Qüestiió - V. 9, n.º 2 (juny 1985) pp. 77-87

NONPARAMETRIC BAYESIAN ESTIMATION AND GOODNESS OF FIT TEST

VICENTE QUESADA PALOMA*, ALFONSO GARCÍA PÉREZ**
* UNIVERSIDAD COMPLUTENSE DE MADRID
**UNIVERSIDAD AUTÓNOMA DE MADRID

We first make a review of prior distributions neutral to the right, and then we get the Bayes rule for the survival function S(t) = 1 - F(t), with quadratic loss, with these prior distributions. We give, after that, the estimator with a special kind of processes neutral to the right, the homogeneous processes.

We get in point four the linear Bayes rule and we give there an interpretation of the parameters.

We finish with a Bayesian generalization of the Kolmogorov-Smirnov goodness of fit test.

Keywords: PROCESSES NEUTRAL TO THE RIGHT, HOMOGENEOUS PROCESSES, LINEAR APPROACH, BAYESIAN GOODNESS OF FIT TEST.

1. INTRODUCTION.

In nonparametric Bayesian estimation, where the parametric space is the set of all probability distributions on a given sample space, the prior distribution is the induced by a stochastic process, and the posterior distribution is nearly impossible to apply. It is only possible to attach statistical problems with a prior distribution indeced by a Dirichlet process. We begin studying a large class of prior distributions, which contains Dirichlet processes, the processes neutral to the right. When we study the survival function estimation problem in such a class, with quadratic loss, and we restrict our attention to a special kind of processes neutral to the right, the homogeneous processes, we get a good interpretation of the parameters involved in the estimation of S(t), the survival function.

We have, however, some unpleasant circunstances, even in this reduced class of prior distributions, that make it adviseable to look for a solution. i.e., a "good estimator" that is, workable with any kind of prior distribution. But, what do we mean by a "good estimator"?. Namely that all the elements in it have clear interpretations and that any person who wants to use it can do so with only two requirements: a prior estimation $S_o(t)$ and a degree of belief in it, l_t. We get the estimator

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

$$\text{with} \quad p_n(t) = \frac{1 - [S_o(t)]^{l_t}}{1 + (n-1)\,[S_o(t)]^{l_t} - n\, S_o(t)}$$

and $S_n(t)$ the empirical survival function, which with a Dirichlet prior is just the posterior mean.

We finish with a Bayesian generalization of the Kolmogorov-Smirnov goodness of fit test.

2.- PRIOR DISTRIBUTIONS NEUTRAL TO THE RIGHT.

The class of random distributions called neutral to the right, introduced by Doksum, is a class of prior distributions on the set of probability distributions on a given

- Vicente Quesada Paloma - Dep. Estadística Matemática e Investigación Operativa de la Facultad de Matemáticas de la Universidad Complutense de Madrid - Madrid 28040
- Alfonso García Pérez - Dep. de Estadística Matemática de la Facultad de Ciencias de la Universidad Autónoma de Madrid - Madrid 28049.

- Article rebut el març del 1985.

sample space, which has good properties in the sense that the support of their elements is large enough and that the posterior distribution of an element of the class is also contained.

Definition 2.1. Doksum (1974):

A stochastic process $\{F(t):t\epsilon R\}$ defined on a probability space $(\Omega,\sigma_\Omega,\lambda)$, with state space $[0,1]$ is said to be a random distribution function neutral to the right if can be written in the form

$$F(t) = 1 - \exp[-Y_t]$$

where Y_t is some separable, a.s. non-decreasing, right continuous a.s., independent increment process with $\lim_{t\to-\infty} Y_t = 0$ a.s. and $\lim_{t\to\infty} Y_t = \infty$ a.s.

Let us suppose that Y_t has no fixed point of discontinuity; then, the increments have infinitely divisible distributions, and so the Lévy representation of the logarithm of the moment generating function of Y_t may be written as

$$\log M_t(\theta) = \log E[e^{-\theta Y_t}] = -\theta\alpha(t) + \theta^2\lambda(t) + \int_{-\infty}^{0}(e^{-\theta z}-1+\frac{\theta z}{1+z^2})\,dL_t(z) + \int_0^{\infty}(e^{-\theta t}-1+\frac{\theta z}{1+z^2})\,dN_t(z)$$

where

$$\Pi_t\{|x|>\varepsilon\}<\infty \quad \text{and} \quad \int_{0<|x|\le\varepsilon} x^2 d\Pi_t(x), <\infty$$

$$\text{If} \quad \Pi_t(B) = \begin{cases} L_t(B), & \text{if } B\subset(-\infty,0) \\ N_t(B), & \text{if } B\subset(0,\infty) \end{cases}$$

but $t\ge 0$ and Y_t is non-decreasing, so if

$$b(t) = \alpha(t) - \int_0^{\infty}\frac{z}{1+z^2}\,dN_t(z)$$

it will be

$$\log M_t(\theta) = \log E[\exp-\theta Y_t] = -\theta b(t) + \int_0^{\infty}(e^{-\theta z}-1)\,dN_t(z)$$

where

$$\int_0^{\infty}\frac{z}{1+z}\,dN_t(z)<\infty.$$

In this context, $N_t(.)$ will be a measure on the Borel subsets of $(0,\infty)$, continuous in t. Let us relax however this condition, allowing the process Y_t to have some fixed points of discontinuity (at most a numerable quantity), provided (still refering to Lévy representation) the lenghts of the jumps at these points have infinitely divisible distributions. We will refer to $N_t(.)$ as the Lévy measure associated to the process neutral to the right, and is used to characterise the process.

At last, let us fix our attention on two facts: the non-random component is now $-\theta b(t)$, and so if it does not exist, it will be

$$\mathcal{P}\{F:F \text{ is a discrete distribution function}\} = 1$$

where $\mathcal{P}$ is the distribution of the process, i.e., $F(t)$ will increase jumping with probability 1, although it will have fixed points of discontinuity if and only if $N_t(.)$ has them.

Survival function estimation:

Let $X_1,\ldots,X_n$ be a sample of size n from F. If $G_u(s)$ denotes the prior distribution of the jump in Y_t at u, $H_u(s)$ denotes the posterior distribution of the jump in Y_t at u, given $X = u$, $M_t^-(\theta) = \lim_{s\uparrow t} M_s(\theta)$, n $F_n(t)$ denotes the number of X_i less than or equal to t, and $S_n(t) = 1 - F_n(t)$. Following Ferguson and Phadia /2/, the Bayes rule for $S(t)$ with loss function

$$L(S,\hat{S}) = \int_0^{\infty}(S(t)-\hat{S}(t))^2\,dW(t)$$

where W is a given finite measure on $(\mathbb{R}^+,\mathcal{B}_{\mathbb{R}^+})$, and observing that the X_i in the product are such that $X_i \leqslant t$, will be

$$E[S(t)/data]=M_t(1/data)=$$

$$= \frac{M_t(nS_n(t)+1)}{M_t(nS_n(t))} \prod_{i=1}^{nF_n(t)} \left[\frac{M^-_{X_i}(n-i+2)}{M^-_{X_i}(n-i+1)} \cdot \frac{M_{X_i}(n-i)}{M_{X_i}(n-i+1)} \cdot \frac{C_{X_i}(n-i+1)}{M_{X_i}(n-2)} \right] \quad (1)$$

where, if X_i is a prior fixed point of discontinuity of Y_t,

$$C_{X_i}(\alpha) = \int_0^\infty e^{-\alpha s}(1-e^{-s})\,dG_{X_i}(s)$$

while, if X_i is not a prior fixed point of discontinuity of Y_t,

$$C_{X_i}(\alpha) = \int_0^\infty e^{-\alpha s}\,dH_{X_i}(s)$$

We see that if we are able to manage the moment generating function M_X in the different points, we also will be able to manage the Bayes rule.

3. THE HOMOGENEOUS PROCESSES:

The Bayes rule given before is nearly impossible to apply (at least from a practical point of view) because of the difficulties encountered in evaluating H_X, and that is why we now restrict our attention to a special kind of process neutral to the right, the homogeneous processes:

A random distribution function F neutral to the right is said to be homogeneous, Ferguson and Phadia /2/, if the independent increment process $Y_t = -\log(1-F(t))$ has as a characteristic function,

$$M_t(\theta) = \exp\{\gamma(t) \int_0^\infty (e^{-\theta z} - 1)\, dN(z)\}$$

where $\gamma(t)$ is continuous nondecreasing, $\lim_{t\to-\infty}\gamma(t) = 0$ and $\lim_{t\to+\infty}\gamma(t) = +\infty$; that is, it has a Lévy measure independent of t. Under these conditions, the homogeneous processes neutral to the right, will be such that the distribution of the process $\mathcal{P}$, the prior distribution in a estimation problem, will pull out a discrete distribution function F with probability 1, without any fixed point of discontinuity.

Because Y_t has not got any fixed point of discontinuity $C_{X_i}(\alpha)$ depends on H_{X_i}, and it is

$$dH_X(z) = \frac{(1-e^{-z})\, dN(z)}{\int_0^\infty (1-e^{-z})\, dN(z)}$$

Definition 3.1:

If N(.) is the Lévy measure of a homogeneous process, let us note by $l(N,\alpha)$ the function, which does not depend on t,

$$l(N,\alpha) = \frac{\int_0^\infty e^{-\alpha z}\, e^{-z}(1-e^{-z})\,dN(z)}{\int_0^\infty e^{-\alpha z}(1-e^{-z})\,dN(z)}, \quad \alpha=0,1,\ldots,n.$$

We see that because

$$M_t(\theta)=E\left[e^{-\theta Y_t}\right]=E\left[(S(t))^\theta\right]= =\exp\{\gamma(t) \int_0^\infty (e^{-\theta z}-1)\, dN(z)\}$$

it is

$$l(N,\alpha) = \frac{\log M_t(\alpha+2) - \log M_t(\alpha+1)}{\log M_t(\alpha+1) - \log M_t(\alpha)} = \frac{\log E[(S(t))^{\alpha+2}] - \log E[(S(t))^{\alpha+1}]}{\log E[(S(t))^{\alpha+1}] - \log E[(S(t))^{\alpha}]}$$

and also, that if we call $S_0(t) = M_t(1)$, it will be

$$\exp\{-\gamma(x_i) \int_0^\infty e^{-(m-i\ 1)z}(1-e^{-z})\,dN(z)\} =$$

$$= \exp\left\{-\gamma(x_i)\left(\int_0^\infty (1-e^{-z})\, dN(z)\right) \cdot \left(\frac{\int_0^\infty e^{-z}(1-e^{-z})\,dN(z)}{\int_0^\infty (1-e^{-z})\,dN(z)}\right) \cdot \ldots \cdot \left(\frac{\int_0^\infty e^{-(n-i\ 1)z}(1-e^{-z})\,dN(z)}{\int_0^\infty (e^{-(n-1)z}(1-e^{-z})\,dN(z)}\right)\right\} =$$

$$= [S_0(x_i)]^{l(N,0)\cdot l(N,1)\cdot\ldots\cdot l(N,n-i)}$$

Survival function estimation:

Because the homogeneous processes are neutral to the right, the Bayes rule will be that given by (1), but now, i=0,1,...,n

$$\frac{M^-_{X_i}(n-1+2)}{M^-_{X_i}(n-i+1)} \cdot \frac{M_{X_i}(n-i)}{M_{X_i}(n-i+1)} =$$

$$= \frac{\exp\left\{-\gamma(t)\int_0^\infty e^{-(n-i+1)z}(1-e^{-z})\,dN(z)\right\}}{\exp\left\{-\gamma(t)\int_0^\infty e^{-(n-i)z}(1-e^{-z})\,dN(z)\right\}} =$$

$$=[S_0(x_i)]^{1(N,-1)\cdot 1(N,0)\cdot 1(N,1)\cdot\ldots\cdot 1(N,n-i-1)[1(N,n-i)-1]}$$

where 1(N,-1) is defined as 1, and is also now

$$\frac{C_{X_i}(n-i+1}{C_{X_i}(n-i)} = 1(N,n-i) = \frac{\int_0^\infty e^{-(n-i+1)z}(1-e^{-z})\,dN(z)}{\int_0^\infty e^{-(n-i)z}(1-e^{-z})\,dN(z)}$$

and

$$\frac{M_t(nS_n(t)\ 1)}{M_t(nS_n(t))} = \exp\left\{-\ (t)\int_0^\infty e^{-nS_n(t)z}(1-e^{-z})\,dN(z)\right\} =$$

$$=[S_0(t)]^{1(N,-1)\cdot 1(N,0)\cdot 1(N,1)\cdot\ldots\cdot 1(N,nS_n(t)-1)}$$

and then, the Bayes rule is with prior distribution homogeneous process neutral to the right and with the quadratic loss function.

$$E[S(t)/data] = [S_0(t)]^{\ell(N,-1)\cdot\ell(N,0)\cdot\ldots\cdot\ell(N,nS_n(t)-1)} \cdot$$

$$\cdot \prod_{i=1}^{nF_n(t)} \ell(N,n-i)[S_0(x_i)]^{\ell(N,-1)\cdot\ell(N.0)\cdot\ldots\cdot\ell(N,n-1+1)[\ell(N,n-i)-1]}$$

Let us see something of its interpretation: Of course, $S_0(t) = E[S(t)]$ reflects our prior guess regarding S(t). With respect to our prior "strength of belief", let us represent $1(N,\alpha)$: if $\mathcal{N}$ is the set of the Lévy measures of homogeneous processes, we have,

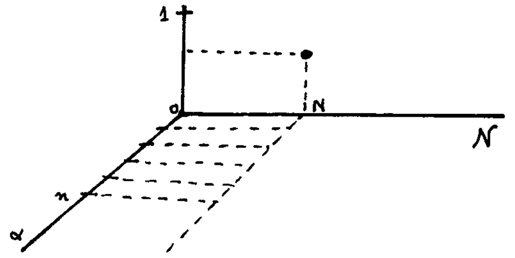

Figure 1 : $l(N,\alpha)$.

and $0 \leq 1(N,0) \leq 1$ can be interpreted like our degree of belief in our prior estimation $S_0(t)$. We shall return to this later.

4. LINEAR APPROACH.

We saw before that the Bayes rule is not specially easy to apply with a particular class of prior distributions, the processes neutral to the right. Even when we consider prior homogeneous processes neutral to the right we have some difficulties with the interpretation of certain parameters. We give here a Bayes rule, which is very easy to apply with all prior distributions, for which the two prior moments E[S(t)] and $E[S^2(t)]$ exist. This is very near to the posterior mean in a lot of situations, and it is the same in some of them, for instance the Dirichlet prior. Namely, we look for the Bayes rule inside the set of linear combinations of some given set of sample functions.

Survival function estimation:

Let S(t) be a random survival function. We make here the classical Bayesian analysis: we select a random sample os size n_1 and look for the Bayes rule $\hat{S}_1(t)$, with quadratic loss, for $S(t)$ inside the set of decision rules like ${}^{1}a_t S_{n_1}(t) + {}^{1}b_t S_o(t)$, with $S_o(t)$ the prior survival function and $S_{m_1}(t)$ the empirical survival function, as before. Then, we select another sample, of size n_2 and again we look for Bayes rule inside the set of decision rules like ${}^{2}a_t\, S_{m_2}(t) + {}^{2}b_t\, \hat{S}_1(t)$. We continue with the process taking samples of sizes $n_3, n_4, \ldots, n$, finding the respective Bayes rules $\hat{S}_3(t), \hat{S}_4(t), \ldots, \hat{S}_n(t)$, and looking each time inside the set of decision rules ${}^{i}a_t\, S_{m_i}(t) + {}^{i}b_t\, \hat{S}_{i-1}(t)$.

Under these conditions, the Bayes rule is,

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

and the minimum associated Bayes risk,

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

where

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

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

The result is followed minimizing the Bayes risk

$$\int_{\mathcal{F}}\int_{\mathcal{X}} (S(t) - a\, S_{n_i}(t) - b\, \hat{S}_{i-1}(t))^2\, dQ(S_{n_i}(t))\, d\mathcal{P}(S)$$

where Q is the distribution of $S_{n_i}(t)$ in the sample and $\mathcal{X}$ the sample space. In the rest we shall note the estimator $\hat{S}_n(t)$ by $\hat{S}(t)$.

<u>Theorem 4.1</u>:

For any prior distribution $\mathcal{P}$ with the two first moments, let l_t be,

$$l_t = \frac{\log E[S^2(t)]}{\log E[S(t)]} - 1$$

then, if $S_o(t) = E[S(t)]$,

(a) $E[S^2(t)] = [S_o(t)]^{\ell_t+1}$

(b) $0 \le l_t \le 1$.

(c) $l_t \to 1 \Leftrightarrow V(S(t)) \to 0$.

(d) $l_t \to 0 \Leftrightarrow V(S(t))$ is the greatest.

(a) is followed by sustitution, (c) because $V(S(t)) = 0 \Leftrightarrow E[S^2(t)] = E^2[S(t)] \Leftrightarrow l_t = 1$. (b) because $V(S(t)) \neq 0 \Leftrightarrow l_t \le 1$, and because $S(t) \le 1$, (d) because $V(S(t)) = [S_o(t)]^{\ell_t+1} - [S_o(t)]^2$ is maximum $\Leftrightarrow l_t$ is minimum $\Leftrightarrow l_t \to 0$.

<u>Remark 1</u>: The smaller $V(S(t))$ the bigger will be l_t, which is a function of $V(S(t))$.

<u>Remark 2</u>: l_t reflects our degree of belief in our prior estimation $S_o(t)$. We shall assign the number <u>o</u> to l_t when and only when we think $S_o(t)$ is the worst prior estimation of $S(t)$. We shall increase l_t when $S_o(t)$ is better, and we shall give the greatest value to l_t, <u>1</u>, when and only when $S_o(t)$ is the best prior estimation of $S_o(t)$.

<u>Remark 3</u>: Now, we can express our linear estimator, for any prior distribution $\mathcal{P}$ with variance, as

$$\hat{S}(t) = \frac{n[S_o(t)]^{\ell_t} - n\,S_o(t)}{1+(n-1)[S_o(t)]^{\ell_t} - n\,S_o(t)}\, S_n(t) + \frac{1 - [S_o(t)]^{\ell_t}}{1+(n-1)[S_o(t)]^{\ell_t} - n\,S_o(t)}\, S_o(t)$$

with minimum associated Bayes risk,

$$R_{min}(n) = \frac{1 - [S_o(t)]^{\ell_t}}{1+(n-1)[S_o(t)]^{\ell_t} - n\,S_o(t)} \cdot \left([S_o(t)]^{\ell_t+1} - S_o^2(t)\right)$$

which only depends on $S_o(t)$, our prior estimation of $S(t)$, and on l_t, the degree of belief in it. When $l_t \to 0$, $\hat{S}(t) \to S_n(t)$, and when $l_t \to 1$, $\hat{S}(t) \to S_o(t)$, so when our prior degree of belief in our prior estimation is near 0 (very little) the estimator $\hat{S}(t)$ does not weigh the prior component

($p_n(t) \to 0$) and we have the usual nonparametric estimator $S_n(t)$. When $l_t \to 1$, which -- means that we think that $S_o(t)$ is a very very good approximation of $S(t)$, $\hat{S}(t)$ pays attention to the prior component ($1-p_n(t) \to 0$) and $\hat{S}(t) \to S_o(t)$. And all this holds true for any prior distribution with finite variance.

Remark 4: l_t in some occasions does not depend on t, for instance, with the homogeneous processes neutral to the right, because in this case, $l_t = l(N,0) = l(N)$. In this case also we have a good interpretarion of $l(N,\alpha)$. A solution for $\alpha \neq 0$ is to consider them equal to $l(N,0)$. Some homogeneous processes, such as the exponential gamma or the simple homogeneous, have a Lévy measure determined by a parameter $c \geq 0$, and in these cases $l(N,\alpha) = l(c,\alpha)$, and c represents our degree of belief.

Remark 5: The prior $\mathcal{P}$ does not necessarily pull out discrete distributions. The reason for which the coefficient of the sample and the prior components add one is given in the next theorem.

Theorem 4.2

Let $\theta \in \Theta$ be a parameter and let Π be a prior distribution on Θ such that $E_\pi[\theta] \neq 0$. If $E_\pi\, E_m[X_i] = E_\pi[\theta]$, $i=1,\ldots,n$, ehere 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

$$a_1 X_1 + \ldots + a_n X_n$$

is such that the estimations $\hat{a}_1,\ldots,\hat{a}_n$ add one, i.e., $\hat{a}_1+\ldots+\hat{a}_n=1$.

The $\hat{a}_1,\ldots,\hat{a}_n$ are such that the Bayes risk $\int_\Theta \int_X (\theta - a_1 x_1 - \ldots - a_n x_n)^2 dQ d\Pi$ must verify $\hat{a}_1 X_1+\ldots+\hat{a}_n X_n = E_\pi[E_m(\theta)]$, and taking expectations we get the result.

Theorem 4.3:

Let $\hat{S}(t)$ be the Bayes rule for $S(t)$, with quadratic loss, looked for inside the set of decision rules $a\,S(t) + b\,S(t)$. Let g 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 quadratic loss and the same prior distribution, for $g(\hat{S}(t))$, with the minimum Bayes risk being equal. The mentioned set is $g(aS_n(t) + bS_o(t))$.

The result os got by writing $g(x) = Ax + B$ and taking minima.

Remark 5: The theorem 4.3 can be extended, considering instead of linear functions, more general ones and using the hyperplane representation with bilinear etc., functions.

Remark 6: The linear estimator for the distribution function F(t) is, $\hat{F}(t) = [1-p_n(t)]F_n(t)+p_n(t)F_o(t)$, where $F_o(t) = 1 - S_o(t)$.

Remark 7: Although sometimes we have supposed that $t \geq 0$, even that times this restriction is unnecessary, and all the results got are valid for a random variable $-\infty < X < +\infty$.

Comparison between the posterior mean and the linear estimator:

We know that the Bayes rule without restrictions 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 without restrictions. This is considered like the regression curve of $S(t)$ on $S_n(t)$, while our estimator $\hat{S}(t)$ represents the regression line of $S(t)$ on $S_n(t)$. So, when we approximate the posterior mean by $\hat{S}(t)$, we really determine the regression line instead of the regression curve, and when $\mathcal{P}$ is the induced by a Dirichlet process, both are the same.

Let us make a more precise comparison with a sample of size one, and in the case of an homogeneous process,

$$\hat{S}(t) = \begin{cases} S_o(t)\dfrac{1-[S_o(t)]^l}{1-S_o(t)} & \text{if } x \le t \\ [S_o(t)]^l & \text{if } x > t \end{cases}$$

and

$$E\,S(t)/x = \begin{cases} S_o(t).l.[S_o(x)]^{l-1} & \text{if } x \le t \\ [S_o(t)]^l & \text{if } x > t \end{cases}$$

which is like this

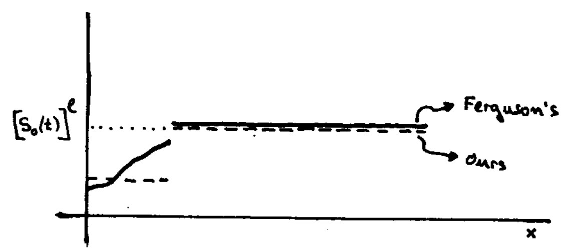

Figura 2 : Estimators.

5. SAMPLE CHARACTERISTICS.

In any easy way we have the next results:

Result 1: $\hat{S}(t)$ takes the value

$$[1-p_n(t)]\,\frac{k}{n} + p_n(t)\; S_o(t) \quad \text{with}$$

probability $\binom{n}{k}[S(t)]^k\;[1-S(t)]^{n-k}$,

$k=0,1,2,\ldots,n.$

Result 2: $E_m[\hat{S}(t)] = S(t)+p_n(t)[S_o(t)-S(t)]$,

and because $p_n(t) \underset{n\to\infty}{\to} 0$, $\hat{S}(t)$ is unbiased asymptotically.

$V_m(\hat{S}(t)=[(1-p_n(t))^2/n]S(t)(1-S(t))$.

Result 3: $\hat{S}(t) \overset{a.s.}{\to} S(t)$, $\hat{S}(t) \overset{P}{\to} S(t)$, $\hat{S}(t)$ is consistent for $S(t)$.

Result 4: $\hat{S}(t) \underset{n\to\infty}{\overset{d}{\to}} N\;(S(t)+p_n(t)[S_o(t)-S(t)];$

$$(1-p_n(t))\sqrt{\frac{S(t)(1-S(t))}{n}})$$

6. GOODNESS OF FIT TEST WITH PRIOR INFORMATION A DIRICHLET PROCESS.

Let us suppose we want to test if the underlying distribution is some specified distribution $F_1(x)$, and also let us suppose we know something about the unknown and true distribution function $F(x)$, i.e., we are able to provide a prior distribution function $F_o(x)$ and a degree of belief in it l_t. We suppose here that the prior distribution $\mathcal{P}$ is the induced by a Dirichlet process with parameter α. In this situation $E[F(t)/\text{data}] = \hat{F}(t)$, the linear estimator, and

$$p_n(t) = p_n = \frac{\alpha(\mathbb{R})}{\alpha(\mathbb{R})+n}$$

$$l_t = [\log\,(\frac{\alpha(\mathbb{R})-\alpha(t)+1}{\alpha(\mathbb{R})+1})]/\;[\log(\frac{\alpha(\mathbb{R})-\alpha(t)}{\alpha(\mathbb{R})})]$$

We suggested here the statistic for the goodness of fit test, to test $H_o: F(x)=F_1(x)$ against $H_1: F(x)\neq F_1(x)$,

$$V_n=(1-p_n)\sup_x|F_n(x)-F(x)|+p_n\;\sup_x|F_o(x)-F(x)|$$

Remark 8: $F_o(x)$ is a very good prior estimation $\Leftrightarrow l \to 1 \Leftrightarrow p_n \to 1 \Leftrightarrow V_n \to \sup_x|F_o(x) - F(x)|$.
So, if $F_o(x)$ is a good prior estimation of $F(x)$, if H_o is true, the differences between $F_o(x)$ and $F_1(x)$ should be small x. Then, large values of V_n, under H_o, tend to descredit the null hypotheses, and so, will reject H_o when $V_n > V_{n,\alpha}$.

$F_o(x)$ is a very bad prior estimation $\Leftrightarrow$ $\Leftrightarrow l \to 0 \Leftrightarrow p_n \to 0 \Leftrightarrow V_n \to D_n$ (the Kolmogorov-Smirnov statistic), and so, if $F_o(t)$ is a very bad prior estimation, our statistic does not weig the prior information and we have the usual goodness of fit test which rejects again H_o for big values of V_n.

In intermediate situations our statistic will take upon itself to weigh the two components in the right way.

So, to test $\quad H_o: F(x) = F_1(x)$
$\qquad\qquad\quad H_1: F(x) \neq F_1(x)$

with significance level α, we will reject H_o if and only if $V_n>V_{n,\alpha}$, where $V_{n,\alpha}$ is such that $\alpha = P\{V_n>V_{n,\alpha}/H_o\}$ =

$$= P\{D_n>(V_{n,\alpha}-cp_n)/(\alpha p_n)/H_o\}$$

where $D_n=\sup_x|F_n(x)-F(x)|$ is the Kolmogorov-Smirnov statistic, $c = \sup_x|F_o(x)| \equiv \text{constant} \le 1$;

and if $D_{n,\alpha}$ is the value such that $P\{D_n > D_{n,}\ /H_o\} = \alpha$, which can be determined as usual trhough the Birnbaum tables, it is,

$$V_{n,\alpha} = (1 - p_n)\ D_{n,\alpha} + c\ p_n$$

Remark 9: Because,

$$P\{V_n > V_{n,\alpha}/H_1\} = P\{D_n > D_{n,\alpha} +$$

$$+[p_n/(1-p_n)][c - \sup_x |F_o(x) - F(x)|]/H_1\} <$$

$$P\{D_n > D_{n,\alpha}/H_1\} \Leftrightarrow$$

$$\Leftrightarrow \text{under } H_1,\quad x \in \mathbb{R}, |F_o(x) - F(x)| < o$$

we conclude that the Kolmogorov-Smirnov test will be more powerful for an alternative F(x) than ours, if and only if $\forall x \in \mathbb{R}$ F(x) is inside the band $F_o(x) - c$, $F_o(x) + c$

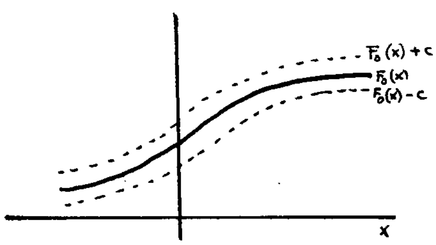

Figura 3 : Band for an alternative.

So, if there exists at least one $x \in \mathbb{R}$ for which F(x) is out of the band, our test will be more powerful than Kolmogorov-Smirnov. We see also that if the null hypotheses $F_1(t)$ is near our prior estimation $F_o(t)$, c will be smaller and so we shall get more powerfull tests. May be if the null hypothesis is accepted, it can be used like a new prior estimation, and in this way, instead of changing our prior belief through Bayes theorem, we can change it through hypothesis testing, where we include our prior estimation $F_o(x)$ and a degree of belief in it, l, and the sample information, $F_n(x)$.

Remark 10: Instead of rejecting H_o when $V_n > V_{n,\alpha}$, a different way of proceeding is to reject H_o when $V_n > D_{n,\alpha}$. In this way, we make better use of the prior information.

Remark 11: In an obvious way we can construct one-sided goodness of fit tests which would detect directional differences, using the statistics

$$V_n^+ = (1-p_n)D_n^+ + p_n\ a = (1-p_n)\sup_x(F_n(x) - F(x)) +$$

$$+ p_n \sup_x(F_o(x) - F(x))$$ and

$$V_n^- = (1-p_n)D_n^- + p_n\ b = (1p_n)\sup_x(F(x) - F_n(x)) +$$

$$+ p_n \sup_x(F(x) - F_o(x)).$$

For the alternative

$$H_{1,+} : F(x) \geq F_1(x) \quad , \quad \forall x \in \mathbb{R}$$

The rejection region is $V_n^+ > V_{n,\alpha}^+$, where $V_{n,\alpha}^+ = (1-p_n)D_{n,\alpha}^+ + a\ p_n$ and for the alternative

$$H_{1,-} : F(x) \leq F_1(x) \quad , \quad \forall x \in \mathbb{R}$$

H_o is rejected when $V_n^- > V_{n,\alpha}^-$, where $V_{n,\alpha}^- = (1-p_n)D_{n,\alpha}^- + p_n\ b$ and $D_{n,\alpha}^-$ and $D_{n,\alpha}^+$ are the usual critical points in the one-sided Kolmogorov-Smirnov goodness of fit test.

Remark 12: When n>35 the value $V_{n,\alpha}$, $V_{n,\alpha}^-$ and $V_{n,\alpha}^+$ can be determined from the asymptotic distribution.

Remark 13: Because of theoretical properties about the necessary constituity of F(x), which is not assumed here because $\mathcal{P}$ is a Dirichlet prior, the tests exposed here are certainly conservative.

Remark 14: Because of the relations betweeen survival functions and distribution function (F(t) = 1 - S(t)), the tests exposed here can easily be extended to test survival functions instead of distributions functions.

7. EXAMPLE.

Let us suppose we want to know the probability law of the fasting blood glucose determinations of nonobese, apparently healthy, adult males which constitutes our population of interest.

We have got some previous information about this variable. We have got 36 determinations made on people with the same characteristics as our population of interest, but of another geografic area. The determinations which constitutes the prior information $F_o(x)$ are shown in table 1.

TABLE 1

value	68	72	75	76	77	78	80	81	84	86	87	92
frequency	2	2	2	2	6	3	6	3	2	2	2	4

These determinations are apparently normally distributed, and give a sample mean of 80,08/100 ml., and a quasi-standard deviation of 6,19. So, it is natural to test the null hypothesis that in our population of interest (which is think to be not very different from the previous data) the characteristic, with cumulative distribution function F(x), is distributed N(80, 6,2). So, we establish the hypothesis

$$\begin{cases} H_o: F(x) = N(80, \; 6,2) \\ H_1: F(x) \neq N(80, \; 6,2) \end{cases}$$

To test the hypothesis, we select a random sample of our population of interest and the values obtained are shown in table 2.

TABLE 2

value	69	71	74	77	80	84	87	91
frequency	1	2	1	2	3	2	2	2

To perform the test, we compute the values $c = \sup_x |F_o(x)-F_1(x)|$ where $F_1(x) \equiv N(80, 6,2)$, (table 3), and $\sup_x |F_n(x)-F_1(x)|$, (table 4).

If we think that our prior information is such that the ponderations p_n and $1-p_n$ are equal to 1/2, we have that

$$V_n = (1-p_n) \sup_x |F_n(x)-F_1(x)| +$$

$$+ p_n \sup_x |F_o(x)-F_1(x)| = 0,1504$$

If our significant level was $\alpha = 0,01$, will be $V_{n,\alpha} = 0,2813$.

Therefore, we are not willing to reject H_o. Indeed, our p-value is

$$\text{p-value} = P\{V_n > 0,1504/H_o\} > 0,32 \; .$$

TABLE 3

x_i	$F_o(x_i)$	$F_1(x_i)$	$\lvert F_o(x_i)-F_1(x_i)\rvert$	$\lvert F_o(x_{i-1})-F_1(x_i)\rvert$
68	0,0556	0,0262	0,0294	
72	0,1111	0,0985	0,0126	0,0429
75	0,1667	0,2090	0,0423	0,0979
76	0,2222	0,2578	0,0356	0,0911
77	0,3889	0,3156	0,0733	0,0934
78	0,4722	0,3745	0,0977	0,0144
80	0,6389	0,5000	0,1389	0,0278
81	0,7222	0,5636	0,1586	0,0753
84	0,7778	0,7422	0,0356	0,0200
86	0,8333	0,8340	0,0007	0,0562
87	0,8889	0,8708	0,0181	0,0375
92	1,0000	0,9738	0,0262	0,0849

TABLE 4

x_i	$F_n(x_i)$	$F_1(x_i)$	$\lvert F_n(x_i)-F_1(x_i)\rvert$	$\lvert F_n(x_{i-1})-F_1(x_i)\rvert$
69	0,0667	0,0384	0,0283	.
71	0,2000	0,0735	0,1265	0,0068
74	0,2667	0,1660	0,1007	0,0340
77	0,4000	0,3156	0,0844	0,0489
80	0,6000	0,5000	0,1000.	0,1000
84	0,7333	0,7422	0,0089	0.1422
87	0,8667	0,8708	0,0041	0,1375
91	1,0000	0,9616	0,0384	0,0949

If we follow the observation of remark 10, we have that we accept H_o again because $D_{n,\alpha} = D_{n,0,01} = 0,404$. Observe that the clasic Kolmogorov-Smirnov test would also accept because $D_n = 0,1422$, so, following remark 10, we move the classic Kolmogorov-Smirnov statistic D_n , introducing the prior information.

8. REFERENCES:

/1/ DOKSUM, K.: "Tailfree and neutral random probabilities and their posterior distributions". Ann. Probability, 1, 183-201. (1974).

/2/ FERGUSON AND PHADIA: "Bayesian nonparametric estimation based on censored data". Ann. Statistics, 7, 163-186 (1979).

/3/ GOLDSTEIN, M.: "A note on some Bayesian nonparametric estimate". Ann. Statistics 3, 736-740. (1975).

/4/ QUESADA, V. and GARCIA PEREZ, A.: "Estimation of random survival functions: A linear approach". QUESTIIO, V.6 nº 2 (June, 1982).