Qüestiió - V. 9, n.º 1 (març 1985) pp. 1-5

On the Selection of the Parameters of an Exponential Gamma Process Prior in Bayesian Nonparametric Estimation

DOMINGO MORALES
UNIVERSIDAD COMPLUTENSE DE MADRID

Consider a nonparametric Bayesian estimation problem where the Statistician has decided to use an exponential gamma process prior. This paper deals with the selection problem of the process parameters.

Some algorithm to determine the parameter c from the prior guess and the strength of belief are given.

The case where this last concept changes with time is also studied.

Keywords: BAYESIAN NONPARAMETRIC ESTIMATION, EXPONENTIAL GAMMA PROCESS, PRIOR GUESS, STRENGTH OF BELIEF, ALGORITHM TO DETERMINE THE SCALE PARAMETER.

AMS 1980 Subject classification: Primary 62C10, Secondary 62G99

1. INTRODUCTION.

Suppose that in a nonparametric Bayesian decision problem the Statistician has decided to use an exponential gamma process (E.G.P.) What is the criterion to select, within the class of all E.G.P., the most advantageous E.G.P. to treat an estimation problem?

An E.G.P. has two parameters: $\Lambda(t)$ and c. So, it is necessary to give some criteria to determine $\Lambda(t)$ and c on the basis of the prior guess and the strength of belief.

$\Lambda(t)$ is chosen in such a way that the marginal expectation of the E.G.P., $E_{\Lambda,c}(F(t))$, is the prior guess $F_o(t)$; and thus

$\Lambda(t)=\ln(1-F_o(t))/\ln(\frac{c}{c+1})^c)$, where c is not already specified.

How to select c?

Ferguson /5/ said that c measures the prior strength of belief in some sense; i.e., if we can rely on our prior guess, c should be "large". Otherwise c should be "small". But, what do we understand by "large"?

Observe that $\hat{F}(t)$ will depend on the sample $(x_1,...,x_n)$, and on the process parameters $(\Lambda(t), c)$. Consequently, the value of c will influence our estimation.

Clearly, it is necessary to have an accurate method to translate the strength of belief in the parameter c.

When we ask for the strength of belief, in a nonparametric Bayesian procedure, we will receive an answer like: bad, moderate, good.

At that time, we will have to translate that qualitative value in a numerical quantity, whether belonging to $(0,\infty)$ if we directly determine c, or belonging to $(0,1)$ if we use:

$$l(c)=\ln(\frac{c+2}{c+1})/\ln(\frac{c+1}{c}).$$

How can we be sure that the chosen numerical value will not cause an important bias on the final estimation?

Our purpose is to give an algorithmical method, which allows us to determine the most advantageous value for the parameter c in a nonparametric Bayesian estimation problem.

- Domingo Morales - Universidad Complutense de Madrid - Dep. de Estadística e I.O. Ciudad Universitaria - Madrid 28040

- Article rebut el novembre de 1984.

2. PRELIMINARIES.

Consider the topological space (R^+, T_{us}), where $R^+=[0,\infty)$ and T_{us} denotes its usual topology. The space $M(R^+)$ is the collection of all probability measures on R^+ and T_d is the topology of the weak convergence on $M(R^+)$. $(M(M(R^+)), T_D)$ is the space of all probability measures $\mathscr{P}$ on $M(R^+)$ (i.e., priors) together with the topology of weak convergence derived from $(M(R^+), T_d)$.

Consider $F, Q \in M(R^+)$, $F(t)=F((-\infty,t]))$. The Levy distance between F and Q is given by:

$$d_L(F,Q)=\inf\{a>0/\forall x \quad F(x-a)-a<G(x)<F(x+a)+a\}.$$

This distance metrizes the topology of the weak convergence. Furthermore, it has the property that $\sqrt{2}d_L(F,Q)$ is the maximum euclidean distance between the graphs of F and Q, measured along the 135º - direction.

Obviously, the following family of subsets is a neighbourhood base of $F\in M(R^+)$ in the topology of the weak convergence:

$$B(F,a)=\{Q\in M(R^+) \ : \ d_L(F,Q)<a\} , \ a>0$$

Consider the sets $N(F,a,k,t_1,\ldots,t_k) = \{Q\in M(R^+)/|F(t_i)-Q(t_i)|<a \quad i=1,\ldots,k\}$, $a>0$, $k\in N$, $t_1,\ldots,t_k \in C(F)$, where $C(F)$ is the set of continuity points of F.

<u>Proposition 2.1</u>.

$N_F=\{N(F,a,k,t_1,\ldots,t_k): a>0,\ k\in N,\ t_1,\ldots,t_k\ C(F)\}$ is a neigbourhood base of $F\in M(R^+)$ in T_d. Furthermore $N(F,a/2,k,t_1,\ldots,t_k)\subset B(F,a)$ $\forall N(F,a/2,k,t_1,\ldots,t_k)$ s.t. $a>0$, $k\in N$, $t_1,\ldots,t_k\in C(F)$, $t_i-t_{i-1}<a$, $F(t_1)<a/2$, $F(t_k)>1-a/2$.

Proof: Proof similar to theorem 2.3.3. in Huber /7/.

The following distance also metrices the topology of the weak convergence.

Consider $F,Q\in M(R^+)$. The Prokhorov distance between F and Q is given by:

$$d_p(F,Q)=\inf\{a>0/F(G)\geq Q(G_a)-a \quad \forall G\in T_{us}\},$$

where $G_a=\{x\in R^+/\inf_{y\in G^c} d(x,y)\geq a\}$

In $(M(M(R^+)),T_D)$, the Prokhorov distance between $\mathscr{P}$ and $\mathscr{A}$ is defined in a similar way:

$$d_p(\mathscr{P},\mathscr{A})=\inf\{a>0/\mathscr{A}(G)\geq \mathscr{P}(G_a)-a \quad \forall G\ T_d\},$$

where $G_a=\{Q\in M(R^+)/\inf_{P\in G^c} d_L(Q,P)\geq a\}$. See /2/ for further details about the topology of the weak convergence.

3. A METHOD TO SELECT THE PARAMETER c OF AN E.G.P.

Suppose that $Q(t)$ is the prior guess. In that case,

$\Lambda(t)=\ln(1-Q(t))/\ln(\frac{c}{c+1})^c$ is selected (where c is not already specified).

The following criterion is proposed: "Given $Q\in M(R^+)$, $\varepsilon>0$, $\alpha>0$. Find $c_o(Q,\varepsilon,\alpha)$ such that $\forall c\geq c_o$ $\mathscr{P}_{\Lambda,c}(B(Q,\varepsilon))>1-\alpha$, where $\mathscr{P}_{\Lambda,c}$ denotes the distribution of the E.G.P. we are looking for".

Remember that $\sqrt{2}d_L(Q,F)$ is the maximum euclidean distance between Q and F, measured along the 135º - direction.

This fact allows us to draw $B(Q,\varepsilon)$. If $Q(t)$ is continuous, the strip is even more easy to draw.

Then, we'll ask the specialist to translate his strength of belief in two numbers $\sqrt{2}\varepsilon$ and α such that the true distribution function we want to estimate, $F(t)$, verifies $d_L(F,Q)<\varepsilon$ with probability $1-\alpha$. Therefore, the specialist will have to give us a strip around $Q(t)$ and his subjective probability of $F(t)$ to be inside the strip.

In this section we find an algorithm which determines c_o from the knowledge of Q,ε,α.

Remember that $B(Q,\varepsilon)\subset N(Q,\varepsilon/2,k,t_1,\ldots,t_k)$ $\forall N(Q,\varepsilon/2,k,t_1,\ldots,t_k)$, verifying the conditions in Prop. 2.1. Therefore, it is enough to find $c_o(Q,\varepsilon,\alpha)>0$ such that

$\forall c \geq c_0 \quad \mathscr{P}_{\Lambda,c}(N(Q,\varepsilon/2,k,t_1,\ldots,t_k)) > 1-\alpha$ for some k and $t_1,\ldots,t_k \in C(Q)$, verifying the above conditions.

This is equivalent to $\mathscr{P}(N^c(Q,\varepsilon/2,k,t_1,\ldots,t_k)) < \alpha$ Now we have

$$\mathscr{P}_{\Lambda,c}(N^c(Q,\varepsilon/2,k,t_1,\ldots,t_k)) = \mathscr{P}_{\Lambda,c}\left(\bigcup_{i=1}^{k} \{F: |F(t_i)-Q(t_i)| \geq \varepsilon/2\}\right) \leq \sum_{i=1}^{k} \mathscr{P}_{\Lambda,c}(\{F: |F(t_i)-Q(t_i)| \geq \varepsilon/2\}) \leq$$

$$< \sum_{i=1}^{k} \frac{4}{\varepsilon^2} E_{\Lambda,c}((F(t_i)-Q(t_i))^2) = \frac{4}{\varepsilon^2} \sum_{i=1}^{k} V_{\Lambda,c}(F(t_i)),$$

where $V_{\Lambda,c}(F(t))$ denotes the marginal variance of the E.G.P. with parameters $(\Lambda(t),c)$.

Remark 3.1.

If $\Lambda(t) = \ln(1-Q(t))/\ln\left(\frac{c}{c+1}\right)^2$, then

$E_{\Lambda,c}(F(t)) = Q(t)$ and

$$V_{\Lambda,c}(F(t)) = \left(\frac{c}{c+2}\right)^{\ln(1-Q(t))/\ln\left(\frac{c}{c+1}\right)} - (1-Q(t))^2$$

The following algorithm is proposed: "Find $k \in N$ and $t_1,\ldots,t_k \in C(Q)$ according to proposition 2.1. Select c_0 such that

$$\frac{4}{\varepsilon^2} \sum_{i=1}^{k} V_{\Lambda,c}(F(t_i)) < \alpha".$$

The following result shows that the algorithm is convergent.

Lema 3.1.

$\forall t > 0$ fixed $g(t,c) = V_{\Lambda,c}(F(t))$ decreases monotonously in $c > 0$.

Proof: Use standard calculus techniques.

4. ON USING E.G.P. WHEN THE STRENGTH OF BELIEF IS A FUNCTION OF THE TIME.

In the previous section we have explained how to express numerically the strength of belief using the Levy distance, then we have given a procedure to obtain the parameter c. The already mentioned method supposes that the strength of belief is constant, and therefore independent of time. Usually, this hypothesis is not true because the reliance the specialists have in their prior guess depends significantly on the different zones of the time axis. If those differences are small, the previous section technique should be used because it is useless to complicate a model unnecessary. However, if those differences are large, the use of the proposed method is questionable.

In this section we will express the strength of belief as a function of the time $2\varepsilon(t)$. We are interested in analizing, in a Bayesian context, the possibility of having a strength of belief which changes significantly with time.

A study of this problem, giving a solution with practical usefulness, will have the following steps:

(1) Decide the type of random probability measures to be used.

(2) Quantify the strength of belief.

(3) Choose a numerical method to determine the E.G.P. parameters, using as initial data the prior guess and the quantification of the prior strength of belief.

(4) Estimate the distribution (survival) function.

In step (1) we have decided to use E.G.P. Concerning step (4) see the derivation of the posterior expectation of an E.G.P. in Doksum /4/. We will develop the remainder steps in the rest of this section.

Suppose our prior guess is $Q(t)$. The prior strength of belief is a qualitative concept, and therefore ambiguous. So, we need a method to numerically translate this concept for every value of time. To solve this problem we propose a positive function $\varepsilon(t)$ ($\varepsilon: R^+ \longrightarrow R^+$) such that the true distribution function $F(t)$ must verify:

$\mathscr{P}_{\Lambda,c}(F \in B(Q,2\varepsilon(t))) > 1-\alpha$, where:

(a) $\mathscr{P}_{\Lambda,c} \in M(M(R^+))$ denotes the distribution of an E.G.P. with parameters $(\Lambda(t),c)$.

(b) $B(Q,2\varepsilon(t)) =$

$= \{G \in M(R^+): \forall t \geqslant 0 \ \inf\{a>0: Q(t-a)-a < G(t) < Q(t+a)+a\} < 2\varepsilon(t)\}$.

Remark 4.1.

$\sqrt{2}(2\varepsilon(t))$ is the maximum euclidean distance, measured along the 135º-direction, between the point $(t,G(t))$ and the graph of $Q(t)$.

Firstly, we consider a function $\varepsilon(t)$ taking only a finite number of values; and secondly, a continuous function. In both cases we give algorithms to determine the parameter c of the E.G.P. we will use in a Bayesian estimation problem.

$\varepsilon(t)$ taking a finite number of values

Let $x_o<x_1<x_2<\ldots x_k<x_{k+1}$ be a partition of R^+, where $x_o=0$ and $x_{k+1}=+\infty$. $\varepsilon(t)=\varepsilon_i$ if and only if $t\varepsilon[x_{i-1},x_i)$, $i=1,\ldots,k+1$.

Under these conditions, the following result is obtained.

Proposition 4.1.

Let $\{t_j^i\}\subset C(Q)$ $(i=1,\ldots,k+1;\ j=1,\ldots,k_i)$ be a partition of R^+ such that:

(1) $t_1^i=x_{i-1}$ $\quad i=1,\ldots,k$

(2) $t_j^i\varepsilon[x_{i-1},x_i)$ $\quad i=1,\ldots,k+1$; $j=1,\ldots,k_i$

(3) $t_j^i-t_{j-1}^i<\varepsilon_i$ $\quad i=1,\ldots,k+1$; $j=2,\ldots,k_i$

(4) $x_i-t_{k_i}^i<\varepsilon_i$ $\quad i=1,\ldots,k$

(5) $Q(t_2^1)<\varepsilon_1/2$

(6) $Q(t_{k_{k+1}}^{k+1})>1-\varepsilon_k/2$

(7) $\varepsilon_{i+1}<2\varepsilon_i$ $\quad i=1,\ldots,k$

Then, $N(Q,\varepsilon(t)/2,\ \{t_j^i\}\ )=$

$=\{F\varepsilon M(R^+):|F(t_j^i)-Q(t_j^i)|<\varepsilon_i/2\quad i=1,\ldots,k+1;$
$j=1,\ldots,k_i\}\subset B(Q,\varepsilon(t))$.

Proof: Proof similar to proposition 2.1.

We want to find the smallest value of the parameter c verifying: $\overline{\mathscr{P}}_{\Lambda,c}(B(Q,2\varepsilon(t))\geq 1-\alpha$, where $\Lambda(t)=\ln(1-Q(t))/\ln(\frac{c}{c+1})^c$.

As $N(Q,\varepsilon(t),\{t_j^i\})\subset B(Q,2\varepsilon(t))$, it suffices to show that $\mathscr{P}_{\Lambda,c}(N^c(Q,\varepsilon(t),\{t_j^i\}))\leq\alpha$

Now, $\mathscr{P}_{\Lambda,c}(N^c(Q,\varepsilon(t),\{t_j^i\}))=$

$$\mathscr{P}_{\Lambda,c}\Big(\bigcup_{i=1}^{k+1}\bigcup_{j=1}^{k_i}\{F\varepsilon M(R^+):|F(t_j^i)-Q(t_j^i)|\geq\varepsilon_i\}\Big)\leq$$

$$\leq\sum_{i=1}^{k+1}\sum_{j=1}^{k_i}\frac{1}{\varepsilon_i^2}V_{\Lambda,c}(F(t_j^i))=(1),\text{ where}$$

$V_{\Lambda,c}(F(t_j^i))$ has been defined in remark 3.1.

Therefore, we propose the following algorithm: "Find k, $k_i\varepsilon N$ and $\{t_j^i\}\subset C(Q)$ according to proposition 4.1. Select c such that $(1)\leq\alpha$".

$\varepsilon(t)$ continuous

Proposition 4.2

Let $\varepsilon(t)$ be a continuous function. Let $0<t_1<\ldots<t_k\varepsilon C(Q)$ be a partition of R^+ such that:

(1) $\varepsilon(t_1)=\varepsilon(t_k)=\varepsilon\leqslant\varepsilon(t)\quad\forall t\varepsilon[t_1,t_k]$

(2) $Q(t_1)<\varepsilon/2$ and $Q(t_k)>1-\varepsilon/2$

(3) $t_{i+1}-t_i<\varepsilon\quad i=1,\ldots,k-1$

(4) $\varepsilon(t)=\varepsilon\quad\forall t\varepsilon R^+-[t_1,t_k]$

(5) $2\varepsilon(t)>\max\{\varepsilon(t_i),\varepsilon(t_{i+1})\}\ \forall t\varepsilon[t_i,t_{i+1})$
$i=1,\ldots,k-1$

Then, $N(Q,\varepsilon(t)/2,t_1,\ldots,t_k)=$
$=\{F\varepsilon M(R^+):|F(t_i)-Q(t_i)|<\varepsilon(t_i)/2$
$i=1,\ldots,k\}\subset B(Q,\varepsilon(t))$.

Proof: Proof similar to proposition 2.1.

We want to find the smallest value of the parameter c verifying $\mathscr{P}_{\Lambda,c}(B(Q,2\varepsilon(t))\geq 1-\alpha$, Where $\Lambda(t)=\ln(1-Q(t))/\ln(\frac{c}{c+1})^c$.

As $N(Q,\varepsilon(t),t_1,\ldots,t_k)\subset B(Q,2\varepsilon(t))$, it suffices to show that $\mathscr{P}_{\Lambda,c}(N^c(Q,\varepsilon(t),t_1,\ldots t_k))\leqslant\alpha$.

Now, $\mathscr{P}_{\Lambda,c}(N^c(Q,\varepsilon(t),t_1,\ldots,t_k))=$

$$=\mathscr{P}_{\Lambda,c}\Big(\bigcup_{i=1}^{k}\{F:|F(t_i)-Q(t_i)|\geq\varepsilon(t_i)\}\Big)\leq$$

$$\leq\sum_{i=1}^{k}\frac{2}{\varepsilon^2(t_i)}V_{\Lambda,c}(F(t_i))=(2),\text{ where}$$

$V_{\Lambda,c}(F(t_i))$ has been defined in remark 3.1.

Therefore, we propose the following algorithm. "Find $k\varepsilon N$ and $t_1,\ldots,t_k\varepsilon C(Q)$ according to proposition 4.2. Select c such that $(2)\leq\alpha$".

5. A NUMERICAL EXAMPLE.

We illustrate the application of the procedure obtained in the third section with an exercise.

Suppose the prior guess is $Q(t)=1-e^{-t}$ in $t\geq 0$ (i.e.; an exponential distribution with parameter $\lambda=1$), and the strength of belief has been quantify by ε=EPS=0.3 and α=ALFA=0.1.

Observe that $Q(t)$ and ε are given by the specialists in the subject under study. According to their accumulated knowledge they should be able to propose a prior guess $Q(t)$ of the true distribution function $F(t)$ and to draw a strip round $Q(t)$ such that $1-\alpha$ is their subjective probability of $F(t)$ to be inside the strip. We recall that this strip is built in such a way that making $\sqrt{2}\varepsilon$ to be the maximum euclidean distance between F and Q, measured along the 135º-direction.

We want to find c_o such that

$$\mathscr{P}_{\Lambda,c_o}(B(Q,0.3))\geq 0.9.$$

According to prop. 2.1., we use the following algorithm to calculate $t_i\in C(Q)$ $(i=1,\ldots,k)$.

Let INV_Q be a PASCAL FUNCTION representing the inverse function of Q.

```
BEGIN
T[1]:=INV_Q(1-EPS/2);
I:=1;
WHILE T[I]<=(1-EPS/2)DO
      BEGIN
      T[I+1]:=T[I]+EPS ;
      I:=I+1 ;
      END ;
K:=I ;
END ;
```

We obtain the following $t_i\in C(Q)$: 0.1625, 0.4625, 0.7625, 1.0625, 1.3625, 1.6625, 1.9625.

Using the following algorithm we evaluate the function

$$FF(c)=\frac{4}{(0.3)^2}\sum_{i=1}^{7} V_{\Lambda,c}(F(t_i)),\quad \text{where}$$

$\Lambda(t)$ and $V_{\Lambda,c}(F(t))$ have been defined in Remark 3.1.

```
BEGIN
FF:=0 ;
FOR I:=1 TO K DO
    FF:=FF+(C/(C+2))**(LN(1-Q(T[I]))/LN(C/
                         /C+1)))-(1-Q(T[I]))**2;
FF:=4/(EPS**2)+FF ;
END ;
```

To calculate c_o, we have used the following algorithm.

```
BEGIN
INCR:=1000 ; PREC:=0.1 ;
C:=0 ;
REPEAT
   IF FF(C+INCR)>ALFA
      THEN C:=C+INCR
      ELSE INCR:=INCR/10 ;
UNTIL INCR<PREC ;
END ;
```

We have obtained c_o=348.

6. BIBLIOGRAPHY.

/1/ ANTONIAK, C.: "Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems" Ann. Statist. 2, 1152-1174. (1974).

/2/ BILLINGSLEY, P.: "Convergence of Probability Measures" Wiley, New York.(1968).

/3/ DALAL, S.R. AND GAINEFORD, J.H.: "On approximating parametric Bayes models by nonparametric Bayes models", Ann. Statist. 8, 664-672.

/4/ DOKSUM, K.: "Tailfree and neutral random probabilities and their posterior distributions". Ann Probability. 2, 183-201 (1974).

/5/ FERGUSON, T.S. AND PHADIA, E.G.: "Bayesian nonparametric estimation based on censored data". Ann Statist. 7, 163-186. (1979).

/6/ FERGUSON, T.S.: "A Bayesian analisys of some nonparametric problems" . Ann. Statist. 1, 209-230.

/7/ HUBER, P.J.: "Robust Statistics". Wiley, New York. (1981).