Ann. Statist. Vol. 28, No. 3, 2000, (837-859)
Research Report No. 400, 1998, *Dept. Theoret. Statist. Aarhus*

Sequential Testing Problems for Poisson Processes

G. PESKIR* *and* A. N. SHIRYAEV*

We present the explicit solution of the Bayesian problem of sequential testing of two simple hypotheses about the intensity of an observed Poisson process. The method of proof consists of reducing the initial problem to a free-boundary differential-difference Stephan problem, and solving the latter by use of the principles of smooth and continuous fit. A rigorous proof of the optimality of the Wald's sequential probability ratio test in the variational formulation of the problem is obtained as a consequence of the solution of the Bayesian problem.

1. Description of the problem

Suppose that at time $t=0$ we begin to observe a Poisson process $X=(X_t)_{t\geq 0}$ with intensity $\lambda>0$ which is either λ_0 or λ_1 where $\lambda_0<\lambda_1$. Assuming that the true value of λ is not known to us, our problem is then to decide as soon as possible and with a minimal error probability (both specified later) if the true value of λ is either λ_0 or λ_1.

Depending on the hypotheses about the unknown intensity λ, this problem admits two formulations. The *Bayesian* formulation relies upon the hypothesis that an a priori probability distribution of λ is given to us, and that λ takes either of the values λ_0 and λ_1 at time $t=0$ according to this distribution. The *variational* formulation (sometimes also called a *fixed error probability* formulation) involves no probabilistic assumptions on the unknown intensity λ. The Wald sequential probability ratio test (SPRT) is known to be optimal in this context for a large class of observable processes (see [5], [6], [2]).

Despite the fact that the Bayesian approach to sequential analysis of problems on testing two statistical hypotheses has gained a considerable interest in the last fifty or so years (see e.g. [15], [16], [3], [8], [4], [13], [14]), it turns out that not many problems of that type have been solved explicitly (by obtaining a solution in closed form). In this respect the case of testing two simple hypotheses about the mean value of a Wiener process with drift is exceptional as the explicit solution to the problem has been obtained in both Bayesian and variational formulation. These solutions (including the proof of the optimality of the SPRT) were found by reducing the initial problem to a free-boundary Stephan problem (for a second order differential operator) which could be solved explicitly (see [12], [13]).

Our main aim in this paper is to present the *explicit solution* of the Poisson intensity problem stated above in the context of a *Bayesian formulation* (Section 2), and then apply this result to deduce the optimality of the method (SPRT) in the context of a variational formulation (Section 3)

*Centre for Mathematical Physics and Stochastics, supported by the Danish National Research Foundation.
MR 1991 Mathematics Subject Classification. Primary 62C10, 60G40, 62L10. Secondary 62L15, 34K10, 60J75.
Key words and phrases: Sequential testing, Bayes decision rule, Poisson process, SPRT (sequential probability ratio test), optimal stopping, a free-boundary differential-difference Stephan problem, the principles of continuous and smooth fit, point (counting) (Cox) process, measure of jumps and its compensator, Itô's formula. © goran@imf.au.dk

with a precise description of the set of all admissible probabilities of a wrong decision ("errors of the first and second kind"). It will be clear from the sequel that the corresponding Stephan problem becomes more delicate, since in the present case one needs to deal with a differential-difference operator, the appearance of which is a consequence of the discontinuous character of the observed (Poisson) process. The problem solved in Section 2 has been open for some time. (In the 1984 paper [6] the authors write that "in the case of Poisson processes, an explicit solution [of the Bayesian and Stephan problem] is not known".)

2. Solution of the Bayesian problem

1. In the Bayesian formulation of the problem (see [13] Ch. 4) it is assumed that at time $t = 0$ we begin observing a trajectory of the *point* process $X = (X_t)_{t \geq 0}$ with the compensator (see [9] Ch. 18) $A = (A_t)_{t \geq 0}$, where $A_t = \lambda t$ and a random intensity $\lambda = \lambda(\omega)$ takes two values λ_1 and λ_0 with probabilities π and $1 - \pi$. (We assume that $\lambda_1 > \lambda_0 > 0$ and $\pi \in [0, 1]$.)

For a precise probability-statistical description of the Bayesian sequential testing problem it is convenient to assume that all our considerations take place on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P_\pi)$, where P_π has the following special structure:

$$P_\pi = \pi P_1 + (1 - \pi) P_0 \tag{2.1}$$

for $\pi \in [0, 1]$. We further assume that the $\mathcal{F}_0$-measurable random variable $\lambda = \lambda(\omega)$ takes two values λ_1 and λ_0 with probabilities $P_\pi(\lambda = \lambda_1) = \pi$ and $P_\pi(\lambda = \lambda_0) = 1 - \pi$. Concerning the observable point process $X = (X_t)_{t \geq 0}$, we assume that $P_\pi(X \in \cdot \mid \lambda = \lambda_i) = P_i(X \in \cdot\ )$, where $P_i(X \in \cdot\ )$ coincides with the distribution of a Poisson process with intensity λ_i for $i = 0, 1$.

Probabilities π and $1 - \pi$ play a role of *a priori* probabilities of the statistical hypotheses:

$$H_1 : \lambda = \lambda_1 \tag{2.2}$$

$$H_0 : \lambda = \lambda_0 . \tag{2.3}$$

2. Based upon the information which is continuously updated through the observation of the point process X , our problem is to test sequentially the hypotheses H_1 and H_0 . For this it is assumed that we have at disposal a class of sequential *decision rules* (τ, d) consisting of *stopping times* $\tau = \tau(\omega)$ with respect to $(\mathcal{F}_t^X)_{t \geq 0}$ where $\mathcal{F}_t^X = \sigma\{X_s \mid s \leq t\}$, and $\mathcal{F}_\tau^X$-measurable functions $d = d(\omega)$ which take values 0 and 1 . Stopping the observation of X at time τ , the *terminal decision function* d indicates that either the hypothesis H_1 or the hypothesis H_0 should be accepted; if $d = 1$ we accept H_1 , and if $d = 0$ we accept that H_0 is true.

3. Each decision rule (τ, d) implies losses of two kinds: the loss due to a cost of the observation, and the loss due to a wrong terminal decision. The average loss of the first kind may be naturally identified with $c E_\pi(\tau)$, and the average loss of the second kind can be expressed as $a\, P_\pi(d = 0, \lambda = \lambda_1) + b\, P_\pi(d = 1, \lambda = \lambda_0)$, where $c, a, b > 0$ are some constants. It will be clear from (2.8) below that there is no restriction to assume that $c = 1$, as the case of general $c > 0$ follows by replacing a and b with a/c and b/c respectively. Thus, the *total average loss* of the decision rule (τ, d) is given by

$$L_\pi(\tau,d) = E_\pi\Big(\, \tau + a\, 1_{(d=0\,,\lambda=\lambda_1)} + b\, 1_{(d=1\,,\lambda=\lambda_0)} \Big)\ . \tag{2.4}$$

Our problem is then to compute

$$V(\pi) = \inf_{(\tau,d)} L_\pi(\tau,d) \tag{2.5}$$

and to find the optimal decision rule (τ_*,d_*), called *the π-Bayes decision rule*, at which the infimum in (2.5) is attained.

Observe that for any decision rule (τ,d) we have:

$$a\, P_\pi(d\!=\!0,\lambda\!=\!\lambda_1) \,+\, b\, P_\pi(d\!=\!1,\lambda\!=\!\lambda_0) \,=\, a\,\pi\,\alpha(d) + b\,(1\!-\!\pi)\,\beta(d) \tag{2.6}$$

where $\alpha(d) = P_1(d\!=\!0)$ is called *the probability of an error of the first kind*, and $\beta(d) = P_0(d\!=\!1)$ is called *the probability of an error of the second kind*.

4. The problem (2.5) can be reduced to an optimal stopping problem for the *a posteriori probability process* defined by

$$\pi_t = P_\pi\big(\lambda = \lambda_1 \mid \mathcal{F}_t^X\big) \tag{2.7}$$

with $\pi_0 = \pi$ under P_π. Standard arguments (see [13] p.166-167) show that

$$V(\pi) = \inf_\tau E_\pi\big(\, \tau + g_{a,b}(\pi_\tau)\,\big) \tag{2.8}$$

where $g_{a,b}(\pi) = a\pi \wedge b(1\!-\!\pi)$; the optimal stopping time τ_* in (2.8) is also optimal in (2.5), and the optimal decision function d_* is obtained by setting

$$\begin{aligned} d_* &= 1 \quad \text{if} \quad \pi_{\tau_*} \geq b/(a\!+\!b) \\ &= 0 \quad \text{if} \quad \pi_{\tau_*} < b/(a\!+\!b)\ . \end{aligned} \tag{2.9}$$

Our main task in the sequel is therefore reduced to solving the optimal stopping problem (2.8).

5. Another natural process, which is in a one-to-one correspondence with the process $(\pi_t)_{t\geq 0}$, is *the likelihood ratio process*; it is defined as the Radon-Nikodym density

$$\varphi_t = \frac{d\,(P_1 \,|\, \mathcal{F}_t^X)}{d\,(P_0 \,|\, \mathcal{F}_t^X)} \tag{2.10}$$

where $P_i \,|\, \mathcal{F}_t^X$ denotes the restriction of P_i to $\mathcal{F}_t^X$ for $i = 0, 1$. Since

$$\pi_t = \pi\, \frac{d\,(P_1 \,|\, \mathcal{F}_t^X)}{d\,(P_\pi \,|\, \mathcal{F}_t^X)} \tag{2.11}$$

where $P_\pi \mid \mathcal{F}_t^X = \pi\, P_1 \mid \mathcal{F}_t^X + (1\!-\!\pi)\, P_0 \mid \mathcal{F}_t^X$, it follows that

$$\pi_t = \left(\frac{\pi}{1\!-\!\pi}\, \varphi_t\right) \bigg/ \left(1 + \frac{\pi}{1\!-\!\pi}\, \varphi_t\right) \tag{2.12}$$

as well as that

$$\varphi_t = \frac{1-\pi}{\pi} \frac{\pi_t}{1-\pi_t} \; . \tag{2.13}$$

Moreover, the following explicit expression is known to be valid (see e.g. [5] or [9] Theorem 19.7):

$$\varphi_t = \exp\left(X_t \log\left(\frac{\lambda_1}{\lambda_0}\right) - (\lambda_1 - \lambda_0)\, t \right) . \tag{2.14}$$

This representation may now be used to reveal the *Markovian* structure in the problem. Since the process $(X_t)_{t\geq 0}$ is a time-homogeneous Markov process having stationary independent increments (Lévy process) under both P_0 and P_1 , from the representation (2.14), and due to the one-to-one correspondence (2.12), we see that $(\varphi_t)_{t\geq 0}$ and $(\pi_t)_{t\geq 0}$ are time-homogeneous Markov processes under both P_0 and P_1 with respect to natural filtrations which clearly coincide with $(\mathcal{F}_t^X)_{t\geq 0}$. Using then further that $E_\pi(Y \mid \mathcal{F}_t^X) = E_1(Y \mid \mathcal{F}_t^X)\, \pi_t + E_0(Y \mid \mathcal{F}_t^X)\, (1-\pi_t)$ for any (bounded) measurable Y , it follows that $(\pi_t)_{t\geq 0}$, and thus $(\varphi_t)_{t\geq 0}$ as well, is a time-homogeneous Markov process under each P_π for $\pi \in [0,1]$. (Observe, however, that although the same argument shows that $(X_t)_{t\geq 0}$ is a Markov process under each P_π for $\pi \in \langle 0,1 \rangle$, it is not a time-homogeneous Markov process unless π equals 0 or 1 .) Note also directly from (2.7) that $(\pi_t)_{t\geq 0}$ is a martingale under each P_π for $\pi \in [0,1]$. Thus, the optimal stopping problem (2.8) falls into the class of optimal stopping problems for Markov processes, and we therefore proceed by finding the infinitesimal operator of $(\pi_t)_{t\geq 0}$. A slight modification of the arguments above shows that all these processes possess a strong Markov property actually.

6. By Itô's formula (see e.g. [10] Ch. 2, §3 or [7] Ch. I, §4) one can verify that processes $(\varphi_t)_{t\geq 0}$ and $(\pi_t)_{t\geq 0}$ solve the following stochastic equations respectively:

$$d\varphi_t = \left(\frac{\lambda_1}{\lambda_0} - 1 \right) \varphi_{t-}\, d\big(X_t - \lambda_0 t\big) \tag{2.15}$$

$$d\pi_t = \frac{(\lambda_1 - \lambda_0)\, \pi_{t-}(1-\pi_{t-})}{\lambda_1 \pi_{t-} + \lambda_0\, (1-\pi_{t-})} \left(dX_t - \Big(\lambda_1 \pi_{t-} + \lambda_0\, (1-\pi_{t-}) \Big) dt \right) \tag{2.16}$$

(cf. formula (19.86) in [9]). The equation (2.16) may now be used to determine the infinitesimal operator of the Markov process $(\pi_t, \mathcal{F}_t^X, P_\pi)_{t\geq 0}$ for $\pi \in [0,1]$. For this, let $f = f(\pi)$ from $C^1[0,1]$ be given. Then by Itô's formula we find

$$\begin{aligned} f(\pi_t) &= f(\pi_0) + \int_0^t f'(\pi_{s-})\, d\pi_s + \sum_{0<s\leq t} \Big(f(\pi_s) - f(\pi_{s-}) - f'(\pi_{s-})\, \Delta \pi_s \Big) \\ &= f(\pi_0) + \int_0^t f'(\pi_{s-}) \Big(-(\lambda_1 - \lambda_0)\, \pi_{s-}(1-\pi_{s-}) \Big)\, ds + \sum_{0<s\leq t} \Big(f(\pi_s) - f(\pi_{s-}) \Big) \\ &= f(\pi_0) + \int_0^t f'(\pi_{s-}) \Big(-(\lambda_1 - \lambda_0) \pi_{s-}(1-\pi_{s-}) \Big)\, ds + \int_0^t \int_0^1 \Big(f(\pi_{s-}+y) - f(\pi_{s-}) \Big)\, \mu^\pi(ds,dy) \\ &= f(\pi_0) + \int_0^t f'(\pi_{s-}) \Big(-(\lambda_1 - \lambda_0) \pi_{s-}(1-\pi_{s-}) \Big)\, ds + \int_0^t \int_0^1 \Big(f(\pi_{s-}+y) - f(\pi_{s-}) \Big)\, \nu^\pi(ds,dy) \\ &\quad + \int_0^t \int_0^1 \Big(f(\pi_{s-}+y) - f(\pi_{s-}) \Big) \Big(\mu^\pi(ds,dy) - \nu^\pi(ds,dy) \Big) \end{aligned} \tag{2.17}$$

$$= f(\pi_0) + \int_0^t (\mathbb{L}f)(\pi_{s-})\, ds + M_t$$

where μ^π is the random measure of jumps of the process $(\pi_t)_{t\geq 0}$ and ν^π is a compensator of μ^π (see e.g. [10] Ch. 3 or [7] Ch. II), the operator $\mathbb{L}$ is given as in (2.19) below, and $M = (M_t)_{t\geq 0}$ defined as

$$M_t = \int_0^t \int_0^1 \Big(f(\pi_{s-} + y) - f(\pi_{s-}) \Big) \Big(\mu^\pi(ds, dy) - \nu^\pi(ds, dy) \Big) \tag{2.18}$$

is a local martingale with respect to $(\mathcal{F}_t^X)_{t\geq 0}$ and P_π for every $\pi \in [0,1]$. It follows from (2.17) that the infinitesimal operator of $(\pi_t)_{t\geq 0}$ acts on $f \in C^1[0,1]$ like

$$(\mathbb{L}f)(\pi) = -(\lambda_1 - \lambda_0)\pi(1-\pi)f'(\pi) + \Big(\lambda_1\pi + \lambda_0(1-\pi)\Big)\left(f\left(\frac{\lambda_1\pi}{\lambda_1\pi + \lambda_0(1-\pi)}\right) - f(\pi)\right) . \tag{2.19}$$

7. Looking back at (2.5) and using explicit expressions (2.4) and (2.6) with (2.1), it is easily verified (cf. [8] p. 105) that the payoff $\pi \mapsto V(\pi)$ is a concave function on $[0,1]$, and thus it is continuous on $\langle 0,1 \rangle$. Evidently, this function is pointwise dominated by $\pi \mapsto g_{a,b}(\pi)$. From these facts and from the general theory of optimal stopping for Markov processes (see e.g. [13]) we may guess that the payoff $\pi \mapsto V(\pi)$ from (2.8) should solve the following *Stephan problem* (for a differential-difference equation defined by the infinitesimal operator):

$$(\mathbb{L}V)(\pi) = -1 \ , \quad A_* < \pi < B_* \tag{2.20}$$

$$V(\pi) = a\,\pi \wedge b\,(1-\pi) \ , \quad \pi \notin \langle A_*, B_* \rangle \tag{2.21}$$

$$V(A_*+) = V(A_*) \ , \ V(B_*-) = V(B_*) \quad \text{(continuous fit)} \tag{2.22}$$

$$V'(A_*) = a \quad \text{(smooth fit)} \tag{2.23}$$

for some $0 < A_* < b/(a+b) < B_* < 1$ to be found. Observe that (2.21) contains two conditions relevant for the system: (i) $V(A_*) = aA_*$ and (ii) $V(\pi) = b(1-\pi)$ for $\pi \in [B_*, S(B_*)]$ with $S = S(\pi)$ from (2.24) below. These conditions are in accordance with the fact that if the process $(\pi_t)_{t\geq 0}$ starts or ends up at some π outside $\langle A_*, B_* \rangle$, we must stop it instantly.

Note from (2.16) that the process $(\pi_t)_{t\geq 0}$ moves continuously towards 0 and only jumps towards 1 at times of jumps of the point process X . This provides some intuitive support for the principle of smooth fit to hold at A_* . However, without a concavity argument it is not a priori clear why the condition $V(B_*-) = V(B_*)$ should hold at B_* . As *Figure 1* below shows, this is a rare property shared only by exceptional pairs (A, B) , and one could think that once A_* is fixed through the "smooth fit", the unknown B_* will be determined uniquely through the "continuous fit". While this train of thoughts sounds perfectly logical, we shall see quite opposite below that the equation (2.19) dictates our travel to solution from B_* to A_* .

Our next aim is to show that the three conditions in (2.22) and (2.23) are sufficient to determine a unique solution of the Stephan problem which in turn leads to the solution of the optimal stopping problem (2.8).

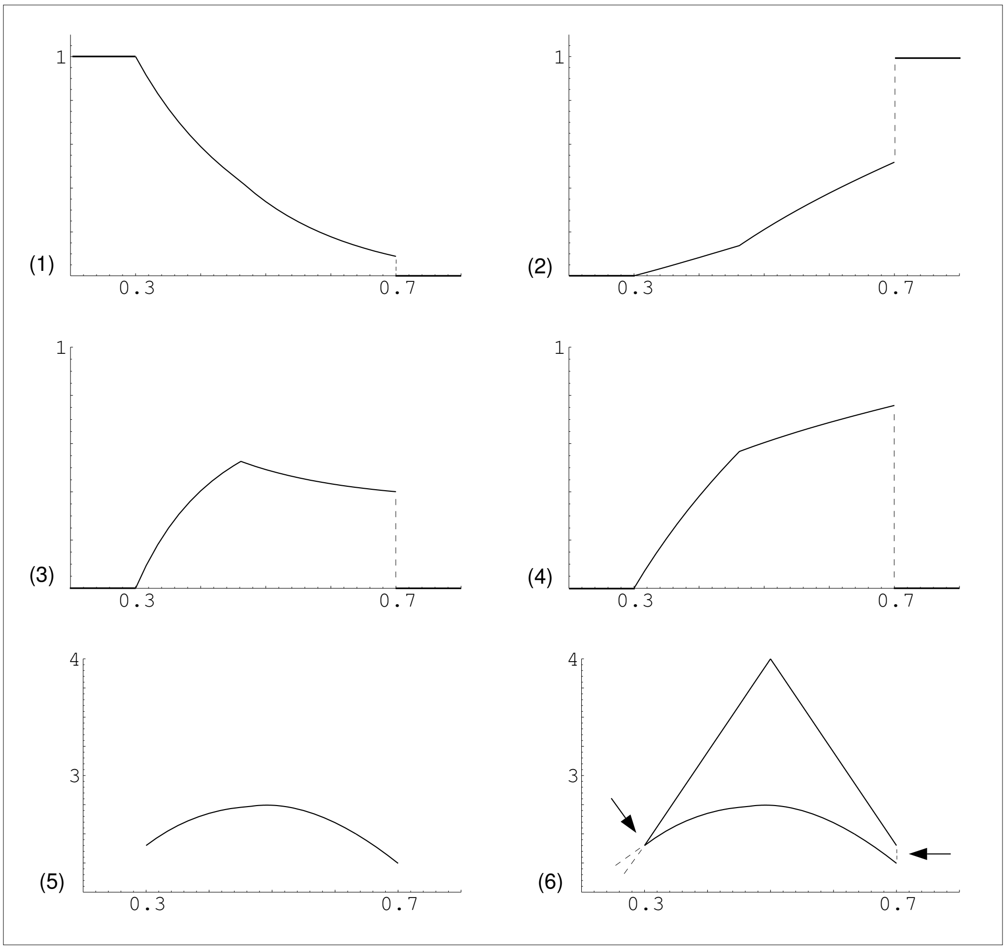

Figure 1. In view of the problem (2.8) and its decomposition via (2.4) and (2.6) with (2.1), we consider $\tau = \inf \{ \, t \geq 0 \mid \pi_t \notin \langle A, B \rangle \, \}$ for $(\pi_t)_{t \geq 0}$ from (2.7)+(2.12)+(2.14) with $\pi \in \langle A, B \rangle$ given and fixed, so that $\pi_0 = \pi$ under P_0 and P_1 ; the computer drawings above show the following functions respectively: (1) $\pi \mapsto P_1(\pi_\tau = A)$; (2) $\pi \mapsto P_0(\pi_\tau \geq B)$; (3) $\pi \mapsto E_1(\tau)$; (4) $\pi \mapsto E_0(\tau)$; (5) $\pi \mapsto \pi E_1(\tau) + (1-\pi) E_0(\tau) + a \pi P_1(\pi_\tau = A) + b\,(1-\pi)\, P_0(\pi_\tau \geq B) = E_\pi(\tau + g_{a,b}(\pi_\tau))$; (6) $\pi \mapsto E_\pi(\tau + g_{a,b}(\pi_\tau))$ and $\pi \mapsto g_{a,b}(\pi)$, where $A = 0.3$, $B = 0.7$, $\lambda_0 = 1$, $\lambda_1 = e$ and $a = b = 8$. Functions (1)-(4) are found by solving systems analogous to the system (3.15)-(3.17); their discontinuity at B should be noted, as well as the discontinuity of their first derivative at $B_1 = 0.46\ldots$ from (2.25); observe that the function (5) is a superposition of functions (1)-(4), and thus the same discontinuities carry over to the function (5), unless something special occurs. The crucial fact to be observed is that if the function (5) is to be the payoff (2.8), and thus extended by the gain function $\pi \mapsto g_{a,b}(\pi)$ outside $\langle A, B \rangle$, then such an extension would generally be discontinuous at B and have a discontinuous first derivative at A ; this is depicted in the final picture (6). It is a matter of fact that the optimal A_* and B_* are to be chosen in such a way that both of these discontinuities disappear; these are the *principles of continuous and smooth fit* respectively. Observe that in this case the discontinuity of the first derivative of (5) also disappears at B_1 , and the extension obtained is C^1 everywhere but at B_* where it is only C^0 generally (see *Figure 3* below).

8. *Solution of the Stephan problem (2.20)-(2.23).* Consider the equation (2.20) on $\langle 0, B]$ with some $B > b/(a+b)$ given and fixed. Introduce the "step" function

$$S(\pi) = \frac{\lambda_1 \pi}{\lambda_1 \pi + \lambda_0 (1-\pi)} \tag{2.24}$$

for $\pi \le B$. Observe that $\pi \mapsto S(\pi)$ is increasing, and find points $\ldots < B_2 < B_1 < B_0 := B$ such that $S(B_n) = B_{n-1}$ for $n \ge 1$. It is easily verified that

$$B_n = \frac{(\lambda_0)^n B}{(\lambda_0)^n B + (\lambda_1)^n (1-B)} \quad (n = 0, 1, \ldots\,). \tag{2.25}$$

Denote $I_n = \langle B_n, B_{n-1}]$ for $n \ge 1$, and introduce the "distance" function

$$d(\pi, B) = 1 + \left[\log\left(\frac{B}{1-B} \, \frac{1-\pi}{\pi} \right) \Big/ \log\left(\frac{\lambda_1}{\lambda_0} \right) \right] \tag{2.26}$$

for $\pi \le B$, where $[x]$ denotes the integer part of x. Observe that d is defined to satisfy

$$\pi \in I_n \iff d(\pi, B) = n \tag{2.27}$$

for all $0 < \pi \le B$.

Consider the equation (2.20) on I_1 upon setting $V(\pi) = b(1-\pi)$ for $\pi \in \langle B, S(B)]$; this is then a first-order linear differential equation which can be solved explicitly, and imposing a continuity condition at B which is in agreement with (2.22), we obtain a *unique* solution $\pi \mapsto V(\pi; B)$ on I_1; move then further and consider the equation (2.20) on I_2 upon using the solution found on I_1; this is then a first-order linear differential equation which can be solved explicitly, and imposing a continuity condition over $I_2 \cup I_1$ at B_1, we obtain a *unique* solution $\pi \mapsto V(\pi; B)$ on I_2; continuing this process by induction, we find the following formula:

$$V(\pi; B) = \frac{(1-\pi)^{\gamma_1}}{\pi^{\gamma_0}} \sum_{k=0}^{n-1} \left(C_{n-k} \, \frac{\beta^k}{k!} \, \log^k \left(\left(\frac{\lambda_1}{\lambda_0} \right)^{k-1} \frac{\pi}{1-\pi} \right) \right) - \left(n \, \frac{\lambda_1 - \lambda_0}{\lambda_0 \lambda_1} + b \right) \pi + \left(\frac{n}{\lambda_0} + b \right) \tag{2.28}$$

for $\pi \in I_n$, where $C_1, \ldots, C_n$ are constants satisfying the following recurrent relation:

$$C_{p+1} = \sum_{k=0}^{p-1} \left(C_{p-k} \left(f_k^{(p)} - f_{k+1}^{(p)} \right) \right) + \frac{(B_p)^{\gamma_0}}{(1-B_p)^{\gamma_1}} \left(\frac{\lambda_1 - \lambda_0}{\lambda_0 \lambda_1} \, B_p - \frac{1}{\lambda_0} \right) \tag{2.29}$$

for $p = 0, 1, \ldots, n-1$, with

$$f_k^{(p)} = \frac{\beta^k}{k!} \, \log^k \left(\left(\frac{\lambda_1}{\lambda_0} \right)^{k-p-1} \frac{B}{1-B} \right) \tag{2.30}$$

and where we set

$$(2.31) \qquad \gamma_0 = \frac{\lambda_0}{\lambda_1 - \lambda_0} \ ; \ \gamma_1 = \frac{\lambda_1}{\lambda_1 - \lambda_0} \ ; \ \beta = \frac{1}{(\lambda_1 - \lambda_0)} \, \frac{(\lambda_0)^{\gamma_1}}{(\lambda_1)^{\gamma_0}} \ .$$

Making use of the distance function (2.26), we may now write down the unique solution of (2.20) on $\langle 0, B]$ satisfying (2.21) on $[B, S(B)]$ and the second part of (2.22) at B :

$$(2.32) \qquad V(\pi; B) = \frac{(1-\pi)^{\gamma_1}}{\pi^{\gamma_0}} \sum_{k=0}^{d(\pi,B)-1} \left(C_{d(\pi,B)-k} \, \frac{\beta^k}{k!} \, \log^k \left(\left(\frac{\lambda_1}{\lambda_0} \right)^{k-1} \frac{\pi}{1-\pi} \right) \right) - \left(d(\pi, B) \, \frac{\lambda_1 - \lambda_0}{\lambda_0 \lambda_1} + b \right) \pi + \left(\frac{d(\pi, B)}{\lambda_0} + b \right)$$

for $0 < \pi \le B$. It is clear from our construction above that $\pi \mapsto V(\pi; B)$ is C^1 on $\langle 0, B \rangle$ and C^0 at B .

Observe that when computing the first derivative of $\pi \mapsto V(\pi; B)$, we can treat $d(\pi, B)$ in (2.32) as not depending on π . This then gives the following explicit expression:

$$(2.33) \qquad V'(\pi; B) = \frac{(1-\pi)^{\gamma_1 - 1}}{\pi^{\gamma_0 + 1}} \sum_{k=0}^{d(\pi,B)-1} \left(C_{d(\pi,B)-k} \, \frac{\beta^k}{k!} \, \log^k \left(\left(\frac{\lambda_1}{\lambda_0} \right)^{k-1} \frac{\pi}{1-\pi} \right) \cdot \cdot \left(k \big/ \log \left(\left(\frac{\lambda_1}{\lambda_0} \right)^{k-1} \frac{\pi}{1-\pi} \right) - (\pi + \gamma_0) \right) \right) - \left(d(\pi, B) \, \frac{\lambda_1 - \lambda_0}{\lambda_0 \lambda_1} + b \right)$$

for $0 < \pi \le B$.

Setting $C = b/(a+b)$ elementary calculations show that $\pi \mapsto V(\pi; B)$ is concave on $\langle 0, B \rangle$, as well as that $V(\pi; B) \to -\infty$ as $\pi \downarrow 0$, for all $B \in [C, 1]$. Moreover, it is easily seen from (2.28) (with $n = 1$) that $V(\pi; 1) < 0$ for all $0 < \pi < 1$. Thus, if for some $\hat{B} > C$, close to C , it happens that $\pi \mapsto V(\pi; \hat{B})$ crosses $\pi \mapsto a\pi$ when π moves to the left from $\hat{B}$, then a uniqueness argument presented in Remark 2.2 below (for different B's the curves $\pi \mapsto V(\pi; B)$ do not intersect) shows that there exists $B_* \in \langle C, 1 \rangle$, obtained by moving B from $\hat{B}$ to 1 or vice versa, such that for some $A_* \in \langle 0, C \rangle$ we have $V(A_*; B_*) = aA_*$ and $V'(A_*; B_*) = a$ (see *Figure 2*). Observe that the first identity captures part (i) of (2.22), while the second settles (2.23).

These considerations show that the system (2.20)-(2.23) has a unique (non-trivial) solution consisting of A_* , B_* and $\pi \mapsto V(\pi; B_*)$, if and only if

$$(2.34) \qquad \lim_{B \downarrow C} V'(B-; B) < a \ .$$

Geometrically this is the case when for $B > C$, close to C , the solution $\pi \mapsto V(\pi; B)$ intersects $\pi \mapsto a\pi$ at some $\pi < B$. It is now easily verified by using (2.28) (with $n = 1$) that (2.34) holds if and only if the following condition is satisfied:

$$(2.35) \qquad \lambda_1 - \lambda_0 > \frac{1}{a} + \frac{1}{b} \ .$$

In this process one should observe that B_1 from (2.25) tends to a number strictly less than C when $B \downarrow C$, so that all calculations are actually performed on I_1 .

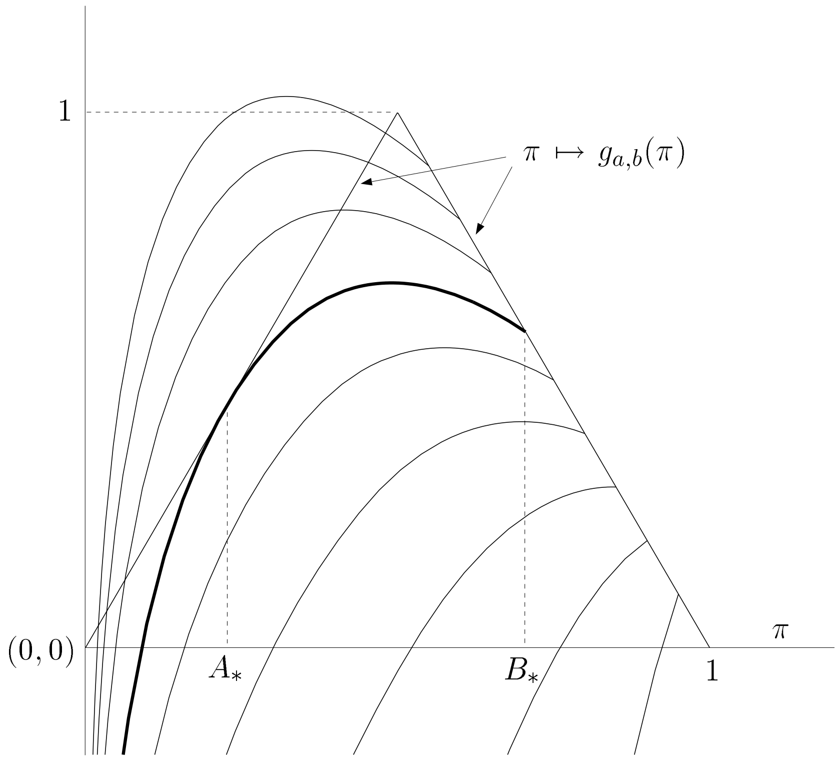

Figure 2. A computer drawing of "continuous fit" solutions $\pi \mapsto V(\pi;B)$ of (2.20), satisfying (2.21) on $[B, S(B)]$ and the second part of (2.22) at B, for different B in $\langle b/(a{+}b), 1\rangle$; in this particular case we took $B = 0.95, 0.80, 0.75, \ldots, 0.55$, with $\lambda_0 = 1$, $\lambda_1 = 5$ and $a = b = 2$. The unique B_* is obtained through the requirement that the map $\pi \mapsto V(\pi;B_*)$ hits "smoothly" the gain function $\pi \mapsto g_{a,b}(\pi)$ at A_* ; as shown above, this happens for $A_* = 0.22\ldots$ and $B_* = 0.70\ldots$; such obtained A_* and B_* are a unique solution of the system (2.38)-(2.39). The solution $\pi \mapsto V(\pi;B_*)$ leads to the explicit form of the payoff (2.8) as shown in *Figure* 3 below.

Thus, the condition (2.35) is necessary and sufficient for the existence of a unique non-trivial solution of the system (2.20)-(2.23); in this case the optimal A_* and B_* are uniquely determined as the solution of the system of transcendental equations $V(A_*;B_*) = aA_*$ and $V'(A_*;B_*) = a$, where $\pi \mapsto V(\pi;B)$ and $\pi \mapsto V'(\pi;B)$ are given by (2.32) and (2.33) respectively; once A_* and B_* are fixed, the solution $\pi \mapsto V(\pi;B_*)$ is given for $\pi \in [A_*, B_*]$ by means of (2.32).

9. *Solution of the optimal stopping problem (2.8).* We shall now show that the solution of the Stephan problem (2.20)-(2.23) found above coincides with the solution of the optimal stopping problem (2.8). This in turn leads to the solution of the Bayesian problem (2.5).

Theorem 2.1

(I): Suppose that the condition (2.35) holds. Then the π-Bayes decision rule (τ_*, d_*) in the problem (2.5) of testing two simple hypotheses H_1 and H_0 is explicitly given by (see Remark 2.3):

(2.36) $$\tau_* = \inf\big\{ \, t \ge 0 \mid \pi_t \notin \langle A_*, B_* \rangle \, \big\}$$

(2.37) $$\begin{aligned} d_* &= 1 \;\; (\textit{accept } H_1) \, , \;\; \textit{if} \;\; \pi_{\tau_*} \ge B_* \\ &= 0 \;\; (\textit{accept } H_0) \, , \;\; \textit{if} \;\; \pi_{\tau_*} = A_* \end{aligned}$$

where the constants A_* *and* B_* *satisfying* $0 < A_* < b/(a+b) < B_* < 1$ *are uniquely determined as solutions of the system of transcendental equations:*

(2.38) $$V(A_*; B_*) \, = \, a A_*$$

(2.39) $$V'(A_*; B_*) \, = \, a$$

with $\pi \mapsto V(\pi; B)$ *and* $\pi \mapsto V'(\pi; B)$ *in (2.32) and (2.33) respectively.*

(II): In the case when the condition (2.35) fails to hold, the π*-Bayes decision rule is trivial: Accept* H_1 *if* $\pi > b/(a+b)$ *, and accept* H_0 *if* $\pi < b/(a+b)$ *; either decision is equally good if* $\pi = b/(a+b)$ *.*

Proof. (I): 1. We showed above that the Stephan problem (2.20)-(2.23) is solvable if and only if (2.35) holds, and in this case the solution $\pi \mapsto V_*(\pi)$ is given explicitly by $\pi \mapsto V(\pi; B_*)$ in (2.32) for $A_* \le \pi \le B_*$, where A_* and B_* are a unique solution of (2.38)-(2.39).

In accordance with the interpretation of the Stephan problem, we extend $\pi \mapsto V_*(\pi)$ to the whole of $[0,1]$ by setting $V_*(\pi) = a\pi$ for $0 \le \pi < A_*$ and $V_*(\pi) = b(1-\pi)$ for $B_* < \pi \le 1$ (see *Figure 3*). Note that $\pi \mapsto V_*(\pi)$ is C^1 on $[0,1]$ everywhere but at B_* where it is C^0 . To complete the proof it is enough to show that such defined map $\pi \mapsto V_*(\pi)$ equals the payoff defined in (2.8), and that τ_* defined in (2.36) is an optimal stopping time.

2. Since $\pi \mapsto V_*(\pi)$ is not C^1 only at one point at which it is C^0 , we can apply Itô's formula to $V_*(\pi_t)$. In exactly the same way as in (2.17) this gives

(2.40) $$V_*(\pi_t) \; = \; V_*(\pi) + \int_0^t (\mathbb{L} V_*)(\pi_{s-}) \, ds + M_t$$

where $M = (M_t)_{t \ge 0}$ is a martingale given by

(2.41) $$M_t \, = \int_0^t \Big(V_*\big(\pi_{s-} + \Delta \pi_s\big) - V_*(\pi_{s-}) \Big) \, d\widehat{X}_s$$

and $\widehat{X}_t = X_t - \int_0^t E_\pi(\lambda \,|\, \mathcal{F}_{s-}^X) \, ds = X_t - \int_0^t (\lambda_1 \pi_{s-} + \lambda_0 (1-\pi_{s-})) \, ds$ is the so-called *innovation process* (see e.g. [9] Theorem 18.3) which is a martingale with respect to $(\mathcal{F}_t^X)_{t \ge 0}$ and P_π whenever $\pi \in [0,1]$. Note in (2.40) that we may extend V_*' arbitrarily to B_* as the time spent by the process $(\pi_t)_{t \ge 0}$ at B_* is of Lebesgue measure zero.

3. Recall that $(\mathbb{L} V_*)(\pi) = -1$ for $\pi \in \langle A_*, B_* \rangle$, and note that due to the smooth fit (2.23) we also have $(\mathbb{L} V_*)(\pi) \ge -1$ for all $\pi \in [0,1] \setminus \langle A_*, B_*]$.

To verify this claim first note that $(\mathbb{L} V_*)(\pi) = 0$ for $\pi \in \langle 0, S^{-1}(A_*) \rangle \cup \langle B_*, 1 \rangle$, since $\mathbb{L} f \equiv 0$ if $f(\pi) = a\pi$ or $f(\pi) = b(1-\pi)$. Observe also that $(\mathbb{L} V_*)(S^{-1}(A_*)) = 0$ and $(\mathbb{L} V_*)(A_*) = -1$ both due to the smooth fit (2.23). Thus, it is enough to verify that $(\mathbb{L} V_*)(\pi) \ge -1$ for $\pi \in \langle S^{-1}(A_*), A_* \rangle$.

For this, consider the equation $I\!\!L V = -1$ on $\langle S^{-1}(A_*), A_*]$ upon imposing $V(\pi) = V(\pi; B_*)$ for $\pi \in \langle A_*, S(A_*)]$, and solve it under the initial condition $V(A_*) = V(A_*; B_*) + c$ where $c \geq 0$. This generates a unique solution $\pi \mapsto V_c(\pi)$ on $\langle S^{-1}(A_*), A_*]$, and from (2.28) we read that $V_c(\pi) = V(\pi; B_*) + K_c(1-\pi)^{\gamma_1}/\pi^{\gamma_0}$ for $\pi \in \langle S^{-1}(A_*), A_*]$ where $K_c = c(A_*)^{\gamma_0}/(1-A_*)^{\gamma_1}$. (Observe that the curves $\pi \mapsto V_c(\pi)$ do not intersect on $\langle S^{-1}(A_*), A_*]$ for different c's.) Hence we see that there exists $c_0 > 0$ large enough such that for each $c > c_0$ the curve $\pi \mapsto V_c(\pi)$ lies strictly above the curve $\pi \mapsto a\pi$ on $\langle S^{-1}(A_*), A_*]$, and for each $c < c_0$ the two curves intersect. For $c \in [0, c_0\rangle$ let π_c denote the (closest) point (to A_*) at which $\pi \mapsto V_c(\pi)$ intersects $\pi \mapsto a\pi$ on $\langle S^{-1}(A_*), A_*]$. Then $\pi_0 = A_*$ and π_c decreases (continuously) in the direction of $S^{-1}(A_*)$ when c increases from 0 to c_0. Observe that the points π_c are 'good' points since by $V_c(\pi_c) = a\pi_c = V_*(\pi_c)$ with $V_c'(\pi_c) > a = V_*'(\pi_c)$ and $V_c(S(\pi_c)) = V(S(\pi_c); B_*) = V_*(S(\pi_c))$ we see from (2.19) that $(I\!\!L V_*)(\pi_c) \geq (I\!\!L V_c)(\pi_c) = -1$. Thus, if we show that π_c reaches $S^{-1}(A_*)$ when $c \uparrow c_0$, then the proof of the claim will be complete. Therefore assume on the contrary that this is not the case. Then $V_{c_1}(S^{-1}(A_*)-) = aS^{-1}(A_*)$ for some $c_1 < c_0$, and $V_c(S^{-1}(A_*)-) > aS^{-1}(A_*)$ for all $c > c_1$. Thus by choosing $c > c_1$ close enough to c_1, we see that a point $\widetilde{\pi}_c > S^{-1}(A_*)$ arbitrarily close to $S^{-1}(A_*)$ is obtained at which $V_c(\widetilde{\pi}_c) = a\widetilde{\pi}_c = V_*(\widetilde{\pi}_c)$ with $V_c'(\widetilde{\pi}_c) < a = V_*'(\widetilde{\pi}_c)$ and $V_c(S(\widetilde{\pi}_c)) = V(S(\widetilde{\pi}_c); B_*) = V_*(S(\widetilde{\pi}_c))$, from where we again see by (2.19) that $(I\!\!L V_*)(\widetilde{\pi}_c) \leq (I\!\!L V_c)(\widetilde{\pi}_c) = -1$. This however leads to a contradiction because $\pi \mapsto (I\!\!L V_*)(\pi)$ is continuous at $S^{-1}(A_*)$ (due to the smooth fit) and $(I\!\!L V_*)(S^{-1}(A_*)) = 0$ as already stated earlier. Thus, we have $(I\!\!L V_*)(\pi) \geq -1$ for all $\pi \in [0,1]$ (upon setting $V_*'(B_*) := 0$ for instance).

4. Recall further that $V_*(\pi) \leq g_{a,b}(\pi)$ for all $\pi \in [0,1]$. Moreover, since $\pi \mapsto V_*(\pi)$ is bounded, and $(X_t - \lambda_i\, t)_{t \geq 0}$ is a martingale under P_i for $i = 0, 1$, it is easily seen from (2.41) with (2.17) upon using the optional sampling theorem, that $E_\pi(M_\tau) = 0$ whenever τ is a stopping time of X such that $E_\pi(\tau) < \infty$. Thus, taking the expectation on both sides in (2.40), we obtain

$$V_*(\pi) \leq E_\pi\big(\, \tau + g_{a,b}(\pi_\tau)\, \big) \tag{2.42}$$

for all such stopping times, and hence $V_*(\pi) \leq V(\pi)$ for all $\pi \in [0,1]$.

5. On the other hand, the stopping time τ_* from (2.36) clearly satisfies $V_*(\pi_{\tau_*}) = g_{a,b}(\pi_{\tau_*})$. Moreover, a direct analysis of τ_* based on (2.12)-(2.14) (see Remark 2.3 below), together with the fact that for a Poisson process $(N_t)_{t \geq 0}$ the exit time of the process $(N_t - \mu\, t)_{t \geq 0}$ from $[\widetilde{A}, \widetilde{B}]$ has a finite expectation for any real μ, shows that $E_\pi(\tau_*) < \infty$ for all $\pi \in [0,1]$. Taking then the expectation on both sides in (2.40), we get

$$V_*(\pi) = E_\pi\big(\, \tau_* + g_{a,b}(\pi_{\tau_*})\big) \tag{2.43}$$

for all $\pi \in [0,1]$. This fact and the consequence of (2.42) stated above show that $V_* = V$, and that τ_* is an optimal stopping time. The proof of the first part is complete.

(II): Although, in principle, it is clear from our construction above that the second part of the theorem holds as well, we shall present a formal argument for completeness.

Suppose that the π-Bayes decision rule is not trivial. In other words, this means that $V(\pi) < g_{a,b}(\pi)$ for some $\pi \in \langle 0,1\rangle$. Since $\pi \mapsto V(\pi)$ is concave, this implies that there are $0 < A_* < b/(a+b) < B_* < 1$ such that $\tau_* = \inf\{\, t > 0 \mid \pi_t \notin \langle A_*, B_*\rangle \}$ is optimal for the prob-

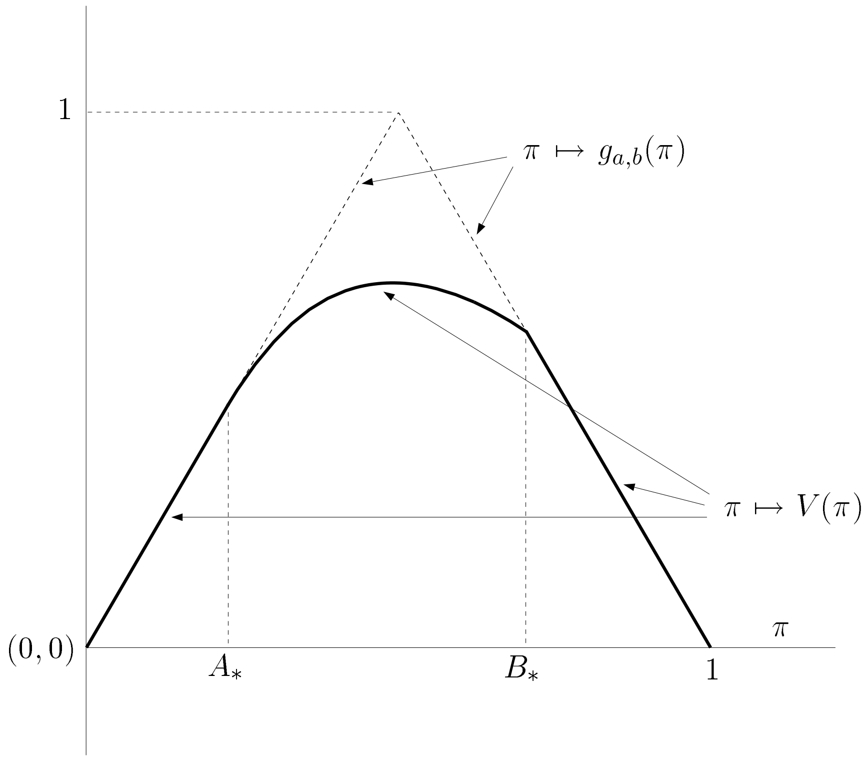

Figure 3. A computer drawing of the payoff (2.8) in the case $\lambda_0 = 1$, $\lambda_1 = 5$ and $a = b = 2$ as indicated in *Figure 2* above. The interval $\langle A_*, B_* \rangle$ is the region of continued observation of the process $(\pi_t)_{t\geq 0}$, while its complement in $[0,1]$ is the stopping region. Thus, as indicated in (2.36), the observation should be stopped as soon as the process $(\pi_t)_{t\geq 0}$ enters $[0,1] \setminus \langle A_*, B_* \rangle$, and this stopping time is optimal in the problem (2.8). The optimal decision function is then given by (2.37).

lems (2.8) and (2.5) respectively, with d_* from (2.9) in the latter case. Thus $V(\pi) = E_\pi(\tau_* + g_{a,b}(\pi_{\tau_*}))$ for $\pi \in [0,1]$, and therefore by the general Markov processes theory, and due to the strong Markov property of $(\pi_t)_{t\geq 0}$, we know that $\pi \mapsto V(\pi)$ solves (2.20) and satisfies (2.21) and (2.22); a priori we do not know if the smooth fit condition (2.23) is satisfied. Nevertheless, these arguments show the existence of a solution to (2.20) on $\langle 0, B_*]$ which is $b(1-B_*)$ at B_* and which crosses $\pi \mapsto a\pi$ at (some) $A_* < b/(a+b)$. But then the same uniqueness argument used in Subsection 8 above (see Remark 2.2 below) shows that there must exist points $\widehat{A}_* \leq A_*$ and $\widehat{B}_* \geq B_*$ such that the solution $\pi \mapsto \widehat{V}(\pi; \widehat{B}_*)$ of (2.20) satisfying $\widehat{V}(\widehat{B}_*; \widehat{B}_*) = b(1-\widehat{B}_*)$ hits $\pi \mapsto a\pi$ smoothly at $\widehat{A}_*$. The first part of the proof above then shows that the stopping time $\widehat{\tau}_* = \inf\{\, t>0 \mid \pi_t \notin \langle \widehat{A}_*, \widehat{B}_* \rangle \,\}$ is optimal. As this stopping time is known to be P_π-a.s. pointwise the smallest possible optimal stopping time (see the proof of Theorem 3.1 below), this shows that τ_* cannot be optimal unless the smooth fit condition holds at A_* , that is, unless $\widehat{A}_* = A_*$ and $\widehat{B}_* = B_*$. In any case, however, this argument implies the existence of a non-trivial solution to the system (2.20)-(2.23), and since this fact is equivalent to (2.35) as shown above, we see that condition (2.35) cannot be violated.

Observe that we have actually proved that if the optimal stopping problem (2.8) has a non-trivial solution, then the principle of smooth fit holds at A_* . An alternative proof of the statement could be done by using Lemma 3 on page 118 in [13]. The proof of the theorem is complete. □

Remark 2.2
The following probabilistic argument can be given to show that the two curves $\pi \mapsto V(\pi, B')$ and $\pi \mapsto V(\pi, B'')$ from (2.32) do not intersect on $\langle 0, B']$ whenever $0 < B' < B'' \leq 1$.

Assume that the two curves do intersect at some $Z < B'$. Let $\pi \mapsto \alpha\pi + \beta$ denote the tangent of the map $V(\,\cdot\,; B')$ at Z . Define a map $\pi \mapsto g(\pi)$ by setting $g(\pi) = (\alpha\pi + \beta) \wedge b(1-\pi)$ for $\pi \in [0,1]$, and consider the optimal stopping problem (2.8) with g instead of $g_{a,b}$. Let $V = V(\pi)$ denote the value function. Consider also the map $\pi \mapsto V_*(\pi)$ defined by $V_*(\pi) = V(\pi; B')$ for $\pi \in [Z, B']$ and $V_*(\pi) = g(\pi)$ for $\pi \in [0,1] \setminus [Z, B']$. As $\pi \mapsto V_*(\pi)$ is C^0 at B' and C^1 at Z , then in exactly the same way as in Subsections 3-5 (Part I) of the proof above we find that $V_*(\pi) = V(\pi)$ for all $\pi \in [0,1]$. However, if we consider the stopping time $\sigma_* = \inf \{\, t > 0 \mid \pi_t \notin \langle Z, B'' \rangle \,\}$, then it follows in the same way as in Subsection 5 (Part I) of the proof above that $V(\pi; B'') = E_\pi(\sigma_* + g(\pi_{\sigma_*}))$ for all $\pi \in [Z, B'']$. As $V(\pi; B'') < V_*(\pi)$ for $\pi \in \langle Z, B']$, this is a contradiction. Thus, the curves do not intersect.

Remark 2.3
1. Observe that the optimal decision rule (2.36)-(2.37) can be equivalently rewritten as follows

$$\tau_* = \inf \{\, t \geq 0 \mid Z_t \notin \langle \widetilde{A}_*, \widetilde{B}_* \rangle \,\} \tag{2.44}$$

$$\begin{aligned} d_* &= 1 \quad (\text{accept } H_1)\,, \quad \text{if } Z_{\tau_*} \geq \widetilde{B}_* \\ &= 0 \quad (\text{accept } H_0)\,, \quad \text{if } Z_{\tau_*} = \widetilde{A}_* \end{aligned} \tag{2.45}$$

where we use the following notation:

$$Z_t = X_t - \mu t \tag{2.46}$$

$$\widetilde{A}_* = \log\left(\frac{A_*}{1-A_*}\,\frac{1-\pi}{\pi}\right)\Big/\log\left(\frac{\lambda_1}{\lambda_0}\right) \tag{2.47}$$

$$\widetilde{B}_* = \log\left(\frac{B_*}{1-B_*}\,\frac{1-\pi}{\pi}\right)\Big/\log\left(\frac{\lambda_1}{\lambda_0}\right) \tag{2.48}$$

$$\mu = (\lambda_1 - \lambda_0)\Big/\log\left(\frac{\lambda_1}{\lambda_0}\right)\,. \tag{2.49}$$

2. The representation (2.44)-(2.45) reveals the structure and applicability of the optimal decision rule in a clearer manner. The result proved above shows that the following sequential procedure is optimal: *While observing X_t , monitor Z_t , and stop the observation as soon as Z_t enters either $\langle -\infty, \widetilde{A}_*]$ or $[\widetilde{B}_*, \infty\rangle$; in the first case conclude $\lambda = \lambda_0$, in the second conclude $\lambda = \lambda_1$.*

In this process the condition (2.35) must be satisfied, and the constants A_* and B_* should be determined as a unique solution of the system (2.38)-(2.39). This system can be successfully treated by means of standard numerical methods if one mimics our travel from B_* to A_* in the construction of our solution in Subsection 8 above. A pleasant fact is that only a few steps by (2.24) will be often needed to recapture A_* if one starts from B_* .

3. After we completed our work we observed that the same problem was treated by different methods in [11]. It is interesting to note that we could not find any later reference to that work. We also observed that the necessary and sufficient condition (2.35) of Theorem 2.1 is different from the condition $a\lambda_1+b(\lambda_0+\lambda_1)<b/a$ found in [11].

3. Solution of the variational problem

In the variational formulation of the problem it is assumed that the sequentially observed process $X=(X_t)_{t\geq 0}$ is a Poisson process with intensity λ_0 or λ_1, and no probabilistic assumption is made about the outcome of λ_0 and λ_1 at time 0. To formulate the problem we shall adopt the setting and notation from the previous section. Thus P_i is a probability measure on $(\Omega, \mathcal{F})$ under which $X=(X_t)_{t\geq 0}$ is a Poisson process with intensity λ_i for $i=0,1$.

1. Given the numbers α, $\beta>0$ such that $\alpha+\beta<1$, let $\Delta(\alpha,\beta)$ denote the class of all decision rules (τ,d) satisfying

(3.1) $$\alpha(d)\leq\alpha \quad \text{and} \quad \beta(d)\leq\beta$$

where $\alpha(d)=P_1(d=0)$ and $\beta(d)=P_0(d=1)$. The variational problem is then to find a decision rule $(\widehat{\tau},\widehat{d}\,)$ in the class $\Delta(\alpha,\beta)$ such that

(3.2) $$E_0(\widehat{\tau})\leq E_0(\tau) \quad \text{and} \quad E_1(\widehat{\tau})\leq E_1(\tau)$$

for any other decision rule (τ,d) from the class $\Delta(\alpha,\beta)$. Note that the main virtue of the requirement (3.2) is its simultaneous validity for both P_0 and P_1.

This formulation of the problem is due to Wald [15]. In the papers [17] and [18] Wald and Wolfowitz proved the optimality of the SPRT in the case of i.i.d. observations and under special assumptions on the admissibility of (α,β) (see [17], [18], [1], [8] for more details and compare it with the admissability notion given below). In the paper [5] Dvoretzky, Kiefer and Wolfowitz considered the problem of optimality of the SPRT in the case of a continuous time and satisfied themselves with the remark that "a careful examination of the results of [17] and [18] shows that their conclusions in no way require that the processes be discrete in time" omitting any further detail and concentrating their attention to the problem of finding the error probabilities $\alpha(d)$ and $\beta(d)$ with expectations $E_0(\tau)$ and $E_1(\tau)$ for the given SPRT (τ,d) defined by "stopping boundaries" A and B in the cases of a Wiener and Poisson process. The explicit solution of the Bayesian problem in the case of a Wiener process was given in [12] (see also [13]). For the general problem of the minimax optimality of the SPRT for the case of a continuous time see [6].

Our main aim in this section is to show how the solution of the variational problem together with a precise description of all admissible pairs (α,β) can be obtained from the Bayesian solution in the previous section. The sequential procedure which leads to the optimal decision rule $(\widehat{\tau},\widehat{d}\,)$ in this process is a SPRT which (as already mentioned earlier) was studied for the first time in [5]. We now describe a well-known procedure of passing from the Bayesian solution to the variational solution with some basic facts from [5] adapted to our aims.

2. Note that the explicit procedure of passing from the Bayesian solution to the variational solution presented in the next three steps is not confined to a Poissonian case but is also valid in

greater generality (cf. [8]).

Step 1 (Construction): Given $\alpha\,,\,\beta>0$ with $\alpha+\beta<1$, find constants A and B satisfying $A<0<B$ such that the stopping time

$$\hat{\tau} = \inf\,\big\{\; t\geq 0 \;\mid\; Z_t \notin \big\langle A,B\big\rangle \;\big\} \tag{3.3}$$

satisfies the following identities:

$$P_1\Big(Z_{\hat{\tau}} = A\Big) \;=\; \alpha \tag{3.4}$$

$$P_0\Big(Z_{\hat{\tau}} \geq B\Big) \;=\; \beta \tag{3.5}$$

where $(Z_t)_{t\geq 0}$ is as in (2.46). Associate with $\hat{\tau}$ the following decision function:

$$\begin{aligned} \widehat{d} &= 1 \;\;\text{(accept } H_1 \text{) , if } Z_{\hat{\tau}}\geq B \\ &= 0 \;\;\text{(accept } H_0 \text{) , if } Z_{\hat{\tau}} = A \;. \end{aligned} \tag{3.6}$$

We shall actually see below that not for all values α and β such A and B exist; a function $G:\big\langle 0,1\big\rangle \rightarrow \big\langle 0,1\big\rangle$ is displayed in (3.24) such that the solution (A,B) to (3.4)-(3.5) exists only for $\beta\in\big\langle 0,G(\alpha)\big\rangle$ if $\alpha\in\big\langle 0,1\big\rangle$. Such values α and β will be called *admissible*.

Step 2 (Embedding): Once A and B are found for admissible α and β , we may respectively identify them with $\widetilde{A}_*$ and $\widetilde{B}_*$ from (2.47) and (2.48). Then, for any $\widehat{\pi}\in\big\langle 0,1\big\rangle$ given and fixed, we can uniquely determine A_* and B_* satisfying $0<A_*<B_*<1$ such that (2.47) and (2.48) hold with $\pi=\widehat{\pi}$. Once A_* and B_* are given, we can choose $a>0$ and $b>0$ in the Bayesian problem (2.4)+(2.5) such that the optimal stopping time in (2.8) is exactly the exit time τ_* of $(\pi_t)_{t\geq 0}$ from $\big\langle A_*,B_*\big\rangle$ as given in (2.36). Observe that this is possible to achieve since the optimal A_* and B_* range through all $\big\langle 0,1\big\rangle$ when a and b satisfying (2.35) range through $\big\langle 0,\infty\big\rangle$. (For this, let any $B_*\in\big\langle 0,1\big\rangle$ be given and fixed, and choose $\widetilde{a}>0$ and $\widetilde{b}>0$ such that $B_*=\widetilde{b}/(\widetilde{a}+\widetilde{b})$ with $\lambda_1{-}\lambda_0=1/\widetilde{a}+1/\widetilde{b}$. Then consider the solution $V(\,\cdot\,;B_*):=V_b(\,\cdot\,;B_*)$ of (2.20) on $\big\langle 0,B_*\big\rangle$ upon imposing $V_b(\pi;B_*)=b(1{-}\pi)$ for $\pi\in[B_*,S(B_*)]$ where $b\geq\widetilde{b}$. To each such a solution there corresponds $a>0$ such that $\pi\mapsto a\pi$ hits $\pi\mapsto V_b(\pi;B_*)$ smoothly at some $A=A(b)$. When b increases from $\widetilde{b}$ to ∞ , then $A(b)$ decreases from B_* to zero. This is easily verified by a simple comparison argument upon noting that $\pi\mapsto V_b(\pi;B_*)$ stays strictly above $\pi\mapsto V(\pi;B_*)+V_b(B_*;B_*)$ on $\big\langle 0,B_*\big\rangle$ (recall the idea used in Remark 2.3 above). As each $A(b)$ obtained (in the pair with B_*) is optimal (recall the arguments used in Subsections 3-5 (Part I) of the proof of Theorem 2.1), the proof of the claim is complete.)

Step 3 (Verification): Consider the process $(\widehat{\pi}_t)_{t\geq 0}$ defined by (2.12)+(2.14) with $\pi=\widehat{\pi}$, and denote by $(\widehat{\tau}_*,\widehat{d}_*)$ the optimal decision rule (2.36)-(2.37) associated with it. From our construction above note that $\widehat{\tau}$ from (3.3) actually coincides with $\widehat{\tau}_*$, as well as that $(\widehat{\pi}_{\hat{\tau}_*}=A_*)=(Z_{\hat{\tau}}=A)$ and $(\widehat{\pi}_{\hat{\tau}_*}\geq B_*)=(Z_{\hat{\tau}}\geq B)$. Thus (3.4) and (3.5) show that

$$P_1\Big(\widehat{d}_*\!=0\Big) \;=\; \alpha \tag{3.7}$$

$$P_0\Big(\widehat{d}_*\!=1\Big) \;=\; \beta \tag{3.8}$$

for the admissible α and β. If now any decision rule (τ, d) from $\Delta(\alpha, \beta)$ is given, then either $P_1(d=0) = \alpha$ and $P_0(d=1) = \beta$, or at least one strict inequality holds. In both cases, however, from (2.4)-(2.6) and (3.7)+(3.8) we easily see that $E_{\widehat{\pi}}(\widehat{\tau}_*) \le E_{\widehat{\pi}}(\tau)$, since otherwise $\widehat{\tau}_*$ would not be optimal. Since $\widehat{\tau}_* = \widehat{\tau}$, it follows $E_{\widehat{\pi}}(\widehat{\tau}) \le E_{\widehat{\pi}}(\tau)$, and letting $\widehat{\pi}$ first to 0 and then to 1, we obtain (3.2) in the case when $E_0(\tau) < \infty$ and $E_1(\tau) < \infty$. If either $E_0(\tau)$ or $E_1(\tau)$ equals ∞, then (3.2) follows by the same argument after a simple truncation (e.g. if $E_0(\tau) < \infty$ but $E_1(\tau) = \infty$ choose $n \ge 1$ such that $P_0(\tau > n) \le \varepsilon$, apply the same argument to $\tau_n := \tau \wedge n$ and $d_n := d\, 1_{\{\tau \le n\}} + 1_{\{\tau > n\}}$, and let ε go to zero in the end.) This solves the variational problem posed above for all admissible α and β.

3. The preceding arguments also show:

(3.9) If either $P_1(d=0) < \alpha$ or $P_0(d=1) < \beta$ for some $(\tau, d) \in \Delta(\alpha, \beta)$ with admissible α and β, then at least one strict inequality in (3.2) holds.

Moreover, since $\widehat{\tau}_*$ is known to be $P_{\widehat{\pi}}$-a.s. the smallest possible optimal stopping time (see the proof of Theorem 3.1 below), from the arguments above we also get:

(3.10) If $P_1(d=0) = \alpha$ and $P_0(d=1) = \beta$ for some $(\tau, d) \in \Delta(\alpha, \beta)$ with admissible α and β, and both equalities in (3.2) hold, then $\tau = \widehat{\tau}$ P_0-a.s. and P_1-a.s.

The property (3.10) characterises $\widehat{\tau}$ as *a unique stopping time of the decision rule with maximal admissible error probabilities having both P_0 and P_1 expectation at minimum.*

4. It remains to determine admissible α and β in (3.4) and (3.5) above. For this, consider $\widehat{\tau}$ defined in (3.3) for some $A < 0 < B$, and note from (2.14) that $\varphi_t = \exp\big(Z_t \log(\lambda_1/\lambda_0)\big)$. By means of (2.10) we find

$$
\begin{aligned}
(3.11) \qquad P_1\Big\{ Z_{\widehat{\tau}} = A \Big\} &= P_1\left\{ \varphi_{\widehat{\tau}} = \exp\left(A \log\left(\frac{\lambda_1}{\lambda_0}\right)\right)\right\} \\
&= \exp\left(A \log\left(\frac{\lambda_1}{\lambda_0}\right)\right) P_0\Big\{ Z_{\widehat{\tau}} = A \Big\} = \exp\left(A \log\left(\frac{\lambda_1}{\lambda_0}\right)\right) \Big(1 - P_0\Big\{ Z_{\widehat{\tau}} \ge B \Big\}\Big) .
\end{aligned}
$$

Using (3.4)-(3.5), from (3.11) we see that

$$
(3.12) \qquad A = \log\left(\frac{\alpha}{1-\beta}\right) \Big/ \log\left(\frac{\lambda_1}{\lambda_0}\right) .
$$

To determine B, let P_0^z be a probability measure under which $(X_t)_{t \ge 0}$ is a Poisson process with intensity λ_0 and $(Z_t)_{t \ge 0}$ starts at z. It is easily seen that the infinitesimal operator of $(Z_t)_{t \ge 0}$ under $(P_0^z)_{z \in \mathbf{R}}$ acts like

$$
(3.13) \qquad (\mathbb{L}_0 f)(z) = -\mu f'(z) + \lambda_0 \Big(f(z+1) - f(z) \Big) .
$$

In view of (3.5), introduce the function

$$
(3.14) \qquad u(z) = P_0^z \Big(Z_{\widehat{\tau}} \ge B \Big) .
$$

Strong Markov arguments then show that $z \mapsto u(z)$ solves the following system:

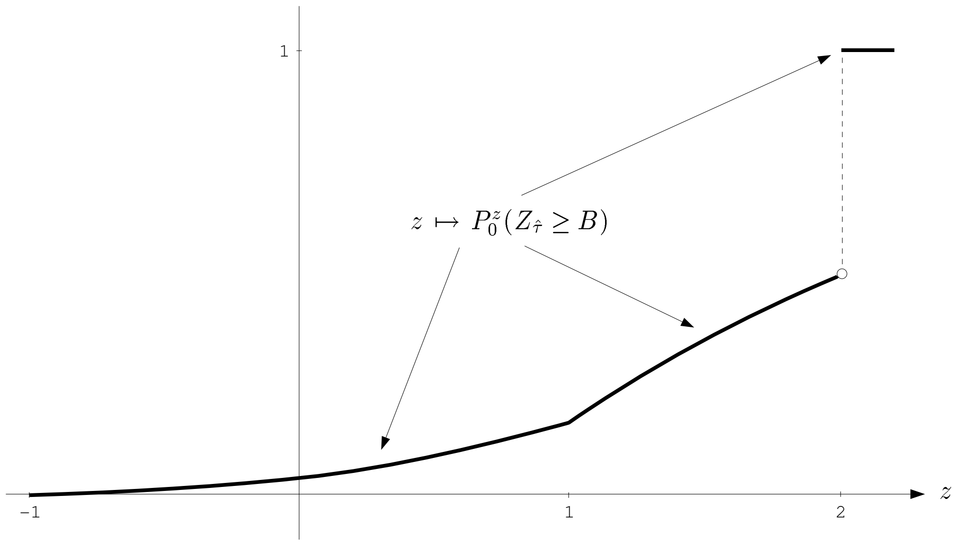

Figure 4. A computer drawing of the map $u(z) = P_0^z(Z_{\hat{\tau}} \geq B)$ from (3.14) in the case $A = -1$, $B = 2$ and $\lambda_0 = 0.5$. This map is a unique solution of the system (3.15)-(3.17). Its discontinuity at B should be noted, as well as the discontinuity of its first derivative at $B-1$. Observe also that $u(A+) = u(A) = 0$. The case of general A , B and λ_0 looks very much the same.

(3.15) $$(\mathbb{L}_0 u)(z) = 0 \quad \text{if} \quad z \in \langle A, B \rangle \setminus \{B-1\}$$

(3.16) $$u(A) = 0$$

(3.17) $$u(z) = 1 \quad \text{if} \quad z \geq B \ .$$

The solution of this system is given in (4.15) of [5]. To display it, introduce the function

(3.18) $$F(x;B) = \sum_{k=0}^{\delta(x,B)} \frac{(-1)^k}{k!} \Big(\big(B-x-k\big)\, \rho\, e^{-\rho} \Big)^k$$

for $x \leq B$, where we denote

(3.19) $$\delta(x,B) = -[x-B+1]$$

(3.20) $$\rho = \log\left(\frac{\lambda_1}{\lambda_0}\right) \Big/ \left(\frac{\lambda_1}{\lambda_0} - 1\right) \ .$$

Setting $J_n = [B-n-1, B-n\rangle$ for $n \geq 0$, observe that $\delta(x,B) = n$ if and only if $x \in J_n$.

It is then easily verified that the solution of the system (3.15)-(3.17) is given by

(3.21) $$u(z) = 1 - e^{-\rho(z-A)}\, \frac{F(z;B)}{F(A;B)}$$

for $A \leq z < B$. Note that $z \mapsto u(z)$ is C^1 everywhere in $\langle A, B \rangle$ but at $B-1$ where it is only C^0 ; note also that $u(A+) = u(A) = 0$, but $u(B-) < u(B) = 1$ (see *Figure 4*).

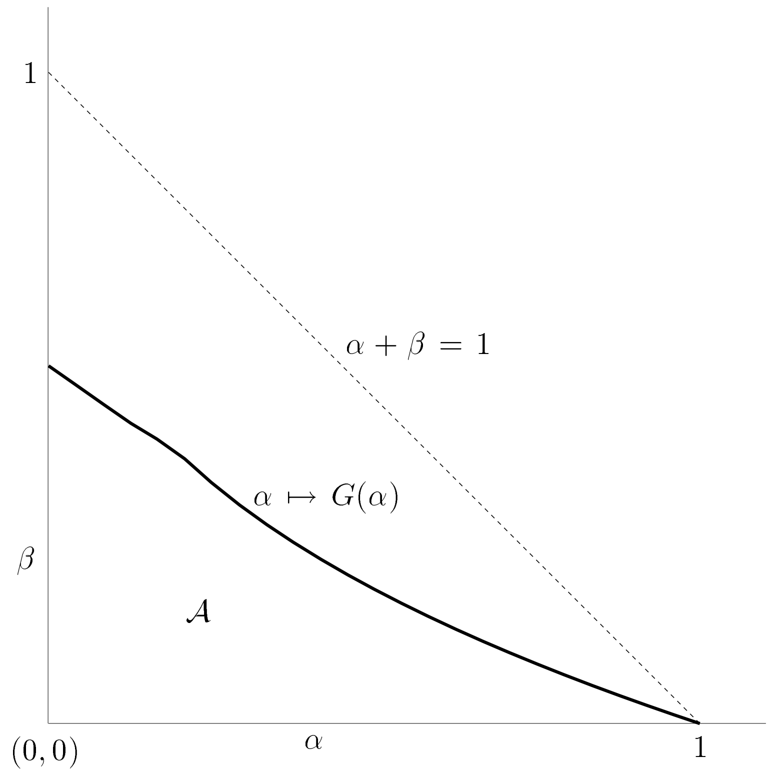

Figure 5. A computer drawing of the map $\alpha \mapsto G(\alpha)$ from (3.24) in the case $\lambda_0 = 1$ and $\lambda_1 = 3$. The area $\mathcal{A}$ which lies below the graph of G determines the set of all admissible α and β . The case of general λ_0 and λ_1 looks very much the same; it can also be shown that $G(0+)$ decreases if the difference $\lambda_1 - \lambda_0$ increases, as well as that $G(0+)$ increases if both λ_0 and λ_1 increase so that the difference $\lambda_1 - \lambda_0$ remains constant; in all cases $G(1-) = 0$. It may seem somewhat surprising that $G(0+) < 1$; observe, however, this is in agreement with the fact that $(Z_t)_{t \geq 0}$ from (2.46) is a supermartingale under P_0 . (A little peak on the graph, at $\hat{\alpha} = 0.19\ldots$ and $\hat{\beta} = 0.42\ldots$ in this particular case, corresponds to the disturbance when A from (3.12) passes through -1 while $B = 0+$; it is caused by a discontinuity of the first derivative of the map from (3.22) at $B-1$ (see Figure 4).)

Going back to (3.5), and using (3.21), we see that

$$P_0\Big(Z_{\hat{\tau}} \geq B\Big) = 1 - e^{\rho A}\, \frac{F(0;B)}{F(A;B)} \; . \tag{3.22}$$

Letting $B \downarrow 0$ in (3.22), and using the fact that the expression (3.22) is continuous in B and decreases to 0 as $B \uparrow \infty$, we clearly obtain a necessary and sufficient condition on β to satisfy (3.5), once $A = A(\alpha, \beta)$ is fixed through (3.12); as $F(0;0) = 1$, this condition reads

$$\beta < 1 - \frac{e^{\rho A(\alpha,\beta)}}{F\big(A(\alpha,\beta);0\big)} \; . \tag{3.23}$$

Note, however, if β increases, then the function on the right-hand side in (3.23) decreases, and thus there exists a unique $\beta_* = \beta_*(\alpha) > 0$ at which equality in (3.23) is attained. (This value can easily be computed by means of standard numerical methods.) Setting

(3.24) $$G(\alpha) = 1 - \frac{e^{\rho A(\alpha,\beta_*(\alpha))}}{F\big(A(\alpha,\beta_*(\alpha));0\big)}$$

we see that admissible α and β are characterised by $0 < \beta < G(\alpha)$ (see *Figure 5*). In this case A is given by (3.12), and B is uniquely determined from the equation

(3.25) $$F(0;B) - (1-\beta)\, F(A;B)\, e^{-\rho A} = 0 \; .$$

The set of all admissible α and β will be denoted by $\mathcal{A}$. Thus, we have

(3.26) $$\mathcal{A} = \Big\{\, (\alpha,\beta) \mid 0 < \alpha < 1 \,,\, 0 < \beta < G(\alpha) \,\Big\} \; .$$

5. The preceding considerations may be summarised as follows (see also Remark 3.2 below).

Theorem 3.1

In the problem (3.1)-(3.2) of testing two simple hypotheses (2.2)-(2.3) based upon sequential observations of the Poisson process $X = (X_t)_{t \ge 0}$ *under* P_0 *or* P_1 *, there exists a unique decision rule* $(\widehat{\tau}, \widehat{d}\,) \in \Delta(\alpha,\beta)$ *satisfying (3.2) for any other decision rule* $(\tau, d) \in \Delta(\alpha,\beta)$ *whenever* $(\alpha,\beta) \in \mathcal{A}$ *. The decision rule* $(\widehat{\tau}, \widehat{d}\,)$ *is explicitly given by (3.3)+(3.6) with* A *in (3.12) and* B *from (3.25), it satisfies (3.9), and is characterised by (3.10).*

Proof. It only remains to prove (3.10). For this, in the notation used above, assume that τ is a stopping time of X satisfying the hypotheses of (3.10). Then clearly τ is an optimal stopping time in (2.8) for $\pi = \widehat{\pi}$ with a and b as in Step 2 above.

Recall that $V_*(\pi) \le g_{a,b}(\pi)$ for all π , and observe that $\widehat{\tau}$ can be written as

(3.27) $$\widehat{\tau} = \inf \Big\{\, t \ge 0 \mid V_*(\widehat{\pi}_t) \ge g_{a,b}(\widehat{\pi}_t) \,\Big\}$$

where $\pi \mapsto V_*(\pi)$ is the payoff (2.8) appearing in the proof of Theorem 2.1. Supposing now that $P_{\widehat{\pi}}(\tau < \widehat{\tau}) > 0$, we easily find by (3.27) that

(3.28) $$E_{\widehat{\pi}}\Big(\, \tau + g_{a,b}\big(\widehat{\pi}_\tau\big) \,\Big) > E_{\widehat{\pi}}\Big(\, \tau + V_*\big(\widehat{\pi}_\tau\big) \,\Big) \; .$$

On the other hand, it is clear from (2.40) with $LV_* \ge -1$ that $(t + V_*(\widehat{\pi}_t))_{t \ge 0}$ is a submartingale. Thus by the optional sampling theorem it follows that

(3.29) $$E_{\widehat{\pi}}\Big(\, \tau + V_*\big(\widehat{\pi}_\tau\big) \,\Big) \ge V_*(\widehat{\pi}) \; .$$

However, from (3.28) and (3.29) we see that τ cannot be optimal, and thus we must have $P_{\widehat{\pi}}(\tau \ge \widehat{\tau}) = 1$. Moreover, since it follows from our assumption that $E_{\widehat{\pi}}(\tau) = E_{\widehat{\pi}}(\widehat{\tau})$, this implies that $\tau = \widehat{\tau}$ $P_{\widehat{\pi}}$-a.s. Finally, as $P_i << P_{\widehat{\pi}}$ for $i = 0, 1$, we see that $\tau = \widehat{\tau}$ both P_0-a.s. and P_1-a.s. The proof of the theorem is complete. □

Observe that the sequential procedure of the optimal decision rule $(\widehat{\tau}, \widehat{d}\,)$ from Theorem 3.1 is precisely the SPRT. The explicit formulas for $E_0(\widehat{\tau})$ and $E_1(\widehat{\tau})$ are given in (4.22) of [5].

Remark 3.2

If $(\alpha, \beta) \notin \mathcal{A}$, that is, if $\beta \geq G(\alpha)$ for some α , $\beta > 0$ such that $\alpha + \beta < 1$, then no decision rule given by the SPRT-form (3.3)+(3.6) can solve the variational problem (3.1)-(3.2).

To see this, let such $(\alpha, \beta^*) \notin \mathcal{A}$ be given, and let (τ, d) be a decision rule satisfying (3.3)+(3.6) for some $A < 0 < B$. Denote $\beta = P_0(Z_\tau \geq B)$ and choose α to satisfy (3.12). Then $\beta < G(\alpha) \leq \beta^*$ by definition of the map G . Given $\beta' \in \langle \beta, G(\alpha) \rangle$, let B' be taken to satisfy (3.5) with β' , and let α' be determined from (3.12) with β' so that A remains unchanged. Clearly $0 < B' < B$ and $0 < \alpha' < \alpha$, and (3.4) holds with A and α' respectively. But then (τ', d') satisfying (3.3)+(3.6) with $A < 0 < B'$ still belongs to $\Delta(\alpha, \beta^*)$, while clearly $\tau' < \tau$ both under P_0 and P_1 . This shows that (τ, d) does not solve the variational problem.

The preceding argument shows that the admissible class $\mathcal{A}$ from (3.26) is exactly the class of all error probabilities (α, β) for which the SPRT is optimal. A pleasant fact is that $\mathcal{A}$ always contains a neighborhood around $(0, 0)$ in $[0, 1] \times [0, 1]$, which is the most interesting case from the point of view of statistical applications.

Acknowledgment. The authors are obliged to a referee and members of the editorial board for some useful remarks which are incorporated in the final version of the paper.

REFERENCES

[1] ARROW, K. J. BLACKWELL, D. *and* GIRSHICK, M. A. (1949). Bayes and minimax solutions of sequential decision problems. *Econometrica* 17 (213-244).

[2] BHAT, B. R. (1988). Optimal properties of SPRT for some stochastic processes. *Contemp. Math.* 80 (285-299).

[3] BLACKWELL, D. *and* GIRSHICK, M. A. (1954). *Theory of Games and Statistical Decisions*. John Wiley and Sons, New York.

[4] CHOW, Y. S. ROBBINS, H. *and* SIEGMUND, D. (1971). *Great Expectations: The Theory of Optimal Stopping*. Houghton Mifflin, Boston, Mass.

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

[6] IRLE, A. *and* SCHMITZ, N. (1984). On the optimality of the SPRT for processes with continuous parameter. *Math. Operationsforsch. Statist. Ser. Statist.* 15 (91-104).

[7] JACOD, J. *and* SHIRYAEV, A. N. (1987). *Limit Theorems for Stochastic Processes*. Springer-Verlag, Berlin Heidelberg.

[8] LEHMANN, E. L. (1959). *Testing Statistical Hypotheses*. John Wiley.

[9] LIPTSER, R. S. *and* SHIRYAYEV, A. N. (1978). *Statistics of Random Processes II*. Springer-Verlag, New York.

[10] LIPTSER, R. Sh. *and* SHIRYAYEV, A. N. (1989). *Theory of Martingales*. Kluwer Acad. Publ.

[11] ROMBERG, H. F. (1972). Continuous sequential testing of a Poisson process to minimize the Bayes risk. *J. Amer. Statist. Assoc.* 67 (921-926).

[12] SHIRYAEV, A. N. (1967). Two problems of sequential analysis. *Cybernetics* 3 (63-69).

[13] SHIRYAEV, A. N. (1978). *Optimal Stopping Rules*. Springer-Verlag, New York.

[14] SIEGMUND, D. (1985). *Sequential Analysis. Tests and Confidence Intervals*. Springer-Verlag, New York.

[15] WALD, A. (1947). *Sequenatial Analysis*. John Wiley and Sons, New York.

[16] WALD, A. (1950). *Statistical Decision Functions*. John Wiley and Sons, New York.

[17] WALD, A. *and* WOLFOWITZ, J. (1948). Optimum character of the sequential probability ratio test. *Ann. Math. Statistics* 19 (326-339).

[18] WALD, A. *and* WOLFOWITZ, J. (1950). Bayes solutions of sequential decision problems. *Ann. Math. Statistics* 21 (82-99).

Goran Peskir
Department of Mathematical Sciences
University of Aarhus, Denmark
Ny Munkegade, DK-8000 Aarhus
home.imf.au.dk/goran
goran@imf.au.dk

Albert N. Shiryaev
Steklov Mathematical Institute
Gubkina str. 8
117966 Moscow
Russia
shiryaev@mi.ras.ru