REVISTA PROYECCIONES Nº 15: 49-74
Diciembre 1988 - ISSN 0716-0917

ESTIMACION BAYESIANA DE FUNCIONES DE DISTRIBUCION: APROXIMACION PARAMETRICA

SEGUNDO ESCALIER SOTO*

GUILLERMO MONDACA ORTIZ*

RESUMEN.

En el planteamiento del problema de decisión estadística, se incorpora el conocimiento a priori o enfoque bayesiano por el par $(\mathcal{S}, P)$. Donde $\mathcal{S}$ es el sistema o forma de definir la medida de probabilidad a priori P , sobre los estados de naturaleza. Enseguida se desarrolla un método alternativo en la búsqueda de una solución al problema de estimación en la estadística bayesiana no paramétrica, denominado "Aproximación Paramétrica". Esta solución paramétrica comprende la solución que se obtiene en el denominado Modelo de Azar Proporcional. Además, cuando se utiliza como conocimiento a priori un proceso de Dirichlet, el método paramétrico genera la misma solución obtenida por el método no paramétrico empleado por Fer

* Académico Departamento Matemáticas, Facultad de Ciencias Básicas, Universidad de Antofagasta - Chile.

•

gurson, sin tener que deducir la distribución a posteriori del proceso. Como conclusión, se observa que el problema de estimar una función de Distribución (o función de Supervivencia) desde el punto de vista bayesiano, es transformado en el problema de tener que probar una relación de monotonía, por lo tanto, se tienen más recursos o técnicas disponibles para abordar esta situación sin tener que deducir la distribución a posteriori.

PALABRAS CLAVES:

Estimación Bayesiana de una Función de Distribución. Probabilidad Aleatoria. Proceso de Dirichlet. Estimación Paramétrica.

I. INTRODUCCION.

Desarrollaremos la situación en la cual se requiere estimar una función de supervivencia asociada a una v.a. $T \geq 0$. Hecho que no resta generalidad alguna.

Presentaremos en un sentido pedagógico la solución para el problema sin información muestral y luego abordaremos el problema con información muestral. También se observará que el Método de Aproximación Paramétrica es válido cuando existe información muestral censurada a la derecha. Puesto que lo que es determinante para construir el estimador propuesto, es sólo el número de observaciones a la derecha de cada instante t .

La idea fundamental del método paramétrico es observar el problema original, como una colección de problemas y obtener una solución para el problema "Proyectado" en un instante t , fijo. Luego, analizar si al variar t obtenemos una solución del problema propuesto. De ser así, denominaremos a ésta "Solución Paramétrica del Problema" o "Aproximación Paramétrica de la Solución". Plantearemos la demostración general cuando se utiliza una supervivencia aleatoria denominada "Neutral por la Derecha, Ho

mogénea". Como ejemplo particular se agrega una situación en que se utiliza un proceso Neutral por la Derecha no Homogéneo, tal como el proceso de Dirichlet, obteniendo la misma solución deducida por Ferguson, sin tener que obtener la distribución a posteriori del proceso.

II. EL METODO PARAMETRICO.

La búsqueda de un método alternativo para solucionar problemas de Decisión Estadística desde el punto de vista Bayesiano, lo podemos enfocar del siguiente modo:

Si el problema consiste de un espacio paramétrico o conjunto de estados de naturaleza (H) , con la característica de que todo elemento de dicho conjunto está relacionado con un conjunto de índices $I \neq \emptyset$, de tal forma que identificar un elemento del espacio paramétrico es equivalente a identificar una colección de valores indicados por el conjunto I que cumplan determinadas condiciones, entonces observamos el problema original como una colección de problemas de estadística bayesiana, donde a cada valor $t \in I$ corresponde un problema de estadística bayesiana.

Ejemplo:

Si $(H) = \{S \mid S : \mathbb{R}_+ \rightarrow [0,1] \text{ Función de supervivencia}\}$, es el espacio paramétrico, entonces los estados de naturaleza están relacionados con el conjunto $I = \mathbb{R}_+$ o pueden identificarse con una colección $\{S_{(t)}\}_{t \in \mathbb{R}_+}$, tal que:

- i) $\lim_{t \rightarrow \infty} S(t) = 0$; ii) $\lim_{t \rightarrow 0} S(t) = 1$;
- iii) $\lim_{u \rightarrow t^+} S(u) = S(t)$; $\forall t \in \mathbb{R}_+$.
- iv) $\forall t_1 \leq t_2, S(t_1) \geq S(t_2)$ monótona no creciente.

Esta idea la incorporamos en la siguiente definición:

DEFINICION 1.

Sea el problema de decisión estadística bayesiano:

$\{(\mathcal{H}, D, R), (\mathcal{S}, P)\}$, tal que existe un conjunto de Índices I relacionado con los estados de naturaleza $\theta \in \mathcal{H}$, de tal forma que identificar un elemento de $\mathcal{H}$ es equivalente a identificar una colección de valores indicados por I que satisfacen unas condiciones dadas. Entonces diremos que:

$\{(\mathcal{H}, D, R), (\mathcal{S}, P)\}$ es una colección de problemas de Decisión Estadística Bayesiana, lo que denotaremos como:

$$\{(\mathcal{H}, D, R), (\mathcal{S}, P)\} = \{(\mathcal{H}_t, D_t, R_t), (\mathcal{S}, P)_t\} \quad t \in I, \text{ con:}$$

- $I \neq \emptyset$ conjunto de índices asociados a $\mathcal{H}$,
- $\{(\mathcal{H}_t, D_t, R_t), (\mathcal{S}, P)_t\}$ Denota el problema de Decisión estadística bayesiana en el instante t , con $\mathcal{H}_t, D_t, R_t$ definidos en forma que $(\mathcal{H}_t, D_t, R_t)$ es un problema de decisión estadística para cada $t \in I$. El par $(\mathcal{S}, P)_t$ denota el enfoque bayesiano en el instante $t \in I$. También lo llamaremos "Conocimiento a Priori" en el instante t .

Enseguida definiremos lo que entenderemos por "Solución Paramétrica" del problema original.

DEFINICION 2.

Sea el problema de estadística bayesiana:

$\{(\mathcal{H}, D, R), (\mathcal{S}, P)\}$ representado por la colección de problemas de estadística bayesiana $\{(\mathcal{H}_t, D_t, R_t), (\mathcal{S}, P)_t\} \quad t \in I$. Diremos que una solución $d^0 \in D$ es una "Solución Paramétrica" si y sólo si $d^0 = \{d_t^0\} \quad t \in I$,

donde $d_t^0 \in D_t$, $\forall t \in I$.

Para determinar una solución del problema $\{(\mathcal{H}_t, D_t, R_t), (\mathcal{S}, P)_t\}$ aplicaremos la técnica denominada Aproximación Paramétrica para cada $t \in I$, fijo. Luego, observamos si al variar $t \in I$, obtenemos una solución del problema original.

Estimaremos la función de supervivencia asociada a una variable aleatoria (v.a) $T \geq 0$, que nos representa el tiempo de muerte de un individuo o el tiempo de fallo de un sistema. La función de pérdida es la cuadrática ponderada y el enfoque bayesiano está expresado por un proceso estocástico $\{S(t)\}$ $t \in \mathbb{R}_+$, que es una supervivencia aleatoria. Así tenemos:

- $\mathcal{H} = \{S \mid S: \mathbb{R}_+ \rightarrow [0,1], \text{ Función de Supervivencia} \}$ es el espacio paramétrico.
- $\mathcal{A} = \mathcal{H}$, por ser un problema de estimación.
- $L = \mathcal{H} \times \mathcal{A} \rightarrow \mathbb{R}_+$, Función de Pérdida Cuadrática Ponderada, tal que, $L(S, \hat{S}) = \int_{\mathbb{R}_+} (S(t) - \hat{S}(t))^2 dw(t)$, para todo $(S, \hat{S}) \in \mathcal{H} \times \mathcal{A}$, con W una medida finita sobre $(\mathbb{R}_+, B(\mathbb{R}_+))$ que se llama comúnmente Medida de Ponderación.
- $(\mathcal{S}, P) = (S = \{S(t)\}_{t \in \mathbb{R}_+}, P)$ denota el enfoque bayesiano o conocimiento a priori asumido, con P medida de probabilidad inducida por S , denominada Medida de Probabilidad a Priori.

Por tanto existe $I = \mathbb{R}_+$ y podemos decir que el problema dado es una colección de problemas, donde:

- $\mathcal{H}_t = \mathcal{A}_t = [0,1] = \{\pi_t(S) = S(t) \in [0,1] \mid S \in \mathcal{H}\}$

$$l_t : (\mathbb{H})_t \times \mathcal{A}_t \rightarrow \mathbb{R}_+$$

$$(\theta, a) \rightarrow l_t(\theta, a) = (\theta - a)^2, \text{ Función de Pérdida Cuadrática Puntual.}$$

- $(\mathcal{S}, P)_t = (S = \{S(t)\}_{t \in \mathbb{R}_+}, P)_t = (S(t), P_t)$, con P_t la medida de probabilidad inducida por la v.a. t -ésima del proceso estocástico $S(t)$, denominada Medida de Probabilidad a Priori para el t -ésimo problema de decisión estadística Bayesiana. Denotamos:

$$(S(t), P_t) = (g_t, P_t), \text{ con } g_t \text{ la función de densidad de probabilidad correspondiente a } S(t).$$

Así, tendremos la notación:

$$\{(\mathbb{H}, \mathcal{A}, L), (\mathcal{S}, P)\} = \{(\mathbb{H})_t, \mathcal{A}_t, l_t, (\mathcal{S}, P)_t\}_{t \in \mathbb{R}_+},$$

donde:

$$\{(\mathbb{H})_t, \mathcal{A}_t, l_t, (\mathcal{S}, P)_t\} = \{([0, 1], [0, 1], (\theta - a)^2), (g_t, P_t)\}, \text{ Para cada } t \in \mathbb{R}_+.$$

a) Solución del problema sin información muestral.

Para cada $t \in \mathbb{R}_+$, tenemos el problema:

$\{([0, 1], [0, 1], (\theta - a)^2), (g_t, P_t)\}$, pero sabemos que para el problema de estimar $\theta \in [0, 1]$ dada la función de pérdida cuadrática $(\theta - a)^2$, el estimador bayes será la media de la distribución a priori, en el problema sin información muestral. Esto es:

$$\tilde{S}_0(t) = E[S(t)]$$

Luego, obtenida la solución $\tilde{S}_0(t) \in \mathcal{A}_t$, $\forall t \in \mathbb{R}_+$, construimos $\tilde{S}_0 = \{\tilde{S}_0(t)\}_{t \in \mathbb{R}_+}$. Pero $\tilde{S}_0 \in \mathcal{U}$ por ser $\tilde{S}_0$ una función de supervivencia, puesto que $S = \{S(t)\}_{t \in \mathbb{R}_+}$ es una supervivencia aleatoria y aplicando resultados de convergencia se deduce que $\tilde{S}_0 \in \mathcal{U}$.

En resumen hemos obtenido:

Proposición 1.

La solución paramétrica del problema de Decisión Bayesiana sin información muestral:

$$\{(\textcircled{H}), \mathcal{A}, L, (\mathcal{S}, P)\}, \quad \text{es}$$

$$\tilde{S}_0 = \{E[S(t)]\}_{t \in \mathbb{R}_+}.$$

b) Solución del problema con información muestral o problema de Decisión Estadística.

En primer término mostraremos como se transforman algunos elementos del problema original con información muestral para un instante $t \in \mathbb{R}_+$, fijo.

La información para el problema original está dada por $X_1, X_2, \dots, X_n$ muestra aleatoria (m.a) de tamaño n , extraída según el proceso $S = \{S(t)\}_{t \in \mathbb{R}_+}$, de la v.a. en estudio $T \geq 0$. Entonces, tal como se expuso antes, el enfoque bayesiano en el instante t , es:

$P[X_i > t/S(t) = \theta] = \theta$, $\forall i = 1, 2, \dots, n$. Esto significa que para el valor t elegido, podemos conocer expresiones, tal como:

$$P [X_1 > t, X_2 > t, \dots, X_k > t, X_{k+1} \leq t, \dots, X_n \leq t / S(t) = \theta] =$$

$$\theta^k (1 - \theta)^{n-k}, \quad \forall k = 0, 1, \dots, n, \quad \text{porque las } X_i \text{ son independientes dado } S(t) = \theta.$$

Esto tiene como consecuencia inmediata que el espacio muestral original $X \subseteq \mathbb{R}_+^n$, donde toman valores $X_1, X_2, \dots, X_n$, es transformado en nuestro espacio muestral para el instante t , denotado por X_t , en $X_t = \{0, 1, 2, \dots, n\}$, que se obtiene por la función:

$$f : X \rightarrow X_t$$

$$\begin{aligned} x &\rightarrow f(x) = \text{Número de observaciones mayores que } t \\ &= \text{Número de coordenadas mayores que } t. \end{aligned}$$

Ahora, nuestra medida de probabilidad inducida por las observaciones condicionada a los estados de naturaleza $S \in \mathcal{H}$, denotada por P_S^X , con $X = (X_1, X_2, \dots, X_n)$, la transformamos en:

$$P_t^{Y(t)}, \quad \text{con } Y(t) = f \circ X = \text{Número de observaciones mayores que } t.$$

En síntesis, la transferencia del espacio muestral original y la medida de probabilidad inducida por X condicionada a los estados de naturaleza $S \in \mathcal{H}$, se transforman para un instante t fijo, en:

$$(X, \sigma(X), P_S^X) \rightarrow (X_t, \sigma(X_t), P_t^{Y(t)}), \quad \text{donde:}$$

$$X_t = \{0, 1, \dots, n\}, \quad \sigma(X_t) = 2^{X_t}, \text{ esto es, la } \sigma\text{-álgebra}$$

en $\mathcal{X}_t$ formada por todos los subconjuntos de $\mathcal{X}_t$, la medida de probabilidad es tal que:

$$P_t^{Y(t)}(A) = P[Y(t) \in A / S(t) = \theta] = \begin{cases} 0 & \text{si } A = \emptyset \\ \sum_{k \in A} \binom{n}{k} \theta^k (1 - \theta)^{n-k} & \text{Si } A \neq \emptyset \end{cases}$$

Para continuar con nuestro enfoque paramétrico en el instante t , debemos mostrar la transformación de las funciones de Decisión. Pero como nuestro espacio muestral es $\mathcal{X}_t$, entonces las funciones de Decisión serán:

$$d : \mathcal{X}_t \rightarrow \mathcal{A}_t = [0, 1], \text{ tal que:}$$

$$\forall \theta \in [0, 1], \quad \int_{\mathcal{X}_t} (\theta - d(k))^2 dP_t^{Y(t)}(k) < \infty,$$

con lo cual obtenemos D_t .

Finalmente, se deduce R_t , la función de riesgo como:

$$R_t(\theta, d) = E[l_t(\theta, d)] = \int_{\mathcal{X}_t} (\theta - d(k))^2 dP_t^{Y(t)}(k)$$

$$, \quad \forall (\theta, d) \in (\mathcal{H})_t \times D_t.$$

Utilizando la notación más usual para denotar el problema de Decisión Estadística Bayesiano, escribimos:

$$\{(\Theta_t, \mathcal{A}_t, \mathbb{P}_t), (g_t, P_t), \{(x_t, \sigma(x_t), P_t^{Y(t)})\}\} =$$

$$\{(\Theta_t, D_t, R_t), (g_t, P_t)\}.$$

Así, el problema de Decisión Estadística Bayesiana original, lo observamos como una colección de problemas de Decisión Estadística Bayesiana, lo cual es expresado como:

$$\{(\Theta, D, R), (g, P)\} = \{(\Theta_t, D_t, R_t), (g_t, P_t)\}_{t \in \mathbb{R}_+}.$$

Finalmente, para el problema en un instante t , se tiene que la función de riesgo bayesiano es:

$$r(P_t, d) = E[R_t(\theta, d)] = \int_{\Theta_t} \int_{\mathcal{X}_t} (\theta - d(k))^2 dP_t^{Y(t)}(k) dP_t(\theta)$$

$$= \int_{\mathcal{X}_t} \int_{\Theta_t} (\theta - d(k))^2 dP_t(\theta/Y(t) = k) dP_t^{Y(t)}(k)$$

De esta expresión se deduce que la solución Bayesiana es:

$\tilde{S}_0(t) = E[S(t)/Y(t) = k]$, con $Y(t) = k$, el número de observaciones mayores que t en la muestra.

Lo que resta es definir la función: $\tilde{S}_0 = \{\tilde{S}_0(t)\}_{t \in \mathbb{R}_+}$, y si $\tilde{S}_0 \in \mathcal{U}$, entonces $\tilde{S}_0$ es la solución paramétrica del problema original de Decisión Estadística Bayesiana denotado por:

$$\{(H, \mathcal{A}, L), (\mathcal{S}, P), \{(X, \sigma(X), P^X)\}_{S \in H}\} = \{(H, D, R), (\mathcal{S}, P)\}.$$

Recordemos que $\mathcal{A} = \{S/S: \mathbb{R}_+ \rightarrow [0, 1] \text{ función de supervivencia}\}$ entonces expresar que $\tilde{S}_0$ es solución paramétrica, equivale a expresar que $\tilde{S}_0$ es una función de supervivencia.

Antes de probar que $\tilde{S}_0 = \{E[S(t)/Y(t)]\}_{t \in \mathbb{R}_+}$ es una función de supervivencia, escribiremos la forma del estimador puntual $E[S(t)/Y(t)]$ para cada valor de $Y(t) = k \in \{0, 1, 2, \dots, n\}$, en una forma adecuada.

El punto es ver que existan las probabilidades condicionales que aparecen y, así caracterizamos el estimador para cada valor de la v.a. $Y(t)$, que nos indica el número de observaciones mayores que t , fijo.

Desde esto escribimos:

$$\tilde{S}_0(t) = E[S(t)/Y(t)] = \sum_{j=0}^n a_j(t) I_{A_j}, \text{ donde } a_j(t) = \frac{E[S(t)I_{A_j}]}{P(A_j)},$$

$A_j = \{Y(t) = j\}$ o equivalente $A_j = \{X \in \mathbb{R}_+^n \mid \text{existen } j \text{ componentes mayores que } t\}$, para cada $j = 0, 1, 2, \dots, n$.

Esto lo haremos porque la relación que existe entre los valores $a_j(t)$ es la expresada por el lema siguiente:

LEMA: $a_0(t) \leq a_1(t) \leq \dots \leq a_n(t); \forall t \in \mathbb{R}_+, \text{ fijo.}$

La demostración de este lema se hace aplicando propiedades de la distribución a posteriori de $S(t)$ dado $Y(t)$ y, observando que para cada valor de $S(t) = \theta \in [0, 1]$, la función $\frac{\theta}{1-\theta}$ es monótona creciente en

$[0,1]$. (Detalles pueden consultarse en [11]).

Enseguida probaremos las propiedades que caracterizan una función de supervivencia, esto es, probaremos que $\tilde{S}_0$ cumple con:

- i) $\lim_{t \rightarrow 0^+} \tilde{S}_0(t) = 1$ c.s.
- ii) $\lim_{t \rightarrow \infty} \tilde{S}_0(t) = 0$ c.s.
- iii) $\lim_{h \rightarrow t^+} \tilde{S}_0(h) = \tilde{S}_0(t)$ para cada $t \in \mathbb{R}_+$.
- iv) Si $t_1 \leq t_2$, entonces $S_0(t_1) \geq S_0(t_2)$ c.s.

Las propiedades i), ii), iii) se cumplen cualquiera sea el conocimiento a priori asumido o enfoque bayesiano, expresado por el proceso estocástico $S = \{S(t)\}_{t \in \mathbb{R}_+}$ que sea una supervivencia aleatoria (que cumpla las condiciones de consistencia de Kolmogorov). La problemática es la propiedad iv) de monotonía no creciente en t (o monotonía no decreciente si estimamos una función de distribución). Las dificultades surgen por varias razones, pero estimamos o podemos decir conjeturamos, que cualquiera sea la supervivencia aleatoria que se asuma como conocimiento a priori la solución paramétrica es efectivamente válida. Lo expresado aquí, está avalado por una serie de resultados empíricos obtenidos mediante simulación. Plantearemos la demostración de la propiedad iv), cuando la supervivencia aleatoria que se utiliza es un proceso neutral por la derecha homogéneo. Pero, como ejemplo obtendremos la solución cuando se utiliza un proceso neutral por la derecha no homogéneo, como es el proceso de Dirichlet.

Las propiedades i), ii) e iii) se deducen aplicando el lema y re

sultados sobre convergencia de v.a. (Detalles pueden verse en [11]).

Cuando el conocimiento a priori es un proceso neutral por la derecha homogéneo, el estimador puede escribirse en la siguiente forma (Ver [11]).

$$S_o(t) = E(S(t)/Y(t) = k] = \frac{\sum_{j=0}^{n-k} (-1)^j \binom{n-k}{j} M_t(n+1-j)}{\sum_{j=0}^{n-k} (-1)^j \binom{n-k}{j} M_t(n-j)}$$

donde $k = 0, 1, 2, \dots, n$, denota el número de observaciones mayores que t en la muestra aleatoria de tamaño n .

Aquí M_t denota la función generatriz de momentos del proceso $W = \{W(t)\}_{t \in \mathbb{R}_+}$ que aparece en la definición de neutral por la derecha homogéneo, esto es, $M_t(u) = E[e^{-uW(t)}]$, (obsérvese el signo menos) recordando que $S(t) = e^{-W(t)}$.

$$\text{Particularmente, } M_t(u) = e^{\gamma(t)} \int_0^\infty (e^{-uz} - 1) dN(z)$$

con $\gamma(t)$ una función continua, creciente definida desde $\mathbb{R}_+$ hacia $\mathbb{R}_+$ tal que $\lim_{t \rightarrow \infty} \gamma(t) = \infty$, $\gamma(0) = 0$, $N(z)$ es una medida finita definida sobre $(\mathbb{R}_+, B(\mathbb{R}_+))$ que se denomina medida de Levy del proceso.

Sean $t_1 \leq t_2$, $t_1 \in \mathbb{R}_+$, $t_2 \in \mathbb{R}_+$, entonces:

$$\tilde{S}_o(t_1) = \frac{\sum_{j=0}^{n-k_1} (-1)^j \binom{n-k_1}{j} M_{t_1}(n+1-j)}{\sum_{j=0}^{n-k_1} (-1)^j \binom{n-k_1}{j} M_{t_1}(n-j)}$$

$$\tilde{S}_o(t_2) = \frac{\sum_{j=0}^{n-k_2} (-1)^j \binom{n-k_2}{j} M_{t_2}(n+1-j)}{\sum_{j=0}^{n-k_2} (-1)^j \binom{n-k_2}{j} M_{t_2}(n-j)}$$

con $Y(t_1) = k_1$, $Y(t_2) = k_2$, número de observaciones mayores que t_1, t_2 , respectivamente.

Para probar que $\tilde{S}_o(t_1) \geq \tilde{S}_o(t_2)$ c.s. $\forall t_1 \leq t_2$ sólo es necesario probar que si $0 \leq t_1 \leq t_2 \leq x_1$ o $x_1 \leq t_1 \leq t_2 \leq x_2$, entonces $\tilde{S}_o(t_1) \geq \tilde{S}_o(t_2)$, ya que para otras situaciones el razonamiento es análogo. Hemos denotado por x_1 el menor valor de las observaciones y x_2 el valor siguiente, esto sólo para efectos pedagógicos.

a) Primero observemos que la situación planteada la podemos resumir gráficamente del modo siguiente:

Es decir, para cualquier $t < x_1 = \min \{X_1, X_2, \dots, X_n\}$ $\tilde{S}_o(t)$ es decreciente en t , porque el estimador toma el valor $a_n(t)$ puesto que las n observaciones son mayores que t , y este comportamiento decreciente en t , es debido a que la función $\gamma(t)$ es creciente (o no decreciente).

En efecto, sean $0 < t_1 \leq t_2 < x_1$, entonces existen n observaciones mayores que t_1, t_2 , esto es $k_1 = k_2 = n$, lo cual implica que:

$$\tilde{S}_o(t_1) = \frac{M_{t_1}(n+1)}{M_{t_1}(n)}, \quad \tilde{S}_o(t_2) = \frac{M_{t_2}(n+1)}{M_{t_2}(n)}, \quad \text{pero}$$

$$\tilde{S}_o(t) = \frac{M_t(n+1)}{M_t(n)} = \frac{e^{\gamma(t)(A+D)}}{e^{\gamma(t)A}} = e^{\gamma(t)D},$$

donde:

$$A = \int_0^\infty (e^{-nz} - 1) dN(z) \leq 0,$$

$$D = \int_0^\infty (e^{-(n+1)z} - e^{-nz}) dN(z) \leq 0, \implies$$

$$\tilde{S}_o(t_1) = e^{\gamma(t_1)D} \geq e^{\gamma(t_2)D} = \tilde{S}_o(t_2) \quad \text{ya que} \quad -D \gamma(t_2) \geq$$

$$-D \gamma(t_1) \geq 0,$$

$$e^{-D} \gamma(t_2) \geq e^{-D} \gamma(t_1) \iff e^D \gamma(t_1) \geq e^D \gamma(t_2)$$

Por lo tanto, $\tilde{S}_0(t_1) \geq \tilde{S}_0(t_2)$ si $0 < t_1 \leq t_2 < x_1$.

b) En segundo lugar se tiene que, $\tilde{S}_0(x_1^-) \geq S_0(x_1^+)$, esto ocurre porque $\tilde{S}_0(x_1^+) = \tilde{S}_0(x_1)$ por la continuidad por la derecha, $S_0(x_1) = a_{n-1}(x_1^-)$ y $\tilde{S}_0(x_1^-) = a_n(x_1)$, luego como $a_n(x_1) \geq a_{n-1}(x_1)$, entonces:

$$\tilde{S}_0(x_1^-) \geq \tilde{S}_0(x_1^+). \quad \text{En general } \tilde{S}_0(x_i^-) \geq \tilde{S}_0(x_i^+), \quad i = 1, \dots, n.$$

c) Finalmente, sólo resta probar que si $x_1 < t_1 \leq t_2 < x_2$, entonces $\tilde{S}_0(t_1) \geq \tilde{S}_0(t_2)$, ya que en cualquier otro caso el razonamiento es análogo. Sean $k_1 = n-1 = k_2$, entonces:

$$\tilde{S}_0(t_1) = \frac{M_{t_1}^{(n+1)} - M_{t_1}^{(n)}}{M_{t_1}^{(n)} - M_{t_1}^{(n-1)}}$$

$$\tilde{S}_0(t_2) = \frac{M_{t_2}^{(n+1)} - M_{t_2}^{(n)}}{M_{t_2}^{(n)} - M_{t_2}^{(n-1)}}$$

Enseguida definimos:

$$A = \int_0^\infty (e^{-nz} - 1) dN(z) \leq 0$$

$$B = \int_0^\infty (e^{-(n-1)z} - e^{-nz}) dN(z) \geq 0$$

$$D = \int_0^{\infty} (e^{-(n+1)z} - e^{-nz}) dN(z) \leq 0, \quad \text{entonces:}$$

$$\tilde{S}_0(t_1) = \frac{e^{(A+D)\gamma(t_1)} - e^{A\gamma(t_1)}}{e^{A\gamma(t_1)} - e^{(A+B)\gamma(t_1)}}$$

$$\tilde{S}_0(t_2) = \frac{e^{(A+D)\gamma(t_2)} - e^{A\gamma(t_2)}}{e^{A\gamma(t_2)} - e^{(A+B)\gamma(t_2)}}.$$

Luego, probaremos que la función de t :

$$\tilde{S}_0(t) = \frac{e^{(A+D)\gamma(t)} - e^{A\gamma(t)}}{e^{A\gamma(t)} - e^{(A+B)\gamma(t)}}$$

es decreciente en t .

Para ésto aplicamos elementos de cálculo, observando que:

$$\tilde{S}_0(t) = \frac{e^{D\gamma(t)} - 1}{1 - e^{B\gamma(t)}}$$

$$\tilde{S}'_0(t) = \frac{(1 - e^{B\gamma(t)})D\gamma'(t)e^{D\gamma(t)} + (e^{D\gamma(t)} - 1)B\gamma'(t)e^{B\gamma(t)}}{(1 - e^{B\gamma(t)})^2}.$$

$$\tilde{S}'_0(t) \leq 0 \iff f(t) = (B-D)e^{(B+D)\gamma(t)} + D e^{D\gamma(t)} - B e^{B\gamma(t)} \leq 0$$

porque $\gamma'(t) \geq 0$, ya que $\gamma(t)$ es creciente.

Entonces tendremos que la función $f(t)$ es tal que:

$$i) \quad f(0) = 0 \quad \text{porque} \quad \gamma(0) = 0$$

$$ii) \quad f'(t) = (B^2 - D^2) \gamma'(t) e^{(B+D)\gamma(t)} + D^2 \gamma'(t) e^{D\gamma(t)} -$$

$$B^2 \gamma'(t) e^{B\gamma(t)}, \quad f(0) = 0; \quad f'(t) \leq 0, \quad \text{porque:}$$

$$B^2 e^{(B+D)\gamma(t)} < B^2 e^{B\gamma(t)}, \quad \text{ya que} \quad B \geq 0 \quad \text{y} \quad D \leq 0$$

$$D^2 e^{(B+D)\gamma(t)} > D^2 e^{D\gamma(t)}, \quad \text{porque} \quad D \leq 0.$$

Esto implica que existe un máximo de la función $f(t)$ en el punto $t = 0$, por lo tanto:

$$f(t) \leq f(0) = 0 \implies f(t) \leq 0, \quad \tilde{S}_0(t) \leq 0.$$

$$\text{Luego, } \tilde{S}_0(t_1) \geq \tilde{S}_0(t_2) \quad \text{si} \quad x_1 < t_1 \leq t_2 < x_2.$$

Desde ésto, podemos afirmar que:

$$\tilde{S}_0(t_1) \geq \tilde{S}_0(t_2) \quad \text{si} \quad 0 \leq t_1 \leq t_2, \quad \text{tal que} \quad t_1 \in \mathbb{R}_+, \quad t_2 \in \mathbb{R}_+.$$

Por lo tanto, desde i), ii), iii) e iv) podemos expresar que la solución paramétrica es efectivamente una función de supervivencia, si el proceso asumido como conocimiento a priori sobre el problema dado es una supervivencia aleatoria neutral por la derecha homogénea.

Esto lo resumimos en:

•

Proposición 2.

Sea el problema de decisión estadística bayesiano:

$\{(\textcircled{H}), D, R), (\mathcal{S}, P)\}$, donde $\mathcal{S} = : S = \{S(t)\}_{t \in \mathbb{R}_+}$ es una supervivencia aleatoria neutral por la derecha homogénea, $\textcircled{H} = \{S/S : \mathbb{R}_+ \rightarrow [0, 1] \text{ función de supervivencia}\}$ y la función de pérdida es la cuadrática ponderada.

Entonces la solución paramétrica:

$$\tilde{S}_0 = \{\tilde{S}_0(t) = E[S(t)/Y(t)]\}_{t \in \mathbb{R}_+}$$

es una función de supervivencia donde $Y(t)$ indica el número de observaciones mayores que t .

Mostraremos, como ya hemos dicho, un proceso que es neutral por la derecha no homogéneo observando que la solución es simple y dicha relación de monotonía es obvia.

ALGUNAS NOTAS SOBRE SOLUCION PARAMETRICA.

1. Como ya hemos expresado, la solución no paramétrica es efectivamente la solución bayesiana que podríamos denominar solución "real" del problema propuesto. Pero dada la difícil aplicabilidad que tiene por la forma y cálculos que requiere, hemos propuesto una aproximación de dicha solución denominada paramétrica. Esta aproximación tiene las características de ser fácil de evaluar y evita el escollo principal que surge en la solución no paramétrica, que es el encontrar la distribución a posteriori del proceso. Otra característica que se deduce de su enunciado es que permite trabajar con cualquier supervivencia aleatoria como conocimiento a priori o enfoque bayesiano.

2. En el intento de aproximarse hacia la solución no paramétrica, cuando se tiene una supervivencia aleatoria como conocimiento a priori, existe ya una solución dada por el método de aproximación lineal. Pero de acuerdo a la formulación matemática del método paramétrico, esta solución se si túa en un "entorno" más cercano que la aproximación lineal respecto a la so lución no paramétrica. (Ver [11]).

3. Desde el enfoque planteado para obtener soluciones paramétricas, se de duce que el método de aproximación lineal, podemos incorporarlo como un ca so particular de solución paramétrica, denominando a este procedimiento "so lución paramétrica por aproximación lineal". Pero, según hemos expresado a lo que llamaríamos solución paramétrica, para afirmar que el método de aproximación lineal es un caso particular de solución paramétrica, deberíamos probar que al variar t , efectivamente obtenemos una función de super vivencia.

Para el caso $n = 1$, esto es inmediato porque la solución lineal coincide con la solución paramétrica, por lo tanto, cuando $n = 1$ (tamaño muestra), podemos denominar a la solución obtenida: "solución bayesiana pa ramétrica por aproximación lineal".

La transformación del proceso:

$$S = \{S(t)\}_{t \in \mathbb{R}_+}$$

Si en el método paramétrico consideramos una transformación del conocimiento a priori reflejado por el proceso estocástico $S = \{S(t)\}_{t \in \mathbb{R}_+}$, es decir, una transformación de cada $S(t)$ v.a. marginal del proceso, de tal forma que la distribución condicional de la v.a. $Y(t)$ dado $S(t)$ sea binomial de parámetros n y $g(S(t)) \forall t \in \mathbb{R}_+$, siendo g la aplicación que transforma nuestro conocimiento a priori, entonces la solución obtenida por el método paramétrico bajo las mismas condiciones de pérdida cuadrá

tica, comprende el procedimiento de estimación de una función de supervivencia asociada a una v.a. $T \geq 0$, cuando existen datos censurados a la derecha y el problema se denomina "azar proporcional".

En efecto, para ello basta considerar que $\exists \beta \geq 0$ conocido y definir $g(S(t)) = S(t)^{\beta+1} \forall t \in \mathbb{R}_+$, entonces la distribución condicional de $Y(t)$ dado $S(t)$ es $B(n, S(t)^{\beta+1})$, siendo $Y(t)$ = número de observaciones mayores que t (sean muertes o censuras).

Ejemplo:

Sea $S = \{S(t)\}_{t \in \mathbb{R}_+}$ un proceso de Dirichlet de parámetro α .

Un proceso de Dirichlet de parámetro α , donde α es una medida no nula sobre $(\mathbb{R}_+, \beta(\mathbb{R}_+))$, queda caracterizado por una constante $\alpha(\mathbb{R}_+)$ y una función de supervivencia:

$S_0(t) = \frac{\alpha(\mathbb{R}_+) - \alpha(t)}{\alpha(\mathbb{R}_+)}$, donde $S_0(t)$ representa en el contexto bayesiano la supervivencia "a priori" esperada, y $\alpha(\mathbb{R}_+)$ es un parámetro que nos indica la confiabilidad en dicha información "a priori" (Ferguson, 1973).

Sabemos, en este caso (por definición y propiedades del proceso de Dirichlet) que $S(t)$, a $t \in \mathbb{R}_+$, fijo, se distribuye como una v.a. beta de parámetros $(\alpha(\mathbb{R}_+) - \alpha(t), \alpha(t))$. Aquí denotamos $\alpha(t) = : \alpha[0, t]$. Entonces, como $Y(t)$ dado $S(t)$ es una binomial, conocemos la distribución "a posteriori" de $S(t)$ dado $Y(t)$, esta será una beta de parámetros apropiados, puesto que la distribución beta es la conjugada de la binomial.

Así tendremos que:

$$S(t) \sim \text{Beta}(\alpha(\mathbb{R}_+) - \alpha(t), \alpha(t))$$

$$S(t)/Y(t) = k \text{ Beta } (\hat{\alpha}, \hat{\beta})$$

$$\text{donde: } \hat{\alpha} = \alpha(\mathbb{R}_+) - \alpha(t) + k$$

$$\hat{\beta} = \alpha(t) + n - k ; \forall k = 0, 1, \dots, n.$$

Por tanto:

$$\begin{aligned} \tilde{S}_0(t) &= E [S(t)/Y(t) = k] = \\ &= \frac{\hat{\alpha}}{\hat{\alpha} + \hat{\beta}} = \frac{\alpha(\mathbb{R}_+) - \alpha(t) + k}{\alpha(\mathbb{R}_+) + n} I(k) \\ &\quad \{0, 1, \dots, n\} \\ &= (1 - p_n) S_n(t) + p_n S_0(t), \end{aligned}$$

$$\text{donde: } S_n(t) = \frac{Y(t)}{n} = \frac{k}{n}$$

$\forall k = 0, 1, \dots, n$, es la supervivencia empírica, $S_0(t)$ es la supervivencia "a priori" (solución del problema sin información) y:

$$p_n = \frac{\alpha(\mathbb{R}_+)}{\alpha(\mathbb{R}_+) + n} \in [0, 1], \forall n \in \mathbb{N} \text{ (fijo)}.$$

Así, el estimador obtenido por el método paramétrico coincide con las soluciones dadas por el método no paramétrico de Ferguson y con la deducida por el método de aproximación lineal de A. García.

Obviamente, este estimador posee las buenas propiedades del estimador de Ferguson. Por ejemplo, sabemos (Ferguson, 1973) que si el tamaño muestral n tiene a infinito, $p_n \rightarrow 0$ y $1 - p_n \rightarrow 1$, es decir, si el tamaño de la muestra crece, el conocimiento "a priori" queda paliado por el co

nocimiento muestral. Este comportamiento asintótico en base al tamaño de la muestra. cuando se trata de un proceso de Dirichlet como conocimiento "a priori", desafortunadamente no es general para cualquier otro proceso neutral a la derecha. Aún, en el método no paramétrico es poco claro un comportamiento tan agradable del estimador, debido a la complejidad de la distribución a posteriori del proceso. Esta característica sólo es consecuencia de las buenas propiedades que posee el proceso de Dirichlet. Por ejemplo, al aplicar el método paramétrico utilizamos inmediatamente la distribución marginal de cada $S(t)$ que es una beta, por lo tanto, la condicional nuevamente es una beta de parámetros apropiados, puesto que es la conjugada de la binomial. Aquí no es necesario utilizar las funciones generatrices de momento del proceso asociado a cada $S(t)$, ya que es mucho más fácil utilizar la distribución conocida de cada $S(t)$. Bien, esta característica ya no es posible aplicar en cualquier otro proceso neutral a la derecha.

Para el análisis respecto al tamaño de muestra n , sólo mencionamos aquí, que por tratarse de un estimador de Bayes respecto a una pérdida cuadrática, el comportamiento de nuestro estimador bajo algunas condiciones de regularidad generales, se aproxima al estimador máximo verosímil de θ cuando n crece. En nuestro caso el estimador máximo verosímil es justamente $S_n(t)$. (Lehmann, 1984).

Si la confiabilidad en nuestro conocimiento "a priori" crece, es decir: Si $\alpha(\mathbb{R}^+)$ tiende a infinito, manteniendo constante $S_0(t)$, el estimador $\tilde{S}_0(t)$ tiende al conocimiento "a priori" esperado $S_0(t)$, ya que $p_n \rightarrow 1$ si $\alpha(\mathbb{R}^+) \rightarrow \infty$. Si la confiabilidad decrece, $\alpha(\mathbb{R}^+)$ tiende a cero, conservando constante $S_0(t)$, el estimador $\tilde{S}_0(t)$ tiende al conocimiento muestral $S_n(t)$, ya que si $\alpha(\mathbb{R}^+) \rightarrow 0$ entonces $p_n \rightarrow 0$.

Estos resultados los expresamos del modo siguiente:

Proposición 3.

Si el conocimiento "a priori" viene dado por un proceso de Dirichlet de parámetro α y hacemos variar $\alpha(\mathbb{R}^+)$ manteniendo constante $S_o(t)$, se tiene que:

$$a) \quad \lim_{\alpha(\mathbb{R}^+) \rightarrow 0} \tilde{S}_o(t) = S_n(t).$$

$$b) \quad \lim_{\alpha(\mathbb{R}^+) \rightarrow \infty} \tilde{S}_o(t) = S_o(t).$$

BIBLIOGRAFIA.

- [1] ASH, R.B. "Real Analysis and Probability", 1972, Academic Press.
- [2] DALAL, S.R. and GAINEFORD, J.H. "On Aproximating Parametric Bayes Models by Nonparametric Bayes Models", 1980. Ann. Statistics, Vol. 8, Nº 3, (664-672).
- [3] DALAL, S.R. "A note on the adequacy of Mixtures of Dirichlet Processes", 1978, Vol. 40, (185-191), SANDHYA.
- [4] DOKSUM, K. "Tailfree and Neutral Random Probabilities and their Pos-

- terior Distributions", 1974, Ann. Probability, Vol. 2, № 2, (183-201).
- [5] DOOB, J.L. "Stochastic Processes", 1953, John Wiley.
- [6] DYSTRA, R.L. and LAUD, P. "A Bayesian Nonparametric Approach to Reliability", 1981, Ann. Statistics, Vol. 9, № 2, (356-367).
- [7] FERGUSON, T. S. "Mathematical Statistics: A Decision Theoretic Approach", 1967, Academic Press.
- "A Bayesian Analysis of Some Nonparametric Problems", 1973, Ann. Statistics, Vol. 1, № 2, (209-230).
- "Prior Distributions on Spaces of Probability Measures", 1974, Ann. Statistics, Vol. 2, № 4, (615-629).
- [8] FERGUSON, T.S. and KLASS, M.J. "A Representation of Independent Processes without Gaussian Components", 1972, Ann. Math. Statistics, Vol. 43, (1634-1643).
- [9] FERGUSON, T.S. and PHADIA, E.G. "Bayesian Nonparametric Estimation Based on Censored Data", 1979, Ann. Statistics, Vol. 7, № 1, (163-186).
- [10] GARCIA, A. "Problemas de Decisión no Paramétrica: Aproximación Lineal", 1982, Univ. Complutense, Madrid, Tesis Doctoral.
- [11] MONDACA, G. "Modelo de Bayes Generalizado: Aproximación Paramétrica", 1986, Universidad Complutense, Madrid, Tesis Doctoral.
- [12] MORALES, D. "Decisión Bayesiana no Paramétrica: Procesos Gamma-Expo

nenciales y Mixturas", 1984, Universidad Complutense, Madrid, Tesis Doctoral.

[13] QUESADA, V. "Métodos Empírico Bayes no Paramétrico", 1972, Universidad Complutense, Madrid, Tesis Doctoral.

[14] QUESADA, V. y GARCIA, A. "Problemas de Decisión Bayesiana no Paramétrica en los Procesos Gamma Exponenciales", 1982, Revista "Estadística Española", I.N.E.

[15] SUSARLA, V. and VAN RYZIN, J. "Nonparametric Bayesian Estimation of Survival Curves from Incomplete Observation", 1976, J.Am. Statistics Assoc. Vol 71, Nº 356, (897-902).

"Empirical Bayes Estimation of a Distribution (Survival) Function from Righth Censored Data", 1978, Ann. Stat., Vol. 8, (740-754).

"Large Sample Theory for an Estimation of the Mean Survival Time from Censored Samples", 1982, Ann. Stat., Vol. 8, Nº 5, (1002-1016).