Trabajos de Estadística y
de Investigación Operativa
Vol. 35. Núm. 2, 1984, pp. 187 a 195

NONPARAMETRIC ESTIMATION: THE SURVIVAL FUNCTION

Alfonso García Pérez
Departamento de Estadística Matemática
Facultad de Ciencias
Universidad Autónoma de Madrid

RESUMEN

Se considera la función de supervivencia desconocia $S(t)$ de una variable aleatoria $T \geqslant 0$. Primero estudiamos las propiedades de $S(t)$ y luego, la estimamos desde un punto de vista Bayesiano, obteniendo, bajo pérdida cuadrática, el estimador

$$\hat{S}(t) = [1 - p_n(t)]S_n(t) + p_n(t)S_0(t)$$

y el riesgo Bayes mínimo asociado

$$R_{\min}(n) = p_n(t)R_{\min}(0)$$

en donde $S_n(t)$ es la función de supervivencia empírica, $S_0(t)$ la función de supervivencia a priori, $R_{\min}(0) = V(S(t))$ el riesgo en el problema sin muestra y

$$p_n(t) = \frac{S_0(t) - E[S^2(t)]}{S_0(t) + (n-1)E[S^2(t)] - nS_0^2(t)}, \qquad 0 \leqslant p_n(t) \leqslant 1$$

siendo $\hat{S}(t) \underset{c.s.}{\rightarrow} S(t)$ y $R_{\min}(n) \underset{n \to \infty}{\longrightarrow} 0$. Comparamos dicho estimador con la media a posteriori y después de ver condiciones generales bajo las cuales los coeficientes, en las estimaciones Bayesianas lineales, suman uno, terminamos dan-

(*) Recibido, Marzo, 1982

do reglas Bayes para las funciones lineales de $S(t)$.

SUMMARY

The unknown survival function $S(t)$ of a random variable $T \geqslant 0$ is considered. First we study the properties of $S(t)$ and then, we estimate it from a Bayesian point of view. We compare the estimator with the posterior mean and we finish giving Bayes rules for linear functions of $S(t)$.

Key words: Bayesian nonparametric estimation, linear approach, survival function estimation.

A.M.S. Subject classification: 62 C 10.

1. INTRODUCTION AND SUMMARY

Let $T \geqslant 0$ be a random variable with an unknown distribution function $F \in \mathfrak{F}$, where $\mathfrak{F}$ is the set of all probability distributions defined over $(\mathbb{R}, \beta)$, the measurable space of the real line with the σ-field of Borel subsets of $\mathbb{R}$.

If $T_1, \ldots, T_n$ is a random sample from F, we shall look for the Bayes estimator of a parameter $g(F)$, when the quadratic loss

$$L(g(F), d) = [g(F) - d]^2$$

and a prior probability $\mathcal{P}$ over $\mathfrak{F}$, are supposed.

We begin, in section 2, studing some properties of the fuction $S(x) = 1 - F(x)$, where $F(x)$ is a random distribution function of a generic random variable X.

We continue, in section 3, estimating the survival function of T,

$$S(x) = 1 - F(t), \qquad t \geqslant 0$$

inside the set of linear combinations of $S_n(t)$ (the empirical survival function) and $S_0(t) = E_{\mathcal{P}}[S(t)]$ (the prior survival function), getting the estimator

$$\hat{S}(t) = [1 - p_n(t)]S_n(t) + p_n(t)S_0(t)$$

and the minimum Bayes risk

$$R_{\min}(n) = p_n(t)V(S(t))$$

which have good properties.

We finish showing that all the linear functions of $S(t)$ can be estimated from $\hat{S}(t)$.

2. THE RANDOM FUNCTION $S(x)$

If X is a random variable and $F(x)$ is a random distribution function of X, Doksum (1974), let $S(x) = 1 - F(x)$ be a random function.

Theorem 2.1. If $S(x)$ is continuous is quadratic mean, g: $\mathbb{R} \to \mathbb{R}^+ \cup \{0\}$ is a nonnegative measurable function, such that

$$\forall n \in \mathbb{N} \qquad \int_{-n}^{n} g^2(s)\, ds < \infty$$

and if

$$\int_{\mathcal{F}} \left(\int_{-\infty}^{+\infty} g(x)S(x)\, dx \right) \left(\int_{-\infty}^{+\infty} g(y)S(y)\, dy \right) d\mathcal{P}(S) < \infty,$$

then, it is

$$\int_{\mathcal{F}} \left(\int_{-\infty}^{+\infty} g(x)S(x)\, dx \right) \left(\int_{-\infty}^{+\infty} g(y)S(y)\, dy \right) d\mathcal{P}(S) = \\ = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x)g(y) \left(\int_{\mathcal{P}} S(x)S(y)\, d\mathcal{P}(S) \right) dx\, dy$$

Proof. $\forall n \in \mathbb{N}$, if we consider the random variables

$$Y_n = \int_{-n}^{n} g(x)S(x)\, dx, \qquad Y = \int_{-\infty}^{+\infty} g(x)S(x)\, dx$$

we get the result by the bounded convergence theorem, and because Davis' theorem.

We extend this result because of the next theorem.

Theorem 2.2. If $S(x)$ is continuous in quadratic mean, g_1: $\mathbb{R} \to \mathbb{R}^+ \cup \{0\}$ and g_2: $\mathbb{R} \to \mathbb{R}^+ \cup \{0\}$ are two non-negative measurable functions, such that

$$\forall n \in \mathbb{N}, \qquad \int_{-n}^{n} g_1(x)\, dx < \infty, \qquad \int_{-n}^{n} g_2(x)\, dx < \infty$$

and if

$$\int_{\mathcal{F}}\left(\int_{-\infty}^{+\infty} g_i(x)S(x)\,dx\right)\left(\int_{-\infty}^{+\infty} g_j(y)S(y)\,dy\right)d\mathcal{P}(S) < \infty,\quad i = 1, 2,\; j = i, 2$$

then, it is

$$\int_{\mathcal{F}}\left(\int_{-\infty}^{+\infty} g_1(x)S(x)\,dx\right)\left(\int_{-\infty}^{+\infty} g_2(y)S(y)\,dy\right)d\mathcal{P}(S) =$$
$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g_1(x)g_2(y)\left(\int_{\mathcal{P}} S(x)S(y)\,d\mathcal{P}(S)\right)dx\,dy$$

Integrating by parts and using theorem 2.2, we have the next result:

Theorem 2.3. Let $S(x)$ be continuous in quadratic mean, and let h_1: $\mathbb{R} \to \mathbb{R}$ and h_2: $\mathbb{R} \to \mathbb{R}$ be two functions of X, such that theirs derivates with respect the Lebesgue measure

$$g_1(x) = \frac{dh_1(x)}{dx} \qquad \text{and} \qquad g_2(y) = \frac{dh_2(x)}{dy}$$

are non-negative measurable functions which verify that

$$\forall n \in \mathbb{N}, \qquad \int_{-n}^{n} g_i(s)\,ds < \infty, \qquad i = 1, 2$$

Let us suppose that

$$c_1 = \lim_{x\to\infty} h_1(x)S_0(x) - \lim_{x\to\infty} h_1(x) < \infty$$

and that

$$c_2 = \lim_{x\to\infty} h_2(x)S_0(x) - \lim_{x\to\infty} h_2(x) < \infty$$

Then, if

$$\int_{\mathcal{F}}\left(\int_{-\infty}^{+\infty} |g_i(x)S(x)|\,dx\right)d\mathcal{P}(S) < \infty, \qquad i, 2.$$

and

$$\int_{\mathcal{F}}\left(\int_{-\infty}^{+\infty} g_i(x)S(x)\,dx\right)\left(\int_{-\infty}^{+\infty} g_j(y)S(y)\,dy\right)d\mathcal{P}(S) < \infty,\; i = 1, 2,\; j = i, 2,$$

it is

$$\int_{\mathfrak{F}} \left(\int_{-\infty}^{+\infty} h_1(x)\, dF(x)\right)\left(\int_{-\infty}^{+\infty} h_2(y)\, dF(y)\right) d\mathcal{P}(F) =$$

$$c_1 \cdot c_2 - c_1 \cdot \int_{-\infty}^{+\infty} g_2(y)\left(\int_{\mathfrak{F}} S(y)\, d\mathcal{P}(S)\right) dy -$$

$$- c_2 \int_{-\infty}^{+\infty} g_1(x)\left(\int_{\mathfrak{F}} S(x)\, d\mathcal{P}(S)\right) dx +$$

$$+ \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g_1(x) g_2(y)\left(\int_{\mathfrak{F}} S(x)S(y)\, d\mathcal{P}(S)\right) dx\, dy$$

where $F(x) = 1 - S(x)$.

3. SURVIVAL FUNCTION ESTIMATION

Let us thick again about the random variable $T \geqslant 0$ and the unknown survival function $S(t)$. If the process $S(t)$ is such that the Kolmogorov consistence conditions are realised, there exists a probability over the trajectory set, which contains $\mathfrak{F}$. So, $\mathcal{P}$ pulls out a $S(t)$ which may be is not a survival function, but because a theorem of Doob (1953), there exists a separable version of $S(t)$ that is a survival function. Then, $\mathcal{P}$ pulls out a function $S(t)$ that can be «arranged» in order to be a survival function. That is why we call $\mathcal{P}$ a prior probability on $\mathfrak{F}$.

We are going to estimate $S(t)$ following the next prior-posterior clasic Bayesian sketch: we pull out a sample of size n_1 from $S(t)$ and we look for the Bayes rule $\hat{S}_1(t)$ inside the set of decision rules

$$aS_{n_1}(t) + bS_0(t).$$

Then, we pull out another sample of size n_2 and we look for the Bayes rule $\hat{S}_2(t)$ inside the set of decision rules

$$aS_{n_2}(t) + b\hat{S}_1(t)$$

We continue with the process till the k-th stage, where if we pull out a sample of size n and if $\hat{S}_{k-1}(t)$ is the Bayes rule of the $(k-1)$-th stage, we look inside the set of decision rules

$$aS_n(t) + b\hat{S}_{k-1}(t).$$

In those conditions, we have the next result.

Theorem 3.1. If $S(t)$ has first and second moments with respect $\mathcal{P}$ and if we follow the before sketch, the Bayes estimator of $S(t)$, is

$$\hat{S}(t) = [1 - p_n(t)]S_n(t) + p_n(t)S_0(t)$$

and the minimum Bayes risk associate,

$$R_{\min}(n) = p_n(t)\{E[S^2(t)] - E^2[S(t)]\} = p_n(t)R_{\min}(0) \tag{4}$$

where

$$p_n(t) = \frac{E[S(t)] - E[S^2(t)]}{E[S(t)] + (n-1)E[S^2(t)] - nE^2[S(t)]}, \qquad 0 \leqslant p_n(t) \leqslant 1$$

and the expectations are calculated with respect $\mathcal{P}$. Like before

$$E[S(t)] = S_0(t).$$

Proof. If we look for the Bayes rule $\hat{S}_1(t)$ inside the set of decision rules like

$$aS_{n_1}(t) + bS_0(t)$$

the a and b which made minimun the Bayes risk

$$R = \int_{\mathcal{F}} \int_0^{\infty} [S(t) - aS_{n_1}(t) - bS_0(t)]^2 \, dQ[S_{n_1}(t)] \, d\mathcal{P}(S)$$

are

$$a = 1 - p_{n_1}(t) \qquad \text{and} \qquad b = p_{n_1}(t)$$

where $Q[S_{n_1}(t)]$ is the distribution of $S_{n_1}(t)$ in the sample. If we continue with the process till the k-th stage we shall get the result.

From the Glivenko-Cantelli theorem we follow the next result because

$$\lim_{n \to \infty} p_n(t) = 0.$$

Proposition 3.1. Whatever the true survival function $S(t)$ be, the Bayes estimate $\hat{S}(t)$ converges to it almost surely, when n goes to ∞ and for each t. Also, the Bayes risk (4) goes to 0.

So, we have an estimator which depends on two components, one the prior information $S_0(t)$ and the other one, the sample information $S_n(t)$, and when n goes to ∞ we have only the sample information.

We know that the Bayes rule inside the set of all Bayes rules is the posterior mean $E_{S(t)/x}[S(t)]$, but because $S_n(t)$ is a sufficient statistic for $S(t)$, $E_{S(t)/z}[S(t)]$, where z is a particular value of $S_n(t)$, is also a Bayes rule inside the set of all Bayes rules and can be considered in a sense, like the regression curve of $S(t)$ on $S_n(t)$, while our estimator $\hat{S}(t)$ can be supposed to be the regression line of $S(t)$ on $S_n(t)$.

So, when we approximate the posterior mean by $\hat{S}(t)$, really we are approximating the regression curve by the regression line, and when $\mathcal{P}$ is the induced by a Dirichlet process both of them are the same.

4. $S(t)$ FUNCTIONS ESTIMATION

First of all, why did the coefficients of the sample and prior components add one? Has the hazard been, by any chance, graceful with us? The answers of these questions are in the next theorem.

Theorem 4.1. Let $\theta \in \Theta$ be a parameter and let π be a prior distribution on Θ such that $E_\pi[\theta] \neq 0$. If $E_\pi[E_m[X_i]] = E_\pi[\theta]$, $i = 1, \ldots, n$, where $E_m$is the expectation with respect the sample distribution, then the Bayes rule for θ, with quadratic loss, inside the set of decision rules

$$a_1X_1 + \ldots + a_nX_n$$

is such that the estimations $\hat{a}_1, \ldots, \hat{a}_n$ add one, i.e.,

$$\hat{a}_1 + \ldots + \hat{a}_n = 1$$

Proof. We must find the $\hat{a}_1, \ldots, \hat{a}_n$ such that the Bayes risk

$$\int_\Theta \int_{\mathfrak{X}} (- \hat{a}_1X_1 - \ldots - \hat{a}_nX_n)^2 \, dQ \, d\pi$$

was minimun. But the d which made minimun

$$\int_\Theta \int_{\mathfrak{X}} (\theta - d)^2 \, dQ \, d\pi$$

is $d = E_\pi[E_m[\theta]]$, so $\hat{a}_1, \ldots, \hat{a}_n$ must be such that

$$\hat{a}_1X_1 + \ldots + \hat{a}_nX_n = E_\pi[E_m[\theta]] = E_\pi[\theta]$$

and taking expectations we have that

$$\hat{a}_1 E_\pi[\theta] + \ldots + \hat{a}_n E_\pi[\theta] = E_\pi[\theta]$$

or

$$\hat{a}_1 + \ldots + \hat{a}_n = 1$$

Theorem 4.1. Let $\hat{S}(t)$ be the Bayes rule for $S(t)$, with quadratic loss, looked for inside the set of decision rules

$$aS_n(t) + bS_0(t)$$

Let $g(x)$ be a linear function on $\mathbb{R}$. Then, there is a set of decision rules such that $g(\hat{S}(t))$ is the Bayes rule, with the same quadratic loss and prior distribution, for $g(S(t))$, being equal the minimun Bayes risk. The mentioned set is

$$g(aS_n(t) + bS_0(t))$$

PROOF. $g(x)$ can be written like

$$g(x) = Ax + B$$

and so,

$$\int\int [g(S(t)) - g(aS_n(t) + bS_0(t))]^2 \, dQ \, d\mathcal{P} = $$
$$= A^2 \int\int [S(t) - aS_n(t) - bS_0(t)]^2 \, dQ \, d\mathcal{P}$$

which is minimun for $\hat{a} = 1 - p_n(t)$ and $\hat{b} = p_n(t)$.

Let us observe that the theorem we have seen before can be extend, considering instead linear functions, more general ones and using the hyperplane representation with bilinear, etc., functions.

Acknowledgments. I would like to thank DR. V. QUESADA, for many helpful discussions.

REFERENCES

DAVIS, M. H. A. (1977): «Linear estimation and stochastic control». Chapman and Hall.

DOKSUM, K. (1974): «Tailfree and neutral random probabilities and their posterior distributions». *Ann. Probab.*, **1**, 183-201.

DOOB, J. L. (1953): «Stochastic Processes». John Wiley.

FERGUSON, T. (1973): «A Bayesian analysis of some nonparametric problems». *Ann. Statist.*, 209-230.

GARCÍA PÉREZ, A. (1982): «Problemas de decisión Bayesiana no paramétrica: Aproximación lineal». *Ph. Doct. Thesis*, Universidad Complutense, Madrid, n.° 12/83.

GARCÍA PÉREZ, A. (1982): «El proceso homogéneo simple y la función de supervivencia». *Estadística Española*, n.° 95, 61-72.

GOLDSTEIN, M. (1975): «Approximate Bayes solutions to some nonparametric problems», *Ann. Statist.*, **3**, 2, págs. 512-517.

QUESADA, V. y GARCÍA PÉREZ, A. (1982): «Estimation of random survival functions». Qüestiió, V.6, n.° 2, 171-178.