Domingo Morales; Leandro Pardo; Vicente Quesada

Sequential estimation of survival functions with a neutral to the right process prior

Applications of Mathematics, Vol. 39 (1994), No. 2, 97–109

Persistent URL: `http://dml.cz/dmlcz/134247`

SEQUENTIAL ESTIMATION OF SURVIVAL FUNCTIONS WITH A NEUTRAL TO THE RIGHT PROCESS PRIOR*

DOMINGO MORALES, LEANDRO PARDO, VICENTE QUESADA, Madrid

(Received April 27, 1991)

Summary. In this work, a parametric sequential estimation method of survival functions is proposed in the Bayesian nonparametric context when neutral to the right processes are used. It is proved that the mentioned method is an 1-SLA rule when Dirichlet processes are used; furthermore, asymptotically pointwise optimal procedures are obtained. Finally, an example is given.

Keywords: Survival analysis, Bayesian nonparametric inference, neutral to the right processes, Dirichlet processes, sequential analysis, stopping rules, 1-SLA rule, asymptotically pointwise optimal procedures

AMS classification: Primary 62L20, secondary 62F15, 62G05

1. INTRODUCTION

Let S be a survival function on the positive real line and let $T_1, \ldots, T_n$ be a random sample from S. There is a positive cost $c > 0$ each time the statistician looks at a new observation. We consider the following loss function

$$L(S, \hat{S}) = \int_{\mathbf{R}^+} \left(S(t) - \hat{S}(t)\right)^2 \mathrm{d}W(t), \tag{1.1}$$

where W is some finite measure on $\mathbf{R}^+$ not having common atoms with the probability measure given by the survival function S_0, which represents the prior knowledge. The purpose of this article is to investigate a sequential nonparametric problem from a Bayesian viewpoint using a neutral to the right process prior (N.R.P.), and the

* The research in this paper was supported in part by DGICYT Grant N. PB91-0387. Their financial support is gratefully acknowledged.

estimation of survival functions on the real line. See Ferguson [4] and Doksum [2] for definitions and properties of the random survival functions.

After each observation, the statistician must decide whether to take another observation or to stop sampling and choose an estimate $\hat{S}$. In this work, we present a sequential estimation method based on a parametric technique given by Morales, Quesada and Pardo [6]. We prove that if a Dirichlet process prior (D.P.) is used, then the parametric procedure is equal to the 1-SLA method (1-stage-look-ahead) proposed by Ferguson [5]. Furthermore, asymptotically pointwise optimal rules (A.P.O.) in the sense of Bickel and Yahav [1] are given.

Only in very exceptional parametric cases it happens that an optimal Bayes stopping rule can be found explicitly. In the Bayesian nonparametric context, approximation to the optimal rule must be found.

2. 1-SPLA ESTIMATION WITH N.R.P. PRIOR

In this section, we give a parametric sequential method to estimate survival functions: 1-stage parametric look ahead method (1-SPLA).

Let $T \geqslant 0$ represent the time taken for an event of interest to occur. We suppose that the prior distribution over the space of survival functions is given by a N.R.P. $S(t) = \exp\big(-Y(t)\big)$, where $Y(t)$ is a stochastic process with independent increments verifying: (1) $Y(t)$ is non-decreasing a.e., (2) $Y(t)$ is right continuous a.e., (3) $\lim_{t\to 0} Y(t) = 0$ a.e., (4) $\lim_{t\to\infty} Y(t) = \infty$ a.e. We will denote the moment generating function of $Y(t)$ by $M_t(a) = E\big(\exp\big(-aY(t)\big)\big)$.

Note that, for each fixed $t > 0$, $S(t)$ is a random variable which can be represented by $\theta_t \in [0,1]$, and whose probability distribution function can be obtained as a marginal distribution of the N.R.P. prior. Let X^t be a random variable such that $X^t = 1$ if $T > t$ and $X^t = 0$ if $T \leqslant t$, then X^t has a Bernoulli distribution with parameter θ_t. After observing a sample $T_1, \dots, T_n$ from the random variable T, we have immediately a sample $X_1^t, \dots, X_n^t$ of X^t. Consequently, for each fixed $t > 0$, we have a parametric Bayesian estimation problem where we observe a random sample from a Bernoulli (θ_t) distribution and θ_t is a random variable whose prior distribution is a marginal distribution from the N.R.P. prior.

Under these hypotheses, we state the problem of sequentially estimating θ_t with quadratic error loss, $L(\theta_t, \hat{\theta}_t) = (\theta_t - \hat{\theta}_t)^2$, and constant cost function $c > 0$. So the problem is to find a stopping rule ψ, and a terminal decision rule $\hat{\theta}_t$, such that $(\psi, \hat{\theta}_t)$ is Bayes with respect to the prior distribution of θ_t. See Ferguson [3] for elementary facts about sequential and Bayesian decision theory.

For each positive integer n, the Bayes terminal decision rule is the Bayes rule in the problem with fixed size sample $X_1^t, \ldots, X_n^t$. The likelihood function of $X_1^t, \ldots, X_n^t$ is

$$f(x_1^t, \ldots, x_n^t \mid \theta_t) = \theta_t^{s_n^t}(1-\theta_t)^{n-s_n^t},$$

where $s_n^t = \sum_{i=1}^n x_i^t$ is the realization of the sufficient statistic $nS_n(t) = \sum_{i=1}^n X_i^t$.

For ease of exposition we omit index t when it is not essential. More concretely, we will write θ instead of θ_t in what follows. Let $g(\theta)$ be the prior distribution of θ, then the posterior distribution is

$$g(\theta \mid x_1^t, \ldots, x_n^t) = g(\theta \mid s_n^t) = \frac{g(\theta) f(x_1^t, \ldots, x_n^t \mid \theta)}{\int_0^1 g(\theta) f(x_1^t, \ldots, x_n^t \mid \theta)\, \mathrm{d}\theta},$$

and the Bayes rule for fixed size sample under squared-error loss is

$$E(\theta \mid x_1^t, \ldots, x_n^t) = E(\theta \mid s_n^t) = \frac{\int_0^1 \theta g(\theta) f(x_1^t, \ldots, x_n^t \mid \theta)\, \mathrm{d}\theta}{\int_0^1 g(\theta) f(x_1^t, \ldots, x_n^t \mid \theta)\, \mathrm{d}\, \theta} = \frac{T(n, s_n^t, 1, t)}{T(n, s_n^t, 0, t)},$$

where

$$T(n, s_n^t, k, t) = \sum_{j=0}^{n-s_n^t} (-1)^{n-s_n^t-j} \binom{n-s_n^t}{j} M_t(n-j+k), \quad k = 0, 1, 2, \ldots .$$

The expression $\hat{S}_n(t) = T(n, s_n^t, 1, t)/T(n, s_n^t, 0, t)$ is a survival function when t varies in $(0, \infty)$ as can be seen in Morales, Quesada and Pardo [6].

For some known N.R.P., the expressions of $\hat{S}_n(t)$ are:

(a) For Dirichlet process with parameter $\alpha(t)$:

$$M_t(a) = \frac{\Gamma(M)\Gamma(M - \alpha(t) + a)}{\Gamma(M - \alpha(t))\Gamma(M + a)},$$

$$\hat{S}_n(t) = \frac{M}{M+n} S_0(t) + \frac{n}{M+n} S_n(t), \quad \text{where } M = \alpha(+\infty)$$

and

$$S_0(t) = E(S(t)) = \frac{M - \alpha(t)}{M}.$$

$\hat{S}_n(t)$ is equal to the Bayes estimator. See Ferguson [4].

(b) For Gamma-exponential process with parameter $(\Lambda(t), c)$:

$$M_t(a) = \left(1 + \frac{a}{c}\right)^{-c\Lambda(t)},$$

$$\hat{S}_n(t) = \frac{\sum_{j=0}^{n-s_n^t}(-1)^{n-s_n^t-j}\binom{n-s_n^t}{j}\left(1+(n+1-j)/c\right)^{-c\Lambda(t)}}{\sum_{j=0}^{n-s_n^t}(-1)^{n-s_n^t-j}\binom{n-s_n^t}{j}\left(1+(n-j)/c\right)^{-c\Lambda(t)}}, \quad c > 0.$$

(c) For Homogeneous simple process with parameter $(\Lambda(t), b)$:

$$M_t(a) = \exp\left\{-b\Lambda(t)\sum_{i=0}^{a-1}\frac{1}{i+b}\right\}, \quad a \in \mathbf{N},\ b > 0,$$

$$\hat{S}_n(t) = \frac{\sum_{j=0}^{n-s_n^t}(-1)^j\binom{n-s_n^t}{j}\exp\left(-b\Lambda(t)\sum_{i=0}^{n-j}\frac{1}{i+b}\right)}{\sum_{j=0}^{n-s_n^t}(-1)^j\binom{n-s_n^t}{j}\exp\left(-b\Lambda(t)\sum_{i=0}^{n-j-1}\frac{1}{i+b}\right)}.$$

For a fixed $t > 0$, the posterior risk is

$$\varrho_n(x_1^t, \ldots, x_n^t, t) = \varrho_n(s_n^t, t) = E\left(\theta^2/nS_n(t) = s_n^t\right) - \left[E\left(\theta/nS_n(t) = s_n^t\right)\right]^2 = \\ = \frac{T(n, s_n^t, 2, t)}{T(n, s_n^t, 0, t)} - \left(\frac{T(n, s_n^t, 1, t)}{T(n, s_n^t, 0, t)}\right)^2.$$

In order to obtain global estimators, in the sense of not depending on t, we give the following definition:

Definition 2.1. The global posterior parametric Bayes risk is

$$R_n(t_1, \ldots, t_n) = \int_0^\infty \varrho_n(s_n^t, t)\,\mathrm{d}W(t),$$

where $(t_1, \ldots, t_n)$ is a realization of $(T_1, \ldots, T_n)$ and $W(t)$ is a finite measure on $\mathbf{R}^+$ not having common atoms with $S_0(t) = E\big(S(t)\big)$.

Our purpose is to obtain a stopping rule of the kind 1-SLA; i.e., such a rule for stopping or continuing sampling, which is optimal among those rules taking at most one more observation at each stage. We need to calculate, for each $t > 0$, the expected Bayes risk when $x_1^t, \ldots, x_n^t$ has been observed and when it has been decided to observe the next X_{n+1}^t. The marginal density function of $X_1^t, \ldots, X_n^t$ is

$$f(x_1^t, \ldots, x_n^t) = \int_0^1 f(x_1^t, \ldots, x_n^t \mid \theta)g(\theta)\,\mathrm{d}\theta = \int_0^1 \theta^{s_n^t}(1-\theta)^{n-s_n^t}g(\theta)\,\mathrm{d}\theta = \\ = T(n, s_n^t, 0, t),$$

and the marginal density of $X_{n+1}^t \mid X_1^t = x_1^t, \ldots, X_n^t = x_n^t$ is

$$f(x_{n+1}^t \mid x_1^t, \ldots, x_n^t) = \frac{f(x_1^t, \ldots, x_{n+1}^t)}{f(x_1^t, \ldots, x_n^t)} = \frac{T(n+1, s_{n+1}^t, 0, t)}{T(n, s_n^t, 0, t)} \quad \text{if } x_{n+1}^t = 1,$$

and

$$f(x_{n+1}^t \mid x_1^t, \ldots, x_n^t) = \frac{T(n+1, s_n^t, 0, t)}{T(n, s_n^t, 0, t)} \quad \text{if } x_{n+1}^t = 0.$$

By taking expectations with respect to $f(x_{n+1}^t \mid x_1^t, \ldots, x_n^t)$, we define, for a fixed $t > 0$, the expected parametric Bayes risk

$$\begin{aligned} A(t_1, \ldots, t_n, t) &= E\big[\varrho_n(X_1^t, \ldots, X_n^t, X_{n+1}^t, t) \mid X_1^t = x_1^t, \ldots, X_n^t = x_n^t\big] = \\ &= \frac{T(n+1, s_n^t, 2, t)T(n+1, s_n^t, 0, t) - T^2(n+1, s_n^t, 1, t)}{T(n+1, s_n^t, 0, t)T(n, s_n^t, 0, t)} + \\ &+ \frac{T(n+1, s_n^t+1, 2, t)T(n+1, s_n^t+1, 0, t) - T^2(n+1, s_n^t+1, 1, t)}{T(n+1, s_n^t+1, 0, t)T(n, s_n^t, 0, t)}. \end{aligned}$$

Now, we define the global expected parametric Bayes risk, when $t_1, \ldots, t_n$ has been observed and when it has been decided to observe the next T_{n+1}, in the following way:

$$RE_n(t_1, \ldots, t_n) = \int_0^\infty A(t_1, \ldots, t_n, t)\, \mathrm{d}W(t),$$

which is smaller than or equal to $R_n(t_1, \ldots, t_n)$, because for every fixed $t \geqslant 0$, we have a different sequential decision problem to estimate the one dimensional parameter $S(t)$. Therefore, we get

$$\varrho(x_1^t, \ldots, x_n^t, t) \geqslant E\big[\varrho(x_1^t, \ldots, x_n^t, X_{n+1}^t, t) \mid X_1^t = x_1^t, \ldots, X_n^t = x_n^t\big],$$

for every fixed $t \geqslant 0$ (see Ferguson [3], chapter 7). Finally, by integrating on $(0, \infty)$ with respect to $W(t)$, the statement follows. Intuitively, the meaning is that the more observations we have the smaller risk we get. We now give the following definition:

Definition 2.2. If $T_1 = t_1, \ldots, T_n = t_n$ has been observed, we define 1-SPLA stopping rule as the rule that stops sampling at stage n if and only if

$$R_n(t_1, \ldots, t_n) - RE_n(t_1, \ldots, t_n) \leqslant c.$$

Intuitively, 1-SPLA rule stops sampling when, through the infinitely many decision problems for each $t > 0$, the mean risk decrease is smaller than the cost of looking at a new observation. Furthermore, as

$$\lim_{n\to\infty} R_n(t_1, \ldots, t_n) = 0 \text{ a.e.} \quad \text{and} \quad R_n(t_1, \ldots, t_n) - RE_n(t_1, \ldots, t_n) \geqslant 0 \text{ a.e.},$$

1-SPLA rule stops with probability one. However, we can not state that 1-SPLA rule equals to Bayes rule; for this it is necessary that if 1-SPLA rule stops at stage n_0, then for each posterior stage n:

$$R_n(t_1,\ldots,t_n) - RE_n(t_1,\ldots,t_n) \leqslant c.$$

3. 1-SPLA estimation with D.P. prior

In this section we prove that if the N.R.P. is a D.P. then 1-SPLA rule equals to 1-SLA rule. Furthermore, we prove that 1-SLA rule and Bayes rule (∞-SLA) can be found among the truncated rules at a certain fixed stage n_0. Analyzing the risks that appear in 1-SPLA and 1-SLA method, we observe that integral symbols can be interchanged because we have bounded nonnegative measurable functions and finite measures.

Let $S(t)$, $t \geqslant 0$, be a D.P. with parameter $\alpha(\cdot)$, where $\alpha(\cdot)$ is a finite non-null measure on $\mathbf{R}_+$ and $M = \alpha(\mathbf{R}_+)$. Let $\mathscr{P}$ be the probability measure given by the D.P. We write $\mathscr{P} \in \mathscr{D}(\alpha(\cdot))$, and similarly, we will denote by $\mathscr{B}e(\cdot,\cdot)$ the family of Beta distributions. For each $t \geqslant 0$, $\alpha(t) = \alpha((0,t]) = MS_0(t)$ and $S(t) \in \mathscr{B}e(M - \alpha(t), \alpha(t))$; furthermore, after observing $T_1 = t_1,\ldots,T_n = t_n$, $\mathscr{P}_{t_1,\ldots,t_n} \in \mathscr{D}(\alpha(\cdot) + \sum_{i=1}^{n} \delta_{t_i})$, where δ_{t_i} is a probability measure giving mass one to the point t_i. Finally, for each $t \geqslant 0$

$$S(t)\Big|_{(T_1=t_1,\ldots,T_n=t_n)} \in \mathscr{B}e\big(M - \alpha(t) + nS_n(t), \alpha(t) + n - nS_n(t)\big).$$

Using the loss function (1.1), we obtain the following expression for the Bayes risk:

$$R_n(\mathscr{P},\hat{S}_n) = \int_{\mathbf{R}_+}\cdots\int_{\mathbf{R}_+}\Big(\int_0^1 \big(S(t)-\hat{S}_n(t)\big)^2\,\mathrm{d}\mathscr{P}_{t_1,\ldots,t_n}\big(S(t)\big)\Big)\,\mathrm{d}W(t)\,\mathrm{d}P(t_1,\ldots,t_n),$$

where $\hat{S}_n(t)$ is the Bayes estimate of $S(t)$ based on $(t_1,\ldots,t_n)$ and $\mathrm{d}P(t_1,\ldots,t_n)$ is the unconditional density of the random sample.

The Bayes stopping rule minimizes the inner integral, then

$$\hat{S}_n(t) = E\big(S(t) \mid T_1 = t_1,\ldots,T_n = t_n\big) = \frac{M}{M+n}S_0(t) + \frac{n}{M+n}S_n(t).$$

Consequently, after observing $T_1 = t_1$, ..., $T_n = t_n$, the posterior Bayes risk is

$$r_n(t_1,\ldots,t_n) = \int_{\mathbf{R}_+} \mathrm{Var}\,\big(S(t) \mid T_1 = t_1,\ldots,T_n = t_n\big)\,\mathrm{d}W(t).$$

Furthermore

$$R_n(t_1,\ldots,t_n) = \int_{\mathbf{R}_+} \varrho_n(x_1^t,\ldots,x_n^t,t)\,\mathrm{d}W(t) =$$
$$= \int_{\mathbf{R}_+} \Big(\int_0^1 (S(t)-\hat{S}_n(t))^2(x_1^t,\ldots,x_n^t)\,\mathrm{d}\mathscr{P}_{x_1^t,\ldots,x_n^t}(S(t))\Big)\,\mathrm{d}W(t),$$

where $\hat{S}(t)(x_1^t,\ldots,x_n^t)$ is the parametric Bayes estimate of $S(t)$ based on $(x_1^t,\ldots,x_n^t)$ and $\mathrm{d}\mathscr{P}_{x_1^t,\ldots,x_n^t}(S(t))$ is the distribution of the marginal random variable at the time t, when it is calculated according to the parametric method, i.e.,

$$\mathrm{d}\mathscr{P}_{x_1^t,\ldots,x_n^t}(S(t)) = g(\theta \mid x_1^t,\ldots,x_n^t)\,\mathrm{d}\theta, \quad \text{with } \theta = S(t).$$

Now, as $\mathscr{P} \in \mathscr{D}(\alpha(\cdot))$, then

$$\mathrm{d}\mathscr{P}_{x_1^t,\ldots,x_n^t}(S(t)) = \mathrm{d}\mathscr{P}_{t_1,\ldots,t_n}(S(t)), \quad \hat{S}_n(t) = \hat{S}(t)(x_1^t,\ldots,x_n^t)$$

and

$$r_n(t_1,\ldots,t_n) = R_n(t_1,\ldots,t_n) = \frac{1}{M+n+1}\int_{\mathbf{R}_+} \hat{S}_n(t)(1-\hat{S}_n(t))\,\mathrm{d}W(t)$$
$$= \frac{1}{(M+n)^2(M+n+1)}\int_{\mathbf{R}_+} (M-\alpha(t)+nS_n(t))(\alpha(t)+n-nS_n(t))\,\mathrm{d}W(t).$$

Similarly $RE_n(t_1,\ldots,t_n)$ can be obtained either by 1-SPLA method or by 1-SLA method; hence, both methods give the same stopping rule. To finish this section, we find a bound for the Bayes stopping rule. First, we observe that the loss function is bounded in the decision problem with $t > 0$ fixed. So $\lim_{n\to\infty} V_0^{(n)} = V_0^{(\infty)}$ (Ferguson [3]), i.e. the Bayes risk of the truncated problem at stage n approximates the Bayes risk of the sequential problem. Furthermore, it is easy to prove that for each $(t_1,\ldots,t_n)$

$$R_n(t_1,\ldots,t_n) - RE_n(t_1,\ldots,t_n) =$$
$$= \frac{1}{(M+n+1)^2}\int_{\mathbf{R}_+} \hat{S}_n(t)(1-\hat{S}_n(t))\,\mathrm{d}W(t) \leqslant \frac{k}{4(M+n+1)^2},$$

where $k = W(\mathbf{R}_+)$ is a positive constant proportionally related to the severity of the errors in estimating $S(t)$.

The problem can be truncated at the stage n such that

$$\frac{k}{4(M+n+1)^2} \leqslant c. \tag{3.1}$$

Finally, conditions for rules 1-SLA to be equal to Bayes rules can be found in Ferguson [5].

4. Asymptotically pointwise optimal (A.P.O.) procedures in sequential analysis with N.R.P.

Bickel and Yahav [1] give the following definitions and results:

Definition 4.1 (B.Y.). Let Y_n, $n \in \mathbf{N}$, be a sequence of random variables on a probability space (Ω, F, P), where Y_n is F_n-measurable and $F_n \subseteq F_{n+1} \subseteq F$ for $n \geqslant 1$. Let $P(Y_n > 0) = 1$, $\lim_{n\to\infty} Y_n = 0$ a.e. and $X_n(c) = Y_n + nc$. Let $\mathscr{T}$ be the class of all stopping times defined on the σ-fields F_n. Abusing our notation, in a fashion long used in large sample theory, use the words "stopping rule" to denote also a function $l(c)$ belonging to the class Q of functions from $(0,\infty)$ to $\mathscr{T}$. We say $q(c) \in Q$ is A.P.O. if and only if for any other $l(c) \in Q$

$$\limsup_{c\to 0} \left(X_{q(c)}(c)/X_{l(c)}(c)\right) \leqslant 1 \text{ a.e.}$$

Theorem 4.2 (B.Y.). *If conditions of Definition* 4.1 *hold and* $\lim_{n\to\infty} nY_n = V$ *a.e., where* V *is a random variable such that* $P(V > 0) = 1$, *then the stopping rule, which is determined by "stop the first time* $(Y_n/n) \leqslant c$*" is A.P.O.*

Our purpose is to adapt these results to the parametric model that we have given to estimate sequentially a survival function with D.P. prior. Let $(\Omega, \mathscr{A}, P)$ be a probability space; $\omega \in \Omega$ determines P-a.e. a survival function $S = S_\omega$, which is a sample of the D.P., and a sample realization; i.e., $(T_1, \ldots, T_n)(\omega) = (t_1, \ldots, t_n)$. Let $(\mathbf{R}_+, \mathscr{B}_+, W)$ be a measurable space where $\mathscr{B}_+$ is the Borel σ-field on $\mathbf{R}_+$ and W has been defined in (1.1).

Analyzing the decision problem for a fixed $t > 0$ and noting that $\theta = S(t)$ and $X_i^t = I_{(t,\infty)}(t_i)$, we observe that $f(x_i^t \mid \theta) = \theta^{x_i^t}(1-\theta)^{1-x_i^t}$ is the Bernoulli probability function with parameter $\theta = S(t)$, which belongs to the one parameter exponential family. From the properties of the Beta distribution, we have that

$$S(t)\Big|_{(X_1^t = x_1^t, \ldots, X_n^t = x_n^t)} \in \mathscr{B}e(M - \alpha(t) + nS_n(t), \alpha(t) + n - nS_n(t))$$

and

$$\begin{aligned} \varrho_n(X_1^t, \ldots, X_n^t, t) &= \operatorname{Var}\left(S(t) \mid X_1^t, \ldots, X_n^t\right) \\ &= \frac{(M - \alpha(t) + nS_n(t))(\alpha(t) + n - nS_n(t))}{(M+n)^2(M+n-1)}. \end{aligned}$$

Now, from the properties of the Dirichlet processes, we have that $\mathscr{P}_{t_1,\dots,t_n} \in \mathscr{D}\big(\alpha(\cdot)+\sum_{i=1}^{n}\delta_{t_i}\big)$, where δ_{t_i} is a probability measure giving mass one to the point t_i. Then, we also have that

$$S(t)\Big|_{(T_1=t_1,\dots,T_n=t_n)} \in \mathscr{B}e\big(M-\alpha(t)+nS_n(t), \alpha(t)+n-nS_n(t)\big)$$

and

$$\operatorname{Var}\big(S(t)\mid T_1,\dots,T_n\big) = \frac{\big(M-\alpha(t)+nS_n(t)\big)\big(\alpha(t)+n-nS_n(t)\big)}{(M+n)^2(M+n-1)}.$$

Let us define $Y_n(t,\omega) = \operatorname{Var}\big(S(t)\mid T_1(\omega),\dots,T_n(\omega)\big)$; hence, $P(N_t)=0$, for each $t>0$, where N_t is the section of

$$N = \Big\{(t,\omega)\in \mathbf{R}_+\times\Omega \mid nY_n(t) \not\to \big(S(t)(1-S(t))(\omega)\big)\Big\},$$

which belongs to the product σ-field $\mathscr{B}_+\times\mathscr{A}$. Consequently, $(W\times P)(N)=0$.

Observe that there exists a set $D\in\mathscr{A}$ such that $P(D)=1$, and for each $\omega\in D$, $S(\cdot)(\omega)$ is a survival function. So, the following results are obtained for any $\omega\in D$:

(a) $\lim_{n\to\infty} nY_n(t,\omega) = \big[S(t)(1-S(t))\big](\omega)\ \forall t\geqslant 0$.

(b) $\big[S(t)(1-S(t))\big](\omega)\leqslant \frac{1}{4}\ \forall t\geqslant 0$, hence $\big[S(\cdot)(1-S(\cdot))\big](\omega)$ is W-integrable.

We now prove that for each $\omega\in D$, $\{nY_n(t)\}_{n\in\mathbf{N}}$ is uniformly bounded.

Lemma 4.1. *For each $\omega\in D$, $n\geqslant 2$, and $t\geqslant 0$, we have that $nY_n(t,\omega)\leqslant 2$.*

Proof. The result follows from the inequality

$$nY_n(t,\cdot) = \frac{n}{M+n-1}\,\frac{[M-\alpha(t)]+nS_n(t)}{M+n}\,\frac{\alpha(t)+n[1-S_n(t)]}{M+n}\leqslant 2.$$

□

Theorem 4.3. *Let us define $Y_n(\omega) = \int_0^\infty Y_n(t,\omega)\,\mathrm{d}W(t)$. For any $\omega\in D$, we have*

$$\lim_{n\to\infty} nY_n(\omega) = \int_0^\infty \big(S(t)(1-S(t))(\omega)\big)\,\mathrm{d}W(t).$$

Proof. Let $\omega\in D$ fixed. As the sequence $nY_n(t,\omega)$ is uniformly bounded, applying the dominated convergence theorem we obtain the result. □

Consequently, by Theorem 4.2, in sequential sampling with N.R.P. prior, the rule "stop at the stage n such that:

$$\frac{1}{n}\int_0^\infty \operatorname{Var}\big(S(t)\mid T_1,\dots,T_n\big)\,\mathrm{d}W(t)\leqslant c"$$

is A.P.O.

5. An example

The purpose of this section is to show an application of the Bayes and 1-SPLA=1-SLA procedures to a particular case where the prior distribution on the space of survival functions is given by a Dirichlet process. We suppose: $S_0(t) = 1 - t$, $W(t) = 40t$, $\alpha(t) = t$, $t \in [0,1]$, $M = \alpha(\mathbf{R}) = 1$, $k = W(\mathbf{R}) = 40$, $c = 1$. Let us define $F(t) = 1 - S(t)$.

First we obtain the Bayes stopping rule. To do this, we make some previous calculations. We consider a sequential random sample $T_1, \ldots, T_n$ drawn at random according to the above Dirichlet process $\{F(t)\}_{t \geqslant 0}$. Remember that the cumulative distribution function of $T_1, \ldots, T_n$ is

$$P(T_1 \leqslant t_1, \ldots, T_n \leqslant t_n) = E\big(F(t_1), \ldots, F(t_n)\big), \quad n \in \mathbf{N}.$$

Distribution of $T_2/T_1 = t_1$.
If $0 < t_1 < t_2 < 1$, then

$$(Y_1, Y_2, Y_3) = \big(F(t_1), F(t_2) - F(t_1), 1 - F(t_2)\big) \in \mathscr{D}(Mt_1, M(t_2 - t_1), M(1 - t_2))$$

and

$$f_{Y_1,Y_3}(y_1, y_3) = \frac{\Gamma(M) y_1^{Mt_1 - 1} y_2^{M(t_2 - t_1) - 1} y_3^{M(1-t_2)-1}}{\Gamma(Mt_1)\Gamma\big(M(t_2 - t_1)\big)\Gamma\big(M(1 - t_2)\big)}$$

in the set $S = \big\{(y_1, y_3) \in \mathbf{R}^2 \mid y_1 \geqslant 0,\ y_3 \geqslant 0,\ y_1 + y_3 \leqslant 1\big\}$, where $y_2 = 1 - y_1 - y_3$, and zero otherwise. Also observe that

$$E(Y_1 Y_3) = E\Big(F(t_1)\big(1 - F(t_2)\big)\Big) = \frac{M}{M+1} t_1(1 - t_2),$$

$$P(T_1 \leqslant t_1) = E\big(F(t_1)\big) = t_1 \quad \text{and} \quad P(T_1 \leqslant t_1, T_2 \leqslant t_2) = E\big(F(t_1)F(t_2)\big).$$

Consequently for any $(t_1, t_2) \in \mathbf{R}_+^2$, we obtain

$$\begin{aligned} P(T_1 \leqslant t_1, T_2 \leqslant t_2) &= 0 && \text{if } t_1 < 0 \text{ or } t_2 < 0, \\ &= \frac{t_1(1 + Mt_2)}{M+1} && \text{if } 0 \leqslant t_1 \leqslant t_2 < 1, \\ &= \frac{t_2(1 + Mt_1)}{M+1} && \text{if } 0 \leqslant t_2 \leqslant t_1 < 1, \\ &= t_1 && \text{if } 0 \leqslant t_1 < 1,\ t_2 \geqslant 1, \\ &= t_2 && \text{if } 0 \leqslant t_2 < 1,\ t_1 \geqslant 1, \\ &= 1 && \text{if } t_1 \geqslant 1,\ t_2 \geqslant 1. \end{aligned}$$

Furthermore, for $M = 1$ and any $t_1 \in (0,1)$, we have that $f(t_2/t_1) = \frac{1}{2}I_{(0,1)}(t_2)$ and $P(T_2 = t_2/T_1 = t_1) = \frac{1}{2}$ if $t_2 = t_1$, i.e., $T_2/T_1 = t_1$ is a mixed random variable giving probability $\frac{1}{2}$ to the point t_1 and with a uniform absolutely continuous part in $(0,1)$.

Truncation rule.

Truncation is at the first n verifying $\frac{k}{4(M+n+1)^2} \leqslant c$. In this example we obtain $n = 2$.

Calculation of $\varrho_2(x_1^t, x_2^t, t)$.

Remember that

$$\varrho_n(x_1^t, \dots, x_n^t, t) = \frac{\big(nS_n(t) + M - \alpha(t)\big)\big(n + \alpha(t) - nS_n(t)\big)}{(M+n)^2(M+n+1)} = $$
$$= \frac{\big(nS_n(t) + 1 - t\big)\big(n + t - nS_n(t)\big)}{(n+1)^2(n+2)},$$

where $nS_n(t) = \sum_{i=1}^{n} X_i^t$. Now , we get that

if $t \geqslant t_2 > t_1$, then $x_1^t = x_2^t = 0$ and $\varrho_2(0,0,t) = \frac{-t^2-t+2}{36}$,
if $t_2 > t \geqslant t_1$, then $x_1^t = 0$, $x_2^t = 1$ and $\varrho_2(0,1,t) = \frac{-t^2+t+2}{36}$,
if $t_2 > t_1 > t$, then $x_1^t = x_2^t = 1$ and $\varrho_2(1,1,t) = \frac{-t^2+3t}{36}$.

Calculation of $R_2(t_1,t_2)$ when $t_1 < t_2$.

$R_2(t_1,t_2) = \int_0^1 \varrho_2(x_1^t, x_2^t, t)\,\mathrm{d}W(t) = \frac{40}{36}\Big[\int_0^{t_1}(-t^2+3t)\,\mathrm{d}t + \int_{t_1}^{t_2}(-t^2+t+2)\,\mathrm{d}t + \int_{t_2}^{1}(-t^2-t+2)\,\mathrm{d}t\Big] = \frac{10}{9}(t_1^2 + t_2^2 - 2t_1 + \frac{7}{6}) = \text{(a)}$.

Symmetrically, if $t_2 < t_1$, then $R_2(t_1,t_2) = \frac{10}{9}(t_2^2 + t_1^2 - 2t_2 + \frac{7}{6}) = \text{(b)}$.

Table for $n = 2$ (see nomenclature in Ferguson [3]).

(T_1,T_2)	$\hat{S}_2(t)$	$R_2(t_1,t_2)$	$V_2^{(2)} = U_2(t_1,t_2)$
$t_1 < t_2$	$\frac{1}{3}(1-t) + \frac{2}{3}S_2(t)$	(a)	$R_2(t_1,t_2) + 2$
$t_1 > t_2$	$\frac{1}{3}(1-t) + \frac{2}{3}S_2(t)$	(b)	$R_2(t_1,t_2) + 2$

Calculation of $E(V_2^{(2)} \mid T_1 = t_1)$ (nonparametric method).

$$E(V_2^{(2)} \mid T_1 = t_1) = 2 + P(T_2 = t_1 \mid T_1 = t_1)\varrho_2(t_1,t_2) + \int_0^{t_1} \frac{(a)}{2}\,\mathrm{d}t =$$
$$= 2 + \frac{20}{9}\Big(t_1^2 - t_1 + \frac{2}{3}\Big) = \text{(c)}.$$

Observe that for making this last calculation we need to know the distributions of the variables $T_n \mid T_1 = t_1, \ldots, T_{n-1} = t_{n-1}$. This point represents a serious obstacle for large values of n. Alternatively, an easier way of making this calculation is given by 1-SPLA procedure.

Calculation of $E(V_2^{(2)} \mid T_1 = t_1)$ (Parametric method).
It is easy to prove that

$$P(X_n^t = 0 \mid X_1^t = x_1^t, \ldots, X_{n-1}^t = x_{n-1}^t) = \frac{\alpha(t) + n - 1 - (n-1)S_{n-1}(t)}{M + n - 1}.$$

Furthermore

$$E(V_2^{(2)} \mid T_1 = t_1) = 2 + E\big(\varrho_2(X_1^t, X_2^t) \mid X_1^t = x_1^t\big) = 2 + (\mathrm{d}).$$

Now, we obtain that
if $t < t_1$, then $E\big(\varrho_2(1, X_2^t, t) \mid X_1^t = 1\big) = P(X_2^t = 1 \mid X_1^t = 1)\varrho(1,1,1) + P(X_2^t = 0 \mid X_1^t = 1)\varrho(1,0,t) = (-t^2 + 2t)/18 = (\mathrm{e})$, and
if $t_1 < t$, then $E\big(\varrho_2(0, X_2^t, t) \mid X_1^t = 0\big) = P(X_2^t = 1 \mid X_1^t = 0)\varrho(0,1,t) + P(X_2^t = 0 \mid X_1^t = 0)\varrho(0,0,t) = (-t^2 + 1)/18 = (\mathrm{f})$.
Hence, $(\mathrm{d}) = 40\big[\int_0^{t_1}(\mathrm{e})\,\mathrm{d}t + \int_{t_1}^1 (\mathrm{f})\,\mathrm{d}t\big] = \frac{20}{9}(t_1^2 - t_1 + \frac{2}{3})$.
Finally, $E(V_2^{(2)} \mid T_1 = t_1) = (\mathrm{c})$.

Calculation of $\varrho_1(x_1^t, t)$, $R_1(t_1)$ and $U_1(t_1)$.
If $t \leqslant t_1$, then $\varrho_1(1,t) = (-t^2 + 2t)/12$.
If $t > t_1$, then $\varrho_1(0,t) = (-t^2 + 1)/12$.
$R_1(t_1) = \int_0^1 \varrho_1(x_1^t, t)\,\mathrm{d}W(t) = \frac{40}{12}\big[\int_0^{t_1}(-t^2 + 2t)\,\mathrm{d}t + \int_{t_1}^1(-t^2 + 1)\,\mathrm{d}t\big] = \frac{10}{3}(t^2 - t_1 + \frac{2}{3})$.
$U_1(t_1) = \varrho_1(t_1) + (\mathrm{c}) = \frac{10}{3}(t_1^2 - t_1 + \frac{2}{3}) + 1 = (\mathrm{g})$.

Table for $n = 1$.

T_1	$\hat{S}_1(t)$	$\varrho_1(t_1)$	$U_1(t_1)$	$E(V_2^{(2)} \mid T_1 = t_1)$	$V_1^{(2)}$	φ_1^0
t_1	$\frac{(1-t)}{2} + \frac{S_1(t)}{2}$	(g) − 1	(g)	(c)	$U_1(t)$	1 (STOP)

As (g) < (c) because $t_1 \in (0,1)$, we stop sampling at stage $n = 1$.

Table for $n = 0$.

Data	$\hat{S}_0(t)$	ϱ_0	U_0	$E(V_1^{(2)})$	$V_0^{(2)}$	φ_1^0
—	$1 - t$	3.33	3.33	2.11	2.11	0 (CONTINUE)

Conclusion.

The Bayes sequential rule is a fixed sample size rule with $n = 1$, and the Bayes risk of the example is 2.11.

We now solve the same problem using 1-SPLA procedure:

If $n = 0$, then we have to stop sampling if and only if

$$\frac{1}{M+1}\int_0^1 S_0(t)\big(1 - S_0(t)\big)\,\mathrm{d}W(t) \leqslant c(M+1).$$

As the numerical values of the right and left hand sizes are $\frac{10}{3}$ and 2 respectively and $\frac{10}{3} > 2$, we look at a new observation.

If $n = 1$, then we have to stop sampling if and only if

$$\frac{1}{M+n+1}\int_0^1 \hat{S}_n(t)\big(1 - \hat{S}_n(t)\big)\,\mathrm{d}W(t) \leqslant c(M+n+1);$$

i.e., if and only if $\frac{10}{3}\big(t_1^2 - t_1 + \frac{2}{3}\big) \leqslant 3$. This last inequality is true if $t_1 \in (0,1)$; hence stop sampling.

Finally, the optimality condition given by Ferguson [5] for the 1-SLA rule, with Dirichlet processes, to be a Bayes rule can be easily checked. So, in this example, 1-SLA rule is equal to Bayes rule. We already know this fact as we have calculated the Bayes rule earlier.

References

[1] *P. J. Bickel, J. A. Yahav*: Asymptotically pointwise optimal procedures in sequential analysis. Proc. Fifth Berkeley Symp. Math. Statist. Prob. I (1965), 401–413.

[2] *K. A. Doksum*: Tailfree and neutral random probabilities and their posterior distributions. Ann. Probability *2* (1974), 183–201.

[3] *T. S. Ferguson*: Mathematical Statistics. Academic Press, 1967.

[4] *T. S. Ferguson*: A Bayesian analysis of some nonparametric problems. Ann. Statistics *1* (1973), 209–230.

[5] *T. S. Ferguson*: Sequential estimations with Dirichlet process priors. Statistical Theory and Related Topics III *1* (1982), 385–401.

[6] *D. Morales, L. Pardo, V. Quesada*: Estimación paramétrica bayesiana no paramétrica de funciones de supervivencias con observaciones parcialmente censuradas. Trabajos de Estadística *1* (1986), 70–87.

Authors' address: Departamento de Estadística e I.O., Facultad de Matemáticas, Universidad Complutense, Madrid 28040, Spain.