \Delta}{G_0(\theta)}, \quad \mathbf{t}_1 = \Delta^{-1} \mathbf{t}_0, \tag{10}$$

where  $\Delta = \text{diag}((\theta \mathbf{I} - T_0)^{-1} \mathbf{t}_0)$ . All these diagonal elements lie in  $(0, 1)$ , as can be seen from the representation of  $G(\theta) \in (0, 1)$  for different initial distributions  $\mathbf{v}_0$ .

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

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

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. The fact that  $W_0$  and  $W_1$  are intimately related is the substance of our next result.

**Proposition 1.** For all  $x \geq 0$ , the scale functions  $W_0(x)$  and  $W_1(x)$  satisfy

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

*Proof.* We show that

$$\begin{aligned} \Delta^{-1} F_0(s - 1) \Delta &= \Delta^{-1} \left( T_0 - \theta \mathbf{I} + \theta s \mathbf{I} + \frac{\mathbf{t}_0 \mathbf{v}_0}{G_0(\theta)} e^{-ds} \right) \Delta \\ &= T_1 + \theta s \mathbf{I} + \mathbf{t}_1 \mathbf{v}_1 e^{-ds} \\ &= F_1(s) \end{aligned}$$

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

$$\begin{aligned} \int_0^\infty e^{-sx} (e^x \Delta^{-1} W_0(x) \Delta) dx &= \Delta^{-1} \int_0^\infty e^{-(s-1)x} W_0(x) dx \Delta \\ &= \Delta^{-1} F_0(s - 1)^{-1} \Delta \\ &= F_1(s)^{-1} \quad \text{for sufficiently large } s. \end{aligned}$$

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 yields the following result.

**Theorem 1.** Let  $f_0$  be a density of a PH distribution with parameters  $T_0$ ,  $\mathbf{v}_0$ , and  $\mathbf{t}_0$ , and let  $f_1$  be the corresponding exponentially tilted density with tilt parameter  $\theta > 0$ . The errors  $\alpha_0$  and  $\alpha_1$  corresponding to the decision boundaries  $a < 0 < b$  in the Wald test of  $f_0$  against  $f_1$  are given by

$$\begin{aligned} \alpha_0 &= 1 - \mathbf{v}_0 W_0(-a) W_0(-a + b + d)^{-1} \mathbf{1}, \\ \alpha_1 &= e^{-b} \mathbf{v}_0 W_0(-a) W_0(-a + b + d)^{-1} (\theta \mathbf{I} - T_0)^{-1} \mathbf{t}_0, \end{aligned}$$

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

$$\int_0^\infty e^{-sx} W_0(x) dx = (T_0 + \theta s \mathbf{I} + \mathbf{t}_0 \mathbf{v}_0 e^{-ds})^{-1} \quad \text{for sufficiently large } s.$$

The transform of  $W_0(x)$  can be inverted in certain cases. In Section 5 we provide an explicit expression for  $W_0(x)$  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  $\mathbf{v}_0 W_0(-a) W_0(-a + b + d)^{-1}$  is a vector of probabilities, and recall that all the entries of  $(\theta \mathbf{I} - T_0)^{-1} \mathbf{t}_0$  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 maximal elements of  $(\theta \mathbf{I} - T_0)^{-1} \mathbf{t}_0$ . Hence,

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

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

**Example 1.** If  $f_0$  is the density of an exponential distribution with rate  $\lambda_0$  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  $-\lambda_0$ , so  $m = M = \lambda_0/(\lambda_0 + \theta) = \lambda_0/\lambda_1$  and  $b = \log[(1 - \alpha_0)/\alpha_1] - \log[\lambda_1/\lambda_0]$ . This simple identity for exponential densities has already been established in [11]. Computation of the boundary  $a < 0$  is more involved, and relies on the identity

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

where  $W_0(\cdot)$  is 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 in which  $f_0$  is the density of an Erlang distribution with  $k$  phases and rate  $\lambda_0$ , i.e.  $f_0(x) = \lambda_0^k x^{k-1} e^{-\lambda_0 x} / (k - 1)!$ , which has Laplace transform  $G_0(\theta) = (\lambda_0/(\lambda_0 + \theta))^k$ . Exponential tilting of  $f_0$  with tilt parameter  $\theta > 0$  results in  $f_1$ , which is another Erlang density on  $k$  phases but with rate  $\lambda_1 = \lambda_0 + \theta$ . Hence, our setup allows us to consider any two Erlang distributions with the same number of phases.

Under the Erlang assumption, the jump size  $d = -k \log(\lambda_0/\lambda_1)$  depends only on the ratio  $\rho = \lambda_0/\lambda_1$  of the two rates and not on their absolute values. Also, since  $\theta \cdot \text{Erlang}(k, \lambda_i) \sim \text{Erlang}(k, \lambda_i/\theta)$ , scaling  $\theta$  to 1 simply scales the process  $X_t$  of (4) in the horizontal direction by the factor  $\theta$  (under both hypotheses) and the law of the random walk  $\Lambda_n$  is unchanged. Hence, Wald's test depends only on the ratio  $\rho$ , and without loss of generality 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  $T_0$ ,  $\mathbf{v}_0$ , and  $\mathbf{t}_0$  of the density  $f_0$ , where  $T_0$  is a  $k \times k$  matrix with  $-\lambda_0$  on the diagonal,  $\lambda_0$  on the upper diagonal, and 0 elsewhere;  $\mathbf{v}_0 = (1, 0, \dots, 0) = \mathbf{e}_1$  and  $\mathbf{t}_0 = (0, \dots, 0, \lambda_0)^\top$ . Algebraic manipulation shows that the vector  $(\theta \mathbf{I} - T_0)^{-1} \mathbf{t}_0$  simplifies to  $(\rho^k, \rho^{k-1}, \dots, \rho^1)^\top$ , and so by Theorem 1 we have

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

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

$$F_0(s) = \begin{pmatrix} s - \lambda_0 & \lambda_0 & 0 & \dots & 0 \\ 0 & s - \lambda_0 & \lambda_0 & \ddots & 0 \\ 0 & 0 & s - \lambda_0 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \lambda_0 e^{-sd} & 0 & 0 & \dots & s - \lambda_0 \end{pmatrix} \text{ for } k \geq 2 \tag{14}$$

and  $F_0(s) = s - \lambda_0 + \lambda_0 e^{-sd}$  for  $k = 1$ . The bounds in (12) for  $b$  now simplify to

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

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  $k$  phases for which  $F_0(s)$  is as in (14) for any  $d > 0$ . Then, for  $x \geq 0$ , the  $ij$ th element of the scale function  $W_0(x)$  is given by

$$W_0(x)_{ij} = \sum_{h=1}^{\lfloor x/d \rfloor} g(\lambda_0(x - dh), hk + j - i), \quad i, j = 1, \dots, k, \tag{16}$$

where  $g(y, m) = (-y)^m e^y / m!$ .

*Proof.* We omit the subscript 0 in this proof. We need to invert the transform

$$\int_0^\infty e^{-sx} W(x) dx = F(s)^{-1}.$$

Application of Cramér’s rule and careful computation of co-factors yields

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

where  $l = (j - i) \bmod k$ . Noting that we can write the fraction on the right-hand side as  $(s - \lambda)^{-k} \sum_{h=0}^\infty (-\lambda)^{hk} (s - \lambda)^{-hk} e^{-sdh}$  for sufficiently large  $s$ , we have

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

Using

$$\int_0^\infty e^{-sx} \frac{x^k}{k!} e^{\lambda x} dx = \frac{1}{(s - \lambda)^{k+1}},$$

we invert  $(-\lambda)^{hk+l} / (s - \lambda)^{hk+l+1}$  to obtain  $(-\lambda x)^{hk+l} e^{\lambda x} / (hk + l)! = g(\lambda x, hk + l)$ . The factor  $e^{-sd(h+1_{\{i>j\}})}$  amounts to shifting  $x$  to  $x - d(h + 1_{\{i>j\}})$ . Hence, for  $j \geq i$ ,

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

Similarly, for  $j < i$ ,  $W(x)_{ij}$  equals

$$\sum_{h=0}^\infty g(\lambda(x - d(h + 1)), hk + k + j - i) 1_{\{x \geq d(h+1)\}} = \sum_{h=1}^{\lfloor x/d \rfloor} g(\lambda(x - dh), hk + j - i),$$

which concludes the proof.

Quantity (16) is intimately connected to the waiting time distribution in an  $E_k/D/1$  queue; see, for instance, [3]. In a risk theory context, for the  $k = 1$  case (which refers to the compound Poisson model with deterministic jumps), (16) can be found in [9, Section 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 this end, some further exit theory of MAPs [6] can be used (and the present context provides an illustrative application of this theory). We also utilize the concept of killing; see, e.g. [5].

Suppose that immediately before every jump  $-d$  and independently for such epochs we kill our MAP  $(X_t, J_t)$  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

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

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

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

where the prefactor  $z$  occurs because the MAP should not be killed at the jump following its first passage epoch over  $b + d$ . Adding these two equations yields  $\mathbb{E}z^N$ .

Importantly, all exit identities still hold for the killed MAP, which is characterized by  $F^z(s) = T + \theta s \mathbf{I} + t \nu z e^{-ds}$ . In particular,

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

where the transform of the scale function  $W^z(x)$  evaluates to  $F^z(s)^{-1}$ . Furthermore, from Corollary 3 of [6] we have

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

where  $Z^z(x) = \mathbf{I} - \int_0^x W^z(y) dy F^z(0)$ . Therefore,

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

Since  $F^z(0)\mathbf{1} = (z - 1)t$ , differentiating with respect to  $z$  and letting  $z \uparrow 1$  gives

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

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. Equation (17) provides both  $\mathbb{E}_0N$  and  $\mathbb{E}_1N$ , where the latter can be expressed through the quantities associated with hypothesis  $H_0$  using Proposition 1 and (10).

### 7. Variational and Bayesian formulations

#### 7.1. Variational formulation: the optimality region

So far, we have considered only the variational formulation of Wald’s test, according to which it suffices to determine the decision boundaries  $a < 0 < b$  that result in given errors  $\alpha_0$  and  $\alpha_1$ . For that purpose, one can solve the two equations of Theorem 1 using numerical methods. When such boundaries exist, they are unique and 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  $(\alpha_0, \alpha_1)$  in  $[0, 1] \times [0, 1]$ , for which there exist decision boundaries resulting in such errors, 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  $R$ .)

- (i) Find the errors  $\alpha_0^*$  and  $\alpha_1^*$  corresponding to  $a = b = 0$ .
- (ii) Fix  $b = 0$ ; for all  $\alpha_0$  in  $[0, \alpha_0^*)$ , determine  $a$  which results in  $\alpha_0$  and then find the corresponding  $\alpha_1 > \alpha_1^*$ .
- (iii) Fix  $a = 0$ ; for all  $\alpha_1$  in  $[0, \alpha_1^*)$ , determine  $b$  which results in  $\alpha_1$  and then find the corresponding  $\alpha_0 > \alpha_0^*$ .

These two continuous curves  $(\alpha_0, \alpha_1)$ , the point  $(\alpha_0^*, \alpha_1^*)$ , and the axes provide the boundary of the optimality region  $R$ .

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

**7.2. Bayesian formulation**

For a Bayesian formulation, assume that  $H_0$  has some prior probability  $\pi \in [0, 1]$ ; see, e.g. [10]. For fixed constants  $c, c_0, c_1 > 0$ , define a penalty (or average loss)

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

this loss is to be minimized. We assert that there always exists a test which is optimal for all  $\pi$ , i.e. it minimizes the penalty among all tests; namely, stop testing 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  $\Lambda_n$ , observe that an equivalent rule is to stop when  $\Lambda_n$  exits

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

see [10]. Recall that, for a given pair  $(a, b)$ , we can find  $\alpha_0, \alpha_1$  and  $\mathbb{E}_0N, \mathbb{E}_1N$  using Theorem 1 and (17), respectively, so we can calculate the penalty  $\gamma$  for a fixed prior  $\pi$ . Then 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 (19).

**8. Numerical illustrations**

In this section we illustrate our results for both the variational and Bayesian formulations. For simplicity, we choose an Erlang distribution with two phases and rate  $\lambda$ , 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 we showed that in such a situation Wald’s test depends only on the single parameter  $\rho = \lambda_0/\lambda_1 \in (0, 1)$  and that the scale function  $W_0(x)$  has an explicit representation.

Consider first the variational formulation. Choose errors  $\alpha_0 = 0.05$  and  $\alpha_1 = 0.025$ , and find the decision boundaries  $a < 0 < b$  by solving (13) numerically. Figure 2(a) depicts  $a$  and  $b$  as functions of  $\rho$  (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 2(b) depicts  $\max(\mathbb{E}_0N, \mathbb{E}_1N)$  for the exact boundaries (solid line) and their Wald bounds (dashed line).

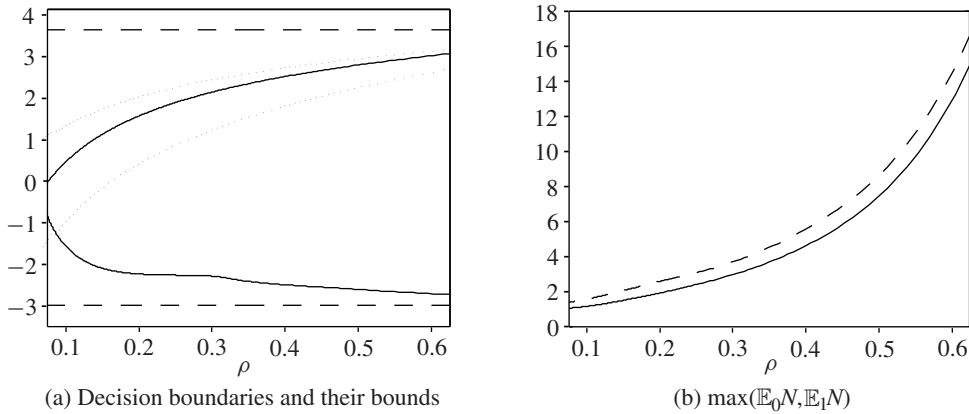


FIGURE 2: Decision boundaries and the maximal expected number of observations.

We comment briefly on the case when  $\rho$  is close to 1, i.e. the test problem is very hard. In this case the increments of the random walk  $\Lambda_n$  decrease in absolute value. This implies that  $\Lambda_N$  is very close to  $a$  or  $b$  (depending on the side of exit), which makes the Wald bounds very tight (see also the discussion in [12]). In Figure 2(a) and (b) the boundaries do indeed come closer to their Wald bounds and the expected number of observations increases as  $\rho \rightarrow 1$ . When  $\rho$  gets close to 1, numerical problems also arise, as, owing to small  $d$ , the number of terms in the representation of  $W_0(x)$  becomes large (cf. Theorem 2).

On the other hand, when  $\rho$  decreases to 0, the test problem becomes simpler. When one of the boundaries hits 0, the Wald test is no longer optimal (cf. Algorithm 1). Figure 3 depicts optimality regions of the Wald test for different values of  $\rho$  for the above Erlang(2) illustrative example.

Consider now the Bayesian formulation as in Section 7. Choose  $c = 0.1$ ,  $c_0 = 1$ , and  $c_1 = 2$  for the penalty  $\gamma$  in (18), and two different values  $\pi = 0.3$  and  $\pi = 0.7$  for the prior. Figure 4(a) depicts the optimal boundaries  $a$  and  $b$  that minimize the penalty. We use these boundaries in (19) to compute the optimal boundaries  $a^*$  and  $b^*$  for the posterior probability

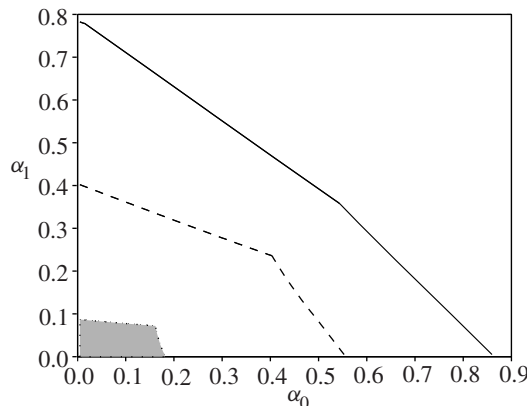


FIGURE 3: Optimality region  $R$  for  $\rho = \frac{1}{6}$  (shaded region),  $\rho = \frac{1}{2}$  (below dashed line), and  $\rho = \frac{5}{6}$  (below solid line).

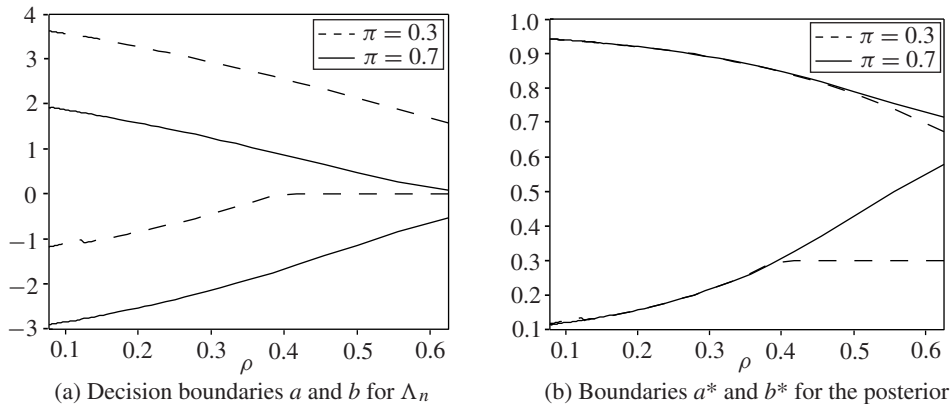


FIGURE 4: Decision boundaries for the Bayesian formulation

(this can be done only if  $(a, b)$  is a unique pair achieving the minimal penalty). The result is depicted in Figure 4(b). Recall that the latter boundaries do not depend on the prior  $\pi$ , and hence, the lines for both  $\pi$  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 (and then  $b$  can be any positive number). The correct values of  $a^*$  and  $b^*$  follow the solid lines corresponding to  $\pi = 0.7$ .

Note that the behaviour of the boundaries  $a$  and  $b$  as functions of  $\rho$  differs substantially between the variational and Bayesian formulations. For increasing  $\rho$ , the distance between the decision boundaries increases in the former case and decreases in the latter for which controlling the number of observations becomes the dominant issue.

### Acknowledgements

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] ASMUSSEN, S. (1989). Exponential families generated by phase-type distributions and other Markov lifetimes. *Scand. J. Statist.* **16**, 319–334.
- [2] ASMUSSEN, S. AND ALBRECHER, H. (2010). *Ruin Probabilities* (Adv. Ser. Statist. Sci. Appl. Prob. **14**), 2nd edn. World Scientific, Hackensack, NJ.
- [3] FRANX, G. J. (2001). A simple solution for the M/D/c waiting time distribution. *Operat. Res. Lett.* **29**, 221–229.
- [4] GERBER, H. U. (1988). Mathematical fun with ruin theory. *Insurance Math. Econom.* **7**, 15–23.
- [5] IVANOV, J. (2013). A note on killing with applications in risk theory. *Insurance Math. Econom.* **52**, 29–34.
- [6] IVANOV, J. AND PALMOWSKI, Z. (2012). Occupation densities in solving exit problems for Markov additive processes and their reflections. *Stoch. Process. Appl.* **122**, 3342–3360.
- [7] KYPRIANOU, A. E. (2006). *Introductory Lectures on Fluctuations of Lévy Processes with Applications*. Springer, Berlin.
- [8] LOTOV, V. I. (1987). Asymptotic expansions for a sequential likelihood ratio test. *Teor. Veroyat. i Primen.* **32**, 62–72.
- [9] SEGERDAHL, C.-O. (1959). A survey of results in the collective theory of risk. In *Probability and Statistics: The Harald Cramér Volume*, Almqvist & Wiksell, Stockholm, pp. 276–299.
- [10] SHIRYAEV, A. N. (1973). *Statistical Sequential Analysis*. American Mathematical Society.
- [11] TEUGELS, J. L. AND VAN ASSCHE, W. (1986). Sequential testing for exponential and Pareto distributions. *Sequent. Anal.* **5**, 223–236.

- [12] WALD, A. (1947). *Sequential Analysis*. John Wiley, New York.
- [13] WALD, A. AND WOLFOWITZ, J. (1948). Optimum character of the sequential probability ratio test. *Ann. Math. Statist.* **19**, 326–339.
- [14] WEISS, L. (1956). On the uniqueness of Wald sequential tests. *Ann. Math. Statist.* **27**, 1178–1181.
- [15] WIJSMAN, R. A. (1960). A monotonicity property of the sequential probability ratio test. *Ann. Math. Statist.* **31**, 677–684.
- [16] WIJSMAN, R. A. (1963). Existence, uniqueness and monotonicity of sequential probability ratio tests. *Ann. Math. Statist.* **34**, 1541–1548.

HANSJÖRG ALBRECHER, *University of Lausanne and Swiss Finance Institute*

Department of Actuarial Science, Faculty of Business and Economics, University of Lausanne, Quartier UNIL-Dorigny, 1015 Lausanne, Switzerland. Email address: hansjoerg.albrecher@unil.ch

PEIMAN ASADI, *University of Lausanne*

Department of Actuarial Science, Faculty of Business and Economics, University of Lausanne, Quartier UNIL-Dorigny, 1015 Lausanne, Switzerland. Email address: peiman.asadi@unil.ch

JEVGENIJS IVANOV, *University of Lausanne*

Department of Actuarial Science, Faculty of Business and Economics, University of Lausanne, Quartier UNIL-Dorigny, 1015 Lausanne, Switzerland. Email address: jevgenijs.ivanovs@unil.ch