



HAL
open science

Bayesian prediction for stochastic processes

Delphine Blanke, Denis Bosq

► **To cite this version:**

| Delphine Blanke, Denis Bosq. Bayesian prediction for stochastic processes. 2012. hal-00750263v1

HAL Id: hal-00750263

<https://hal.science/hal-00750263v1>

Preprint submitted on 9 Nov 2012 (v1), last revised 28 Dec 2013 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

BAYESIAN PREDICTION FOR STOCHASTIC PROCESSES

DELPHINE BLANKE AND DENIS BOSQ

ABSTRACT. In this paper, we adopt a Bayesian point of view for predicting real stochastic processes. We give two equivalent definition of a Bayesian predictor and study some properties: admissibility, prediction sufficiency, unbiasedness, comparison with efficient predictors. Prediction of Poisson process and prediction of Ornstein-Uhlenbeck process in the continuous and sampled situations are considered. Various simulations illustrate comparison with non-Bayesian predictors.

1. INTRODUCTION

A lot of papers are devoted to Bayesian estimation but Bayesian prediction does not appear very much in statistical literature. However, this topic is important in particular if the number of data is small. In this paper, we study some properties of Bayesian predictors and give examples of applications to prediction of stochastic processes. Various simulations illustrate the results. Our goal is to compare efficiency of Bayesian predictors with non-Bayesian predictors, especially if one has few data at his disposal.

Section 2 presents the general prediction model ; in this context estimation appears as a special case of prediction. The main point of the theory is the fact that, given the data X , a statistical predictor of Y is an approximation of the conditional expectation $\mathbb{E}_\theta(Y|X)$, where θ is the unknown parameter. Section 3 deals with Bayesian prediction : we give two equivalent definitions of a Bayesian predictor linked with the equivalence of predicting Y and $\mathbb{E}_\theta(Y|X)$. However, in some situations, it is difficult to get an explicit form of the Bayesian estimator $\mathbb{E}(\theta|X)$, thus it is more convenient to substitute the conditional expectation of θ with the conditional mode of θ , also called maximum a posteriori (MAP). We recall some properties of the MAP and underscore its link with the maximum likelihood estimator. In Section 4, we study some properties of Bayesian predictors : admissibility, connection with sufficiency and unbiasedness, case where the conditional expectation admits a special form. Section 5 considers the regular case of Poisson process prediction. We compare the unbiased efficient predictor with the Bayesian and the MAP ones. Concerning diffusion processes, Thompson and Vladimirov (2005) study Bayesian prediction but, they obtain an intricate result, difficult to handle ; they don't try to compare their results with classical predictors. For the Ornstein-Uhlenbeck process, we deal with prediction in Section 6 for the centered and non-centered case and with various priors while Section 7 is devoted to the sampled case. Some asymptotic results are given along the paper, but, since the non asymptotic case is the most important in the Bayesian perspective, theoretical and numerical comparisons focus on this point.

2000 *Mathematics Subject Classification.* Primary 62M20, 62F15.

Key words and phrases. Bayesian prediction, MAP, Comparing predictors, Poisson process, Ornstein-Uhlenbeck process.

2. THE PREDICTION MODEL

In the sequel, each space is equipped with a suitable σ -algebra and each application is supposed to be measurable with respect to these σ -algebras. Now, let $(\Omega, \mathcal{A}, \mathbb{P}_\theta, \theta \in \Theta)$ be a statistical model, where θ is the unknown parameter, and let $(X, Y) : (\Omega, \mathcal{A}) \mapsto (E \times F, \mathcal{E} \otimes \mathcal{F})$ be a random vector. One observes X and wants to predict Y . Actually, it is convenient to consider the more general problem : “ predict $g = g(X, Y, \theta)$ given X ” (cf Yatracos, 1992).

In this paper, we suppose that g is \mathbb{R} -valued and denote $p = p(X)$ a statistical predictor. We suppose that $\mathbb{E}_\theta(g^2) < \infty$ and $\mathbb{E}_\theta(p^2) < \infty$, $\theta \in \Theta$. If p and q are two predictors, a classical preference relation is

$$p \prec q \text{ (} g \text{)} \iff \mathbb{E}_\theta(p - g)^2 \leq \mathbb{E}_\theta(q - g)^2, \theta \in \Theta$$

where “(g)” means “for predicting g ”.

Now, let $\mathbb{E}_\theta^X(g) = \mathbb{E}_\theta(g|X)$ be the conditional expectation of g given X . The next lemma is simple but important.

Lemma 2.1. *We have $p \prec q \text{ (} g \text{)} \iff p \prec q \text{ (} \mathbb{E}_\theta^X(g) \text{)}$.*

Proof. It suffices to apply the Pythagoras theorem :

$$\mathbb{E}_\theta(p(X) - g)^2 = \mathbb{E}_\theta(p(X) - \mathbb{E}_\theta^X(g))^2 + \mathbb{E}_\theta(\mathbb{E}_\theta^X(g) - g)^2$$

and

$$\mathbb{E}_\theta(q(X) - g)^2 = \mathbb{E}_\theta(q(X) - \mathbb{E}_\theta^X(g))^2 + \mathbb{E}_\theta(\mathbb{E}_\theta^X(g) - g)^2,$$

the result follows. \square

Prediction theory has some similarity but also some difference with estimation theory. In the following, we will only recall some necessary definitions and results. We refer to Bosq and Blanke (2007), chapters 1 and 2, for a more complete exposition.

3. BAYESIAN PREDICTION

3.1. Bayesian predictors. In a Bayesian context, one considers the random vector (X, Y, \mathbb{T}) . The prior distribution of \mathbb{T} over (Θ, \mathcal{T}) is defined by τ and the distribution \mathbb{P} of (X, Y, \mathbb{T}) is defined as

$$\mathbb{P}(X \in A, Y \in B, \mathbb{T} \in C) = \int_C \mathbb{P}_\theta(X \in A, Y \in B) d\tau(\theta), \quad A \in \mathcal{E}, B \in \mathcal{F}, C \in \mathcal{T}.$$

The Bayesian risk is given by

$$\mathbb{E}(p(X) - Y)^2 = \int_\Theta \mathbb{E}_\theta(p(X) - Y)^2 d\tau(\theta) := r(p, Y),$$

hence, the Bayesian predictor $p_0(X) = \underset{p}{\operatorname{argmin}} r(p, Y) = \mathbb{E}(Y|X)$ where $\mathbb{E}(\cdot|\cdot)$ is the conditional expectation associated with \mathbb{P} . We will say that p_0 is *unique* if two versions of p_0 differ only on \mathcal{N} with $\mathbb{P}_\theta(\mathcal{N}) = 0$, $\theta \in \Theta$. Now, we consider the following regularity assumption :

Assumption 3.1. *There exists a common version $m(X, \theta)$ of $\mathbb{E}_\theta(Y|X)$, $\theta \in \Theta$.*

Note that Assumption 3.1 is satisfied if (X, Y, \mathbb{T}) has a strictly positive density with respect to the σ -finite measure $\lambda \otimes \mu \otimes \nu$, say $L(x, y, \theta)\varphi(\theta)$, L denoting the conditional density of $(X, Y | \mathbb{T} = \theta)$. Then $p_0(X) = \int y f_{Y|X}(y|X) d\mu(y)$ where

$$f_{Y|X}(y|X) = \int L(x, y, \theta)\varphi(\theta) d\nu(\theta) / \iint L(x, y, \theta)\varphi(\theta) d\mu(y)d\nu(\theta).$$

In that situation p_0 is unique.

We now derive an alternative form of p_0 :

Lemma 3.1. *If Assumption 3.1 holds, then $m(X, \mathbb{T}) = \mathbb{E}(Y|X, \mathbb{T})$, and $p_0(X) = \mathbb{E}(m(X, \mathbb{T})|X)$.*

Proof. Let $h(X, \theta)$ be a real bounded function. Then, since $m(X, \theta)$ is a version of $\mathbb{E}_\theta(Y|X)$ for all θ , we have

$$\mathbb{E}_\theta\left((Y - m(X, \theta))h(X, \theta)\right) = 0, \theta \in \Theta,$$

hence

$$\mathbb{E}\left((Y - m(X, \mathbb{T}))h(X, \mathbb{T})\right) = \int \mathbb{E}_\theta\left((Y - m(X, \theta))h(X, \theta)\right) d\tau(\theta) = 0.$$

It follows that $m(X, \mathbb{T}) = \mathbb{E}(Y|X, \mathbb{T})$. Now, the Bayesian risk has the decomposition

$$r(Y, p(X)) = \mathbb{E}(Y - m(X, \mathbb{T}))^2 + r(m(X, \mathbb{T}), p(X)),$$

thus it is equivalent to minimize $r(Y, p(X))$ and $r(m(X, \mathbb{T}), p(X))$, consequently

$$p_0(X) = \mathbb{E}(m(X, \mathbb{T})|X).$$

□

Clearly, Lemma 3.1 holds if Y is replaced by $g(X, Y, \theta)$.

3.2. The MAP predictor. An alternative method of Bayesian prediction is based on the conditional mode : one may compute the mode of the distribution of Y , given X , with respect to \mathbb{P} . If a strictly positive density does exist, the distribution of (X, Y) has marginal density

$$f(x, y) = \int_{\Theta} L(x, y, \theta)\varphi(\theta) d\theta$$

and, in fact, it suffices to compute $\operatorname{argmax}_y f(x, y)$ (x fixed). Another method consists in determining the mode of \mathbb{T} given X and to plug it in the conditional expectation $\mathbb{E}_\theta(Y|X)$. This mode (also called *maximum a posteriori*, MAP) has the expression

$$\tilde{\theta}(x) = \operatorname{argmax}_{\theta} \frac{\ell(x, \theta)\varphi(\theta)}{\int_{\Theta} \ell(x, \theta)\varphi(\theta) d\tau(\theta)} = \operatorname{argmax}_{\theta} \ell(x, \theta)\varphi(\theta)$$

where $\ell(x, \theta) = \int L(x, y, \theta) d\mu(y)$, hence the MAP predictor

$$\tilde{p}(X) = \mathbb{E}_\theta(Y|X) \Big|_{\theta=\tilde{\theta}(X)} = m(X, \tilde{\theta})$$

under the Assumption 3.1. It is noteworthy that, if $\Theta = \mathbb{R}$ and one chooses the improper prior $1 \cdot \lambda$, where λ is Lebesgue measure, the obtained estimator is the maximum likelihood (MLE). Note also that, if $\ell(x, \theta)\varphi(\theta)$ is symmetric with respect to $\tilde{\theta}(X)$, the MAP and the Bayes estimator of θ coincide. Finally, it is clear that, under classical regularity conditions,

the MAP and the MLE have the same asymptotic behaviour as well almost surely as in distribution. Now, the MAP has some drawbacks : it is often difficult to compute and uniqueness is not guaranteed. We will use the MAP in Sections 5 to 7.

4. PROPERTIES OF BAYESIAN PREDICTORS

We give below some useful properties of Bayesian predictors.

4.1. Admissibility.

Proposition 4.1. *A Bayesian predictor is admissible as soon as it is unique.*

Proof. If p_0 is not admissible, there exists a predictor p such that

$$\mathbb{E}_\theta(p - g)^2 \leq \mathbb{E}_\theta(p_0 - g)^2, \theta \in \Theta$$

Integrating with respect to τ entails $r(p, g) \leq r(p_0, g)$, but, since p_0 is Bayesian, it follows that $r(p, g) = r(p_0, g)$ and uniqueness of p_0 gives $p = p_0$ (\mathbb{P}_θ a.s. for all θ). \square

4.2. g -Sufficiency. A statistic $S = S(X)$ is said to be g -sufficient (or sufficient for predicting g) if

- (a) S is sufficient in the statistical model associated with X : there exists a version of the conditional distribution of X given S , say Q^S , that does not depend on θ .
- (b) X and g are conditionally independent given S .

Note that this does not imply that $\mathbb{E}_\theta(Y|S(X))$ is constant with respect to θ since Y is not in the model associated with X (see example of the Poisson process in Section 6). If S is g -sufficient, it is then possible to derive a Rao-Blackwell theorem as well as a factorization theorem (cf Bosq and Blanke, 2007). Now, we have

Lemma 4.1. *If S is Y -sufficient and if Assumption 3.1 holds, then*

$$(4.1) \quad \mathbb{E}_\theta(Y|X) = \mathbb{E}_\theta(Y|S(X)), \theta \in \Theta.$$

Proof. Note first that $m(X, \theta)$, respectively $\int m(X, \theta) dQ^S$, is a version of $\mathbb{E}_\theta(Y|X)$, resp. of $\mathbb{E}_\theta(Y|S(X))$. For sake of clarity, we use notation $\mathbb{E}_\theta(Y|X)$ and $\mathbb{E}_\theta(Y|S(X))$. Now,

$$\mathbb{E}_\theta(Y|S(X)) = \mathbb{E}_\theta\left(\mathbb{E}_\theta(Y|X)|S(X)\right),$$

and applying (b) to Y and $\mathbb{E}_\theta(Y|X)$ we obtain

$$\mathbb{E}_\theta(Y \cdot \mathbb{E}_\theta(Y|X)|S(X)) = \mathbb{E}_\theta(Y|S(X)) \cdot \mathbb{E}_\theta\left(\mathbb{E}_\theta(Y|X)|S(X)\right) = \left(\mathbb{E}_\theta(Y|S(X))\right)^2.$$

Taking expectation and noting that

$$\mathbb{E}_\theta(Y \cdot \mathbb{E}_\theta(Y|X)) = \mathbb{E}_\theta\left(\left(\mathbb{E}_\theta(Y|X)\right)^2\right)$$

entails

$$\mathbb{E}_\theta\left(\left(\mathbb{E}_\theta(Y|X)\right)^2\right) = \mathbb{E}_\theta\left(\left(\mathbb{E}_\theta(Y|S(X))\right)^2\right),$$

that is $\|\mathbb{E}_\theta(Y|X)\|_{L^2(\mathbb{P}_\theta)}^2 = \|\mathbb{E}_\theta(Y|S(X))\|_{L^2(\mathbb{P}_\theta)}^2$. Relation (4.1) follows since $\mathbb{E}_\theta(Y|S(X))$ is the projection of $\mathbb{E}_\theta(Y|X)$ on $L^2_{S(X)}$. \square

Note that, if (X, Y) has a strictly positive density of the form $L(S(x), y, \theta)$, one obtains (4.1) by a direct computation. Concerning the Bayesian predictor, we have

Proposition 4.2. *If p_0 is unique and S is Y -sufficient, then $p_0(X) = \mathbb{E}(Y|S(X))$.*

Proof. Since $p_0(X) = E(Y|X)$, the Rao-Blackwell theorem for prediction (cf Bosq and Blanke, 2007, p. 15) entails $p_1(X) := \mathbb{E}^{S(X)}(\mathbb{E}(Y|X)) \prec p_0(X)$ where $\mathbb{E}^{S(X)}$ is conditional expectation with respect to Q^S in (a). Now, from Proposition 4.1, p_0 is admissible, thus $p_1 = p_0$, then p_0 has the form $p_0(X) = \mathbb{E}(Y|X) = \psi(S(X))$. Taking conditional expectation with respect to \mathbb{P} , one obtains the result. \square

4.3. Decomposition of the conditional expectation. We now consider the special case where the conditional expectation admits the following decomposition :

$$(4.2) \quad \mathbb{E}_\theta(Y|X) = A(X) + B(\theta)C(X) + D(\theta), \quad \theta \in \Theta$$

where $A, B \otimes C, D \in L^2(E \times \Theta, \mathcal{E} \otimes \mathcal{T}, \mathbb{P}_{(X, \mathbb{T})})$, then the Bayesian predictor has also a special form :

Proposition 4.3. *If $\mathbb{E}_\theta(Y|X)$ satisfies (4.2), the associated Bayesian predictor is given by*

$$(4.3) \quad p_0(X) = A(X) + \mathbb{E}(B(\mathbb{T})|X) \cdot C(X) + \mathbb{E}(D(\mathbb{T})|X).$$

In particular, if X and Y are independent and $D(\theta) = \mathbb{E}_\theta(Y)$, the predictor reduces to the estimator $p_0(X) = \mathbb{E}(D(\mathbb{T})|X)$.

Proof. Relation (4.2) entails

$$m(X, \mathbb{T}) = A(X) + B(\mathbb{T}) \cdot C(X) + D(\mathbb{T}),$$

and Lemma 3.1 gives $p_0(X) = \mathbb{E}(m(X, \mathbb{T})|X)$ hence (4.3) from the properties of conditional expectation. The last assertion is a special case of (4.3). \square

4.4. Unbiasedness. A predictor p of g is said to be unbiased if

$$\mathbb{E}_\theta p = \mathbb{E}_\theta g, \quad \theta \in \Theta.$$

A Bayesian estimator is, in general, not unbiased, in fact we have the following :

Lemma 4.2 (Blackwell-Girschick). *Let $\widehat{\varphi}(X)$ be an unbiased Bayesian estimator of $\varphi(\theta)$, then*

$$\mathbb{E}(\widehat{\varphi}(X) - \varphi(\mathbb{T}))^2 = 0$$

where \mathbb{E} is expectation taken from $\mathbb{P}_{(X, \mathbb{T})}$.

Proof. See Lehmann and Casella (1998) p. 234. \square

The situation is more intricate concerning a Bayesian predictor. Note first that, if

$$(4.4) \quad \mathbb{E}_\theta p_0 = \mathbb{E}_\theta g, \quad \theta \in \Theta$$

then, p_0 is an unbiased estimator of $\mathbb{E}_\theta g$ but it is not necessarily a Bayesian estimator of $\mathbb{E}_\theta g$. Recall that the Bayesian interpretation of (4.4) is :

$$\mathbb{E}(p_0|\mathbb{T} = \theta) = \mathbb{E}(g|\mathbb{T} = \theta), \quad \theta \in \Theta.$$

Now, we have the following result :

Proposition 4.4. *If the Bayesian risk satisfies*

$$(4.5) \quad \mathbb{E}(p_0(X) - m(X, \mathbb{T}))^2 = 0$$

then $p_0(X)$ is unbiased for predicting Y . Conversely if

$$(4.6) \quad m(X, \theta) = A(X) + D(\theta)$$

and if $p_0(X)$ is unbiased then (4.5) holds.

Proof. Relation (4.5) implies $p_0(X) = m(X, \mathbb{T})$ $\mathbb{P}_{(X, \mathbb{T})}$ a.s., that is

$$\mathbb{E}(Y|X) = \mathbb{E}(Y|X, \mathbb{T}) \quad \mathbb{P}_{(X, \mathbb{T})} \text{ a.s.}$$

Conditioning with respect to \mathbb{T} gives

$$\mathbb{E}(p_0(X)|\mathbb{T}) = \mathbb{E}(Y|\mathbb{T})$$

which means that $p_0(X)$ is unbiased. Conversely (4.6) and (4.3) in Proposition 4.3 imply

$$p_0(X) = A(X) + \mathbb{E}(D(\mathbb{T})|X).$$

Now, since p_0 is unbiased, and $\mathbb{E}(Y|X, \mathbb{T}) = m(X, \mathbb{T})$ implies

$$\mathbb{E}(Y|\mathbb{T}) = \mathbb{E}(m(X, \mathbb{T})|\mathbb{T}),$$

one gets

$$\mathbb{E}(A(X)|\mathbb{T}) + \mathbb{E}\left(\mathbb{E}(D(\mathbb{T})|X)|\mathbb{T}\right) = \mathbb{E}(Y|\mathbb{T})$$

and

$$\mathbb{E}(Y|\mathbb{T}) = \mathbb{E}(A(X)|\mathbb{T}) + \mathbb{E}(D(\mathbb{T})|\mathbb{T})$$

by (4.6). This means that the Bayesian estimator of $D(\theta)$ is also unbiased. Then Lemma 4.2 gives

$$\mathbb{E}(p_0(X) - m(X, \mathbb{T}))^2 = \mathbb{E}(\mathbb{E}(D(\mathbb{T})|X) - D(\mathbb{T}))^2 = 0.$$

□

In the more general case where $\mathbb{E}_\theta(Y|X)$ has the form (4.2) with non-null $B(\theta)C(X)$, it is possible to find an unbiased Bayesian predictor with a non-vanishing Bayesian risk (cf Bosq, 2012).

Now, let us define a ‘‘Bayesian type’’ predictor by

$$(4.7) \quad p_0(X) = \alpha p(X) + (1 - \alpha)m(X, \theta_0), \quad (0 < \alpha < 1, \theta_0 \in \Theta)$$

where $p(X)$ is an unbiased predictor of Y . For these specific predictors, our previous result may be extended as follows.

Proposition 4.5. *Suppose that Assumption 3.1 holds and consider a predictor $p_0(X)$ of the form (4.7). Then, if $p_0(X)$ is unbiased, it follows that*

$$(4.8) \quad \mathbb{E}_\theta(m(X, \theta)) = \mathbb{E}_\theta(m(X, \theta_0)), \quad \theta \in \Theta,$$

if, in addition, there exists a Y -sufficient complete statistic then $m(X, \theta) = m(X, \theta_0)$, $\theta \in \Theta$ and the problem of prediction is degenerated.

Proof. If $p_0(X)$ is an unbiased predictor of Y , one has

$$\mathbb{E}_\theta(p_0(X)) = \mathbb{E}_\theta(m(X, \theta)), \quad \theta \in \Theta,$$

and taking expectation in (4.7) yields

$$\mathbb{E}_\theta(m(X, \theta)) = \alpha \mathbb{E}_\theta(p(X)) + (1 - \alpha) \mathbb{E}_\theta(m(X, \theta_0))$$

hence, since $p(X)$ is unbiased, (4.8) follows. Now, if $S(X)$ is a Y -sufficient statistic, Lemma 4.1 entails $m(X, \theta) = \mathbb{E}_\theta(Y|S(X))$, thus, (4.8) implies

$$\mathbb{E}_\theta\left(\mathbb{E}_\theta(Y|S(X)) - \mathbb{E}_{\theta_0}(Y|S(X))\right) = 0, \quad \theta \in \Theta,$$

and, since $S(X)$ is complete, one obtains the last result. \square

4.5. Comparing predictors. The following elementary lemma allows to compare Bayesian predictors with the classical unbiased predictor. We will use it in the next sections.

Lemma 4.3. *Suppose that*

$$m(X, \theta) = A(X) + d \cdot \theta \quad (d \neq 0)$$

and let $p(X)$ be an unbiased predictor of Y taking the form

$$p(X) = A(X) + d \cdot \bar{\theta}(X).$$

Consider the ‘‘Bayesian type’’ predictor

$$p_0(X) = \alpha p(X) + (1 - \alpha)m(X, \theta_0)$$

where $\alpha \in]0, 1[$ and $\theta_0 \in \Theta$. Then

$$(4.9) \quad p_0 \prec p \iff |\theta - \theta_0| \leq \left(\frac{1 + \alpha}{1 - \alpha}\right)^{\frac{1}{2}} \cdot \left(\mathbb{E}_\theta(\bar{\theta}(X) - \theta)^2\right)^{\frac{1}{2}}.$$

Proof. We have

$$p_0(X) - m(X, \theta) = \alpha(p(X) - m(X, \theta)) + (1 - \alpha)(m(X, \theta_0) - m(X, \theta))$$

then, since p is unbiased,

$$\mathbb{E}_\theta(p_0(X) - m(X, \theta))^2 = \alpha^2 \mathbb{E}_\theta(p(X) - m(X, \theta))^2 + (1 - \alpha)^2 d^2 (\theta_0 - \theta)^2$$

thus

$$p_0 \prec p \iff d^2(1 - \alpha)^2(\theta - \theta_0)^2 + \alpha^2 \mathbb{E}_\theta(p(X) - m(X, \theta))^2 \leq \mathbb{E}_\theta(p(X) - m(X, \theta))^2$$

and (4.9) follows. \square

Remarks : If $X = X_{(n)} = (X_1, \dots, X_n)$ and

$$\mathbb{E}_\theta(\bar{\theta}(X) - \theta)^2 = \frac{v}{n}$$

then the condition is

$$|\theta - \theta_0| \leq \left(\frac{1 + \alpha}{1 - \alpha}\right)^{\frac{1}{2}} \cdot \left(\frac{v}{n}\right)^{\frac{1}{2}}.$$

If one may find $\alpha = \alpha_n$ such that

$$\inf_{n \geq 1} \left(\frac{1 + \alpha_n}{1 - \alpha_n}\right)^{\frac{1}{2}} \cdot \left(\frac{v}{n}\right)^{\frac{1}{2}} \geq b > 0,$$

it follows that $|\theta - \theta_0| \leq b$ implies $p_0(X_{(n)}) \prec p(X_{(n)})$ for all $n \geq 1$. Moreover, the choice $A(X) \equiv 0$ in Lemma 4.3 provides an alternative formulation for comparing Bayesian estimators of θ versus non Bayesian ones.

5. APPLICATION TO POISSON PROCESS

5.1. The Bayesian predictor. Let $(N_t, t \geq 0)$ be an homogeneous Poisson process with intensity $\theta > 0$, $X = (N_t, 0 \leq t \leq S)$ is observed and one wants to predict $Y = N_{S+h}$ ($h > 0$), ($S > 0$). Lemma 2.1 shows that it is equivalent to predict $m(X, \theta) = \theta h + N_S$. The unbiased efficient predictor is obtained by replacing θ with $\frac{N_S}{S}$ (Bosq and Blanke, 2007) for obtaining $p(N_S) = \frac{S+h}{S} N_S =: N_S + \theta_S h$.

Concerning the Bayesian predictor, a classical prior is $\tau = \Gamma(a, b)$ with density

$$\frac{b^a}{\Gamma(a)} \theta^{a-1} \exp(-b\theta) \mathbb{1}_{]0, +\infty[}(\theta), \quad (a > 0, b > 0).$$

Since N_S is N_{S+h} -sufficient, Lemma 4.1 entails

$$\mathbb{E}_\theta(N_{S+h}|N_t, 0 \leq t \leq S) = \mathbb{E}_\theta(N_{S+h}|N_S)$$

and Proposition 4.2 gives

$$p_0(N_t, 0 \leq t \leq S) = \mathbb{E}(N_{S+h}|N_S).$$

The same property holds for the Bayes estimator given by

$$\hat{\theta}_S = \mathbb{E}(\mathbb{T}|N_S) = \frac{a + N_S}{b + S},$$

and, from Proposition 4.3, the Bayesian predictor is

$$\hat{p}_0(N_S) = \frac{a + N_S}{b + S} \cdot h + N_S.$$

To compare \hat{p}_0 with p , note that

$$\hat{\theta}_S = \frac{S}{b + S} \cdot \theta_S + \left(1 - \frac{S}{b + S}\right) \cdot \frac{a}{b}.$$

We deduce that

$$\hat{p}_0(N_S) = \alpha_S p(N_S) + (1 - \alpha_S)(N_S + \theta_0 h)$$

with $\alpha_S = \frac{S}{b+S}$ and $\theta_0 = \frac{a}{b}$. Since $\mathbb{E}_\theta(\theta_S - \theta)^2 = \frac{\theta}{S}$, a straightforward consequence of Lemma 4.3 is

$$(5.1) \quad \hat{p}_0 \prec p \iff (\theta - \theta_0)^2 \leq \left(\frac{1}{S} + \frac{2}{b}\right)\theta.$$

Solving (5.1) in θ , we get that $p_0 \prec p$ iff

$$\theta^2 - 2\theta\left(\theta_0 + \frac{1}{2S} + \frac{1}{b}\right) + \theta_0^2 < 0$$

that is when

$$\theta \in \left] \theta_0 + \frac{1}{2S} + \frac{1}{b} - \sqrt{\Delta}, \theta_0 + \frac{1}{2S} + \frac{1}{b} + \sqrt{\Delta} \right[$$

with $\Delta = \left(\theta_0 + \frac{1}{2S} + \frac{1}{b}\right)^2 - \theta_0^2$.

Also, from (5.1), a sufficient condition, holding for all S , is $(\theta - \theta_0)^2 \leq \frac{2}{b}\theta$ which gives

$$\theta \in \left] \theta_0 + \frac{1}{b} - \sqrt{\widetilde{\Delta}}, \theta_0 + \frac{1}{b} + \sqrt{\widetilde{\Delta}} \right[$$

with $\widetilde{\Delta} = \frac{1}{b}(2\theta_0 + \frac{1}{b})$, that is

$$\theta \in \left] \theta_1, \theta_2 \right[:= \left] \frac{a+1}{b} - \frac{\sqrt{2a+1}}{b}, \frac{a+1}{b} + \frac{\sqrt{2a+1}}{b} \right[.$$

Clearly, one obtains the same result for comparing $\widehat{\theta}_S$ with θ_S . If $a = 1$, $\theta_0 = \frac{1}{b}$ then $\theta_1 = \frac{2-\sqrt{3}}{b}$ and $\theta_2 = \frac{2+\sqrt{3}}{b}$. If b is large, $\theta_2 - \theta_1$ is large but θ also !

Turning to the MAP estimator, one has to compute $\operatorname{argmax}_{\theta} L(\theta)$ which is equal to

$$\operatorname{argmax}_{\theta} e^{-\theta S} \frac{(\theta S)^{N_S}}{N_S!} \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-\theta b}.$$

We have

$$\frac{\partial \ln L(\theta)}{\partial \theta} = \frac{\partial}{\partial \theta} \left(-\theta(S+b) + \frac{N_S + a - 1}{\theta} \right)$$

hence $\widetilde{\theta}_S = \frac{N_S + a - 1}{b + S}$ where we choose $a \geq 1$ for convenience, inducing the predictor :

$$\widetilde{p}_0(N_S) = \frac{N_S + a - 1}{b + S} h + N_S.$$

Replacing a with $a - 1$, the previous discussion about \widehat{p}_0 holds and one gets, for all S , the sufficient condition

$$\widetilde{p}_0 \prec p \iff \frac{a}{b} - \frac{\sqrt{2a-1}}{b} < \theta < \frac{a}{b} + \frac{\sqrt{2a-1}}{b}.$$

Finally, another method consists in computing the marginal distribution of (N_S, N_{S+h}) and then to determine the conditional mode of N_{S+h} given N_S . In this case, one obtains a similar predictor. Details are left to the reader.

5.2. Simulations. In this section, we compare the unbiased (UP), the Bayesian (BP) and the MAP predictors for various Poisson processes. First, we simulate $N = 10^5$ homogeneous Poisson processes with intensity θ varying in $\{0.5, 1, 2, 5, 10\}$. Next, for S in $\{10, 15, 20, 25, 30, 40, 50, 75, 100\}$ and horizon of prediction h in $\{0.5, 1, 2, 5\}$, we compute an approximation of the empirical L^2 -error of prediction :

$$\frac{1}{N} \sum_{j=1}^N (N_{S+h}^{(j)} - \widehat{p}(N_S^{(j)}))^2$$

where $N_t^{(j)}$ stands for the j -th replicate of the process at time t and $\widehat{p}(N_S^{(j)})$ is the predictor under consideration (Bayesian and MAP predictors are computed with a $\Gamma(a, 1)$ distribution

for the prior). We will also consider the empirical L^2 -error of estimation (with respect to the probabilistic predictor $\mathbb{E}_\theta(N_{S+h}|N_S)$) defined by

$$\frac{1}{N} \sum_{j=1}^N (N_S^{(j)} + \theta h - \widehat{p}(N_S^{(j)}))^2.$$

In Table 1, we give the rounded L^2 -errors of estimation according to S as well as prediction errors (enclosed in parentheses) for the unbiased predictor when $\theta = h = 1$. To help the comparison, only the percentage variations of BP and MAP errors (relatively to the UP ones) are reported for $a = 1, 2, 4$. Namely, since $\theta = 1$, it is expected from (5.1) that $a = 4$ represents a bad choice of prior (while $a = 1$ corresponds to the best one, and $a = 2$ is acceptable). From Table 1, we observe that :

- as expected, all errors decrease as S increases ;
- for all errors and any value of S , Bayesian and MAP predictors are better than the unbiased one for $a = 1, 2$, with a clearly significant gain for small values of S in the estimation framework ;
- the bad choice $a = 4$ clearly penalizes the predictor, with a significant impact on the L^2 -error of estimation. Concerning the prediction error, this effect is smaller : unfortunately, the probabilistic error is predominant !

TABLE 1. L^2 estimation (prediction) error for UP and percentage variation of L^2 estimation (prediction) error for BP and MAP, in the case where $\theta = 1$ and $h = 1$.

	S=15			S=20			S=30		
UP	0.066 (1.066)			0.050 (1.050)			0.033 (1.036)		
	a=1	a=2	a=4	a=1	a=2	a=4	a=1	a=2	a=4
BP %	-12.1(-.74)	-6.3(-.42)	40.8(2.42)	-9.3(-.43)	-4.8(-.23)	31.5(1.45)	-6.3(-.21)	-3.2(-.11)	21.8(.69)
MAP %	-6.1(-.33)	-12.1(-.74)	11.3(.64)	-4.7(-.19)	-9.3(-.43)	8.8(.39)	-3.2(-.10)	-6.3(-.21)	6.2(.19)
	S=40			S=50			S=100		
UP	0.025 (1.027)			0.020 (1.025)			0.010 (1.015)		
	a=1	a=2	a=4	a=1	a=2	a=4	a=1	a=2	a=4
BP %	-4.8(-.12)	-2.4(-.05)	16.8(.42)	-3.9(-.08)	-1.9(-.03)	13.7(.28)	-2.0(-.02)	-1.0(-.01)	6.7(.06)
MAP %	-2.5(-.06)	-4.8(-.12)	4.8(.13)	-2.0(-.04)	-3.9(-.08)	4.0(.09)	-0.9(-.01)	-2.0(-.02)	1.9(.02)

In Figure 1, the L^2 -error of prediction is plotted as a function of a for $\theta = 1$ and $S = 20$. As expected by (5.1), parabolic curves are obtained and BP (resp. MAP) is better than UP for a in the interval $\left] 0, 1 + \sqrt{S^{-1} + 2} \right[$ (resp. $\left] 2 - \sqrt{S^{-1} + 2}, 2 + \sqrt{S^{-1} + 2} \right[$). Same conclusions hold for other choices of h and/or θ (see the selected results in Table 2). Errors increase as h and/or θ increase, and a good choice of the prior has a significative impact on the estimation error. In conclusion, at least for this simulated case, it appears that the L^2 -error of prediction can be quite large, especially for small values of S and/or large values of h and θ .

6. BAYESIAN INFERENCE FOR THE ORNSTEIN-UHLENBECK PROCESS

Consider a stationary version of the Ornstein-Uhlenbeck process (O.U.) defined by $X_t = m + \int_{-\infty}^t e^{-\theta(t-s)} dW(s)$, $t \in \mathbb{R}$, ($m \in \mathbb{R}$, $\theta > 0$) where W is a standard bilateral Wiener

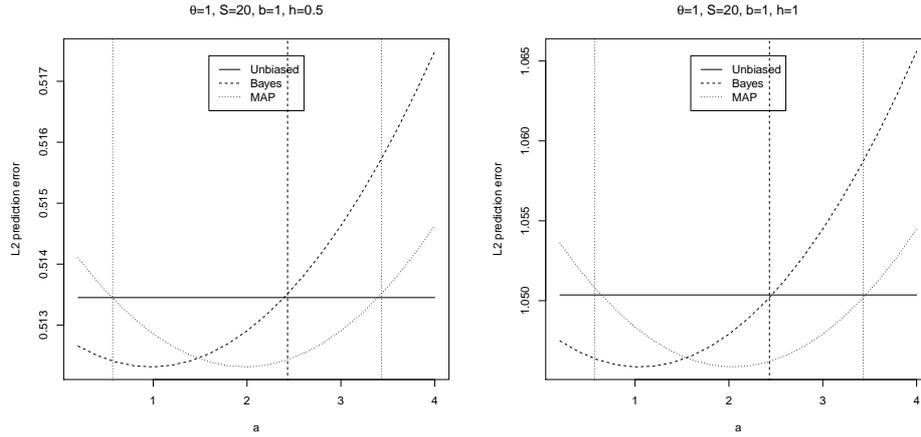


FIGURE 1. L^2 prediction error for $\theta = 1$ in terms of a with $\Gamma(a, 1)$ prior : UP (plain horizontal), BP (dashes), MAP (dots) for $S = 20$. Vertical lines corresponds to $a = 1 + \sqrt{S-1+2}$ (dashes) and $a = 2 \pm \sqrt{S-1+2}$ (dots). On the left : $h = 0.5$, on the right : $h = 1$.

TABLE 2. L^2 estimation (prediction) error, in the case $S = 20$, for UP and percentage variations of L^2 estimation (prediction) error for BP_i and MAP_i , where i refers to $a = i$.

	$\theta=0.5$			$\theta=5$			$\theta=10$		
	$h=0.5$	$h=1$	$h=2$	$h=0.5$	$h=1$	$h=2$	$h=0.5$	$h=1$	$h=2$
UP	.01 (.3)	.02 (.5)	.1 (1.1)	.06 (2.6)	.25 (5.3)	1 (11)	.12 (5.1)	.5 (10.4)	1.99 (22.1)
BP1 %	-6.9 (-.16)	-6.9 (-.33)	-6.9 (-.61)	5.1 (.2)	5.1 (.34)	5.1 (.61)	28 (.75)	28 (1.48)	28 (2.7)
BP2 %	11.5 (.31)	11.5 (.55)	11.5 (1.1)	-1.2 (.02)	-1.2 (.02)	-1.2 (.0)	20.2 (.55)	20.2 (1.09)	20.2 (1.98)
BP4 %	103.3 (2.58)	103.3 (4.89)	103.3 (9.52)	-8.4 (-.19)	-8.4 (-.37)	-8.4 (-.73)	7.4 (.23)	7.4 (.45)	7.4 (.78)
MAP1 %	-7.1 (-.19)	-7.1 (-.34)	-7.1 (-.66)	13.2 (.41)	13.2 (.75)	13.2 (1.39)	36.7 (.97)	36.7 (1.91)	36.7 (3.5)
MAP2 %	-6.9 (-.16)	-6.9 (-.33)	-6.9 (-.61)	5.1 (.2)	5.1 (.34)	5.1 (.61)	28 (.75)	28 (1.48)	28 (2.7)
MAP4 %	48.3 (1.22)	48.3 (2.29)	48.3 (4.48)	-5.7 (-.1)	-5.7 (-.22)	-5.7 (-.44)	13.3 (.38)	13.3 (.75)	13.3 (1.34)

process. Set $X_{0,t} = X_t - m$, $t \in \mathbb{R}$, then the likelihood of $X_{(S)} = (X_t, 0 \leq t \leq S)$ with respect to $X_{0,(S)} = (X_{0,t}, 0 \leq t \leq S)$ is given by

$$(6.1) \quad L(X_{(S)}; m, \theta) = \exp \left(-\frac{\theta m^2}{2}(2 + \theta S) + \theta m(X_0 + X_S + \theta \int_0^S X_t dt) \right)$$

(cf Grenander, 1981, p. 128-129) where $X_{(S)}$ and $X_{0,(S)}$ take their values in the space $C([0, S])$, ($S > 0$).

6.1. **Estimating m .** We suppose that θ is known and $m \in \mathbb{R}$ is unknown. In order to construct a Bayesian estimator of m and a Bayesian predictor of X_{S+h} ($h > 0$) given $X_{(S)}$, we consider the random variable \mathbb{M} with prior distribution $\mathcal{N}(m_0, u^2)$ ($u > 0$), and suppose that \mathbb{M} is independent from W . Using (6.1), it follows that the posterior density of \mathbb{M} given $X_{(S)}$ is $\mathcal{N}(\frac{B}{A}, \frac{1}{A})$ where $A = \theta(2 + \theta S) + \frac{1}{u^2}$ and $B = \theta Z_S + \frac{m_0}{u^2}$ with $Z_S = (X_0 + X_S + \theta \int_0^S X_t dt)$.

Hence the Bayesian estimator of m :

$$\widehat{m}_S = \frac{B}{A} = \frac{Z_S + m_0 \theta^{-1} u^{-2}}{2 + \theta S + \theta^{-1} u^{-2}}$$

when the maximum likelihood estimator (MLE) is $m_S = \frac{Z_S}{2 + \theta S}$. Consequently

$$(6.2) \quad \widehat{m}_S = \alpha_S m_S + (1 - \alpha_S) m_0$$

with $\alpha_S = (1 + \theta^{-1}(2 + \theta S)^{-1} u^{-2})^{-1} \in]0, 1[$. Note that $\lim_{u \rightarrow 0} \widehat{m}_S = m_0$ and $\lim_{u \rightarrow \infty} \widehat{m}_S = m_S$.

Asymptotic efficiency. The MLE m_S is efficient (cf Bosq and Blanke, 2007, p. 28) and \widehat{m}_S is asymptotically efficient since, from (6.2),

$$\frac{\mathbb{E}_m(\widehat{m}_S - m)^2}{\mathbb{E}_m(m_S - m)^2} = \alpha_S^2 + (1 - \alpha_S)^2 \frac{(m_0 - m)^2}{\mathbb{E}_m(m_S - m)^2}$$

with $\alpha_S^2 \rightarrow 1$ as $S \rightarrow \infty$, $(1 - \alpha_S)^2 = \mathcal{O}(S^{-2})$ and $\mathbb{E}_m(m_S - m)^2 = \mathcal{O}(S^{-1})$.

Prediction. We have $\mathbb{E}_m(X_{S+h} | X_{(S)}) = \mathbb{E}_m(X_{S+h} | X_S) = e^{-\theta h}(X_S - m) + m$. The unbiased predictor associated with the MLE is

$$p_S := p(X_{(S)}) = m_S(1 - e^{-\theta h}) + e^{-\theta h} X_S,$$

and using Proposition 4.2, one obtains the Bayesian predictor

$$\widehat{p}_{0,S} := p_0(X_{(S)}) = \widehat{m}_S(1 - e^{-\theta h}) + e^{-\theta h} X_S.$$

We get

$$\widehat{p}_{0,S} = \alpha_S p_S + (1 - \alpha_S)(m_0(1 - e^{-\theta h}) + e^{-\theta h} X_S) = \alpha_S p_S + (1 - \alpha_S)p(X_S, m_0).$$

Concerning efficiency, again we deduce that p_S is efficient and $\widehat{p}_{0,S}$ is asymptotically efficient. Now, in order to compare $\widehat{p}_{0,S}$ with p_S , we use Lemma 4.3 for obtaining the following result.

Proposition 6.1. *We have that $\widehat{p}_{0,S} \prec p_S$ is equivalent to $|m - m_0| \leq \left(\frac{1}{\theta(2 + \theta S)} + 2u^2\right)^{\frac{1}{2}}$ and $|m - m_0| \leq u\sqrt{2}$ implies $\widehat{p}_{0,S} \prec p_S$ for all $S > 0$.*

The proof is straightforward since one has $\mathbb{E}_m(m_S - m)^2 = (\theta(2 + \theta S))^{-1}$. Of course, the result is strictly the same if one compares \widehat{m}_S with m_S since $\widehat{m}_S \prec m_S$ is equivalent to $\widehat{p}_{0,S} \prec p_S$.

6.2. Estimating θ . Suppose now that θ is unknown and m is known ; one may take $m = 0$. The likelihood of $X_{(S)}$ with respect to $W_{(S)}$ has the form

$$L(X_{(S)}) = \exp\left(-\frac{1}{2}(X_S^2 - X_0^2 - S) - \frac{\theta^2}{2} \int_0^S X_t^2 dt\right),$$

see Liptser and Shiryaev (2001). Even if θ is positive, it is convenient to take $\mathcal{N}(\theta_0, v^2)$ (with $\theta_0 > 0$ and $v^2 > 0$) as prior distribution of \mathbb{T} . Then, the marginal distribution of $X_{(S)}$ has density $\varphi(x_{(S)}) = \frac{1}{\sqrt{\alpha v^2}} \exp\left(-\frac{\theta_0^2}{2v^2} + \frac{\beta^2}{2\alpha}\right)$ where $\alpha = \int_0^S x_s^2 ds + \frac{1}{v^2}$ and $\beta = \frac{S - x_S^2 + x_0^2}{2} + \frac{\theta_0}{v^2}$.

It follows that the conditional distribution of \mathbb{T} given $X_{(S)}$ is $\mathcal{N}(\frac{\beta}{\alpha}, \frac{1}{\alpha})$, hence the bayesian estimator of θ : $\hat{\theta}_S = \frac{\beta}{\alpha} = \frac{\frac{1}{2}(S-X_S^2+X_0^2)+\theta_0 v^{-2}}{\int_0^S X_t^2 dt + v^{-2}}$ when the MLE is $\theta_S = \frac{\frac{1}{2}(S-X_S^2+X_0^2)}{\int_0^S X_t^2 dt}$, consequently

$$(6.3) \quad \hat{\theta}_S = \gamma_S \theta_S + (1 - \gamma_S) \theta_0 \quad \text{with} \quad \gamma_S = \frac{\int_0^S X_t^2 dt}{\int_0^S X_t^2 dt + v^{-2}},$$

and $\lim_{v^2 \rightarrow 0} \hat{\theta}_S = \theta_0$ while $\lim_{v^2 \rightarrow \infty} \hat{\theta}_S = \theta_S$.

Concerning prediction, we have $\mathbb{E}_\theta(X_{S+h}|X_{(S)}) = e^{-\theta h} \cdot X_S$, so it is necessary to compute the Bayesian estimator of $e^{-\theta h}$. We get

$$\mathbb{E}(e^{-\mathbb{T}h}|X_{(S)}) = \int_{\mathbb{R}} e^{-\theta h} \sqrt{\frac{\alpha}{2\pi}} e^{-\frac{\alpha}{2}(\theta - \frac{\beta}{\alpha})^2} d\theta = \exp(-\frac{2\beta - h}{2\alpha} \cdot h),$$

hence the Bayesian predictor $\hat{p}_0(X_{(S)}) = \exp(-\frac{2\beta - h}{2\alpha} \cdot h) \cdot X_S$. The predictor associated with the MLE is $p(X_{(S)}) = e^{-\theta_S \cdot h} \cdot X_S$ and finally, an alternative form of the predictor, associated with the MAP, should be $\tilde{p}(X_{(S)}) = e^{-\hat{\theta}_S \cdot h} \cdot X_S$.

Remark : In order to compare $\hat{\theta}_S$ with θ_S , one may use an *approximation* of (6.3) by setting $\gamma_{\text{app},S} = (1 + \frac{2\theta}{Sv^2})^{-1}$. Then, the equivalent form of Lemma 4.3 for estimation, gives the ‘‘approximate’’ condition

$$\hat{\theta}_S < \theta_S \iff |\theta - \theta_0| \leq (1 + \frac{Sv^2}{\theta})^{\frac{1}{2}} (\mathbb{E}_\theta(\theta_S - \theta)^2)^{\frac{1}{2}},$$

note, however that θ_S is *not* unbiased.

Finally, one may consider alternative priors, as well as, the translated exponential distribution with density $\varphi(\theta) = \eta \exp(-\eta(\theta - \theta_0)) \mathbf{1}_{] \theta_0, +\infty[}(\theta)$, ($\eta > 0$, $\theta_0 \geq 0$). If ψ denotes the density of $\mathcal{N}(-\frac{a}{2b}, \frac{1}{b})$, with $a = x_S^2 - x_0^2 - S + 2\eta$ and $b = \int_0^S x_t^2 dt$, the Bayesian estimator is implicit and given by $\hat{\theta}_S = \int_{\theta_0}^{\infty} \theta \psi(\theta) d\theta / \int_{\theta_0}^{\infty} \psi(\theta) d\theta$. Derivation is left to the reader.

7. ORNSTEIN-UHLENBECK PROCESS FOR SAMPLED DATA

We now consider the more realistic case where only $X_0, X_\delta, \dots, X_{n\delta}$ are observed and one wants to predict $X_{(n+h)\delta}$, ($h > 0$).

7.1. Estimation of m . If θ is known, and $m \in \mathbb{R}$ unknown, the associated model is

$$(7.1) \quad X_{n\delta} - m = e^{-\theta\delta} (X_{(n-1)\delta} - m) + \varepsilon_{n\delta}, \quad n \in \mathbb{Z}$$

and

$$(7.2) \quad \text{Var}(\varepsilon_{n\delta}) = \frac{1 - e^{-2\theta\delta}}{2\theta} =: \sigma_{\delta,\theta}^2$$

If $\delta > 0$ is fixed, we deal with a classical AR(1), so we will focus on the case where $\delta = \delta_n$ is ‘‘small’’. One may use various condition as $n \rightarrow \infty$: $\delta_n \rightarrow 0$ and $n\delta_n \rightarrow \infty$ or $\delta_n \rightarrow 0$ and $n\delta_n \rightarrow S > 0$ for example. Two approaches are possible : either considering the likelihood or the conditional likelihood (X_0 is arbitrary but non random) which has a simpler form.

7.1.1. *Unconditional estimation.* Since $X_0 - m, \varepsilon_{\delta_n}, \dots, \varepsilon_{n\delta_n} \sim \mathcal{N}(0, (2\theta)^{-1}) \otimes \mathcal{N}(0, \sigma_{\delta, \theta}^2)^{\otimes n}$, one may deduce that $(X_0 - m, X_{\delta} - m, \dots, X_{n\delta} - m)$ has the density

$$f(x_0, x_1, \dots, x_n) = \left(\frac{\theta}{\pi}\right)^{\frac{1}{2}} \frac{1}{(\sigma_{\delta, \theta} \sqrt{2\pi})^n} \times \exp\left(-\theta(x_0 - m)^2 - \sum_{i=1}^n \frac{(x_i - e^{-\theta\delta_n} x_{i-1} - m(1 - e^{-\theta\delta_n}))^2}{2\sigma_{\delta, \theta}^2}\right).$$

This yields

$$(7.3) \quad m_n = \frac{X_0 + X_{n\delta_n} + (1 - e^{-\theta\delta_n}) \sum_{i=1}^{n-1} X_{i\delta_n}}{n(1 - e^{-\theta\delta_n}) + 1 + e^{-\theta\delta_n}}$$

for the MLE, while if $\mathbb{M} \sim \mathcal{N}(m_0, u^2)$, one has

$$L(X_0, X_{\delta_n}, \dots, X_{n\delta_n}, \mathbb{M}) = \left(\frac{\theta}{\pi}\right)^{\frac{1}{2}} \frac{1}{\sigma_{\delta, \theta} \sqrt{2\pi}} \times \exp\left(-\theta(X_0 - \mathbb{M})^2 - \sum_{i=1}^n \frac{(X_{i\delta_n} - \rho_{\delta, \theta} X_{(i-1)\delta_n} - \mathbb{M}(1 - e^{-\theta\delta_n}))^2}{2\sigma_{\delta, \theta}^2}\right) \times \frac{1}{u\sqrt{2\pi}} \exp\left(-\frac{1}{2u^2}(\mathbb{M} - m_0)^2\right)$$

giving

$$(7.4) \quad \widehat{m}_n = \frac{X_0 + X_{n\delta_n} + (1 - e^{-\theta\delta_n}) \sum_{i=1}^{n-1} X_{i\delta_n} + (1 + e^{-\theta\delta_n}) \frac{m_0}{2\theta u^2}}{n(1 - e^{-\theta\delta_n}) + (1 + e^{-\theta\delta_n}) \left(1 + \frac{1}{2\theta u^2}\right)}.$$

Again, we have $\widehat{m}_n = \alpha_n m_n + (1 - \alpha_n) m_0$ with

$$\alpha_n = \frac{n(1 - e^{-\theta\delta_n}) + 1 + e^{-\theta\delta_n}}{n(1 - e^{-\theta\delta_n}) + (1 + e^{-\theta\delta_n})(1 + (2\theta u^2)^{-1})}.$$

Since $\mathbb{E}(X_{(n+h)\delta_n} | X_{n\delta_n}) = e^{-\theta h \delta_n} (X_{n\delta_n} - m) + m$, the derived predictors of $X_{(n+h)\delta_n}$, $h \geq 1$ are given by $p_n(X_{n\delta_n}) = m_n(1 - e^{-\theta h \delta_n}) + e^{-\theta h \delta_n} X_{n\delta_n}$ while $\widehat{p}_{0,n}(X_{n\delta_n}) = \widehat{m}_n(1 - e^{-\theta h \delta_n}) + e^{-\theta h \delta_n} X_{n\delta_n}$, and Lemma 4.3 implies that

$$\widehat{p}_{0,n} \prec p_n \iff$$

$$(m - m_0)^2 \leq \frac{2n(1 - e^{-\theta\delta_n}) + (1 + e^{-\theta\delta_n})(2 + (2\theta u^2)^{-1})}{(1 + e^{-\theta\delta_n})(2\theta u^2)^{-1}} \times \mathbb{E}_m(m_n - m)^2.$$

Next, easy but tedious computation gives $\mathbb{E}_m(m_n - m)^2 = \frac{1 + e^{-\theta\delta_n}}{2\theta(n(1 - e^{-\theta\delta_n}) + 1 + e^{-\theta\delta_n})}$ yielding the

equivalence : $\widehat{p}_{0,n} \prec p_n \iff (m - m_0)^2 \leq 2u^2 + \frac{1 + e^{-\theta\delta_n}}{2\theta(1 + e^{-\theta\delta_n} + n(1 - e^{-\theta\delta_n}))}$. Asymptotically, we get

if $\delta_n \xrightarrow[n \rightarrow \infty]{} 0$, $n\delta_n \xrightarrow[n \rightarrow \infty]{} S > 0$, $\widehat{p}_{0,n} \prec_{n \rightarrow \infty} p_n$ is equivalent to $(m - m_0)^2 \leq 2u^2 + \frac{1}{\theta(S+2)}$.

The condition $S \rightarrow \infty$ implying in turn the equivalence $\widehat{p}_{0,n} \prec_{n \rightarrow \infty} p_n \iff (m - m_0)^2 \leq 2u^2$, which are the same results as in the continuous case (cf Proposition 6.1). If $n\delta_n \rightarrow S > 0$, note that our estimators of m are no more consistent ! But still in this case, a good choice of the prior should allow reductions of risks of estimation and prediction.

7.1.2. *Conditional likelihood.* In this part, we use conditional likelihood on X_0 , and choosing $\mathbb{M} \sim \mathcal{N}(m_0, u^2)$, ($u > 0$), we obtain the “density” of $(X_{\delta_n}, \dots, X_{n\delta_n}, \mathbb{M})$:

$$\begin{aligned} \tilde{L}(X_{\delta_n}, \dots, X_{n\delta_n}, \mathbb{M}) &= \frac{1}{(\sigma_{\delta, \theta} \sqrt{2\pi})^n} \exp \left(-\frac{1}{2\sigma_{\delta, \theta}^2} \sum_{i=1}^n \left((X_{i\delta_n} - \rho_\delta X_{(i-1)\delta_n}) \right. \right. \\ &\quad \left. \left. + \mathbb{M}(\rho_\delta - 1) \right)^2 \right) \times \frac{1}{u\sqrt{2\pi}} \exp \left(-\frac{1}{2u^2} (\mathbb{M} - m_0)^2 \right), \end{aligned}$$

where $\rho_\delta := \rho(\theta, \delta_n) = \exp(-\theta\delta_n)$ and $\sigma_{\delta, \theta}^2$ is defined in (7.2). Since we are in the Gaussian case, the conditional mode and the conditional expectation coincide, now :

$$\ln \tilde{L} = c - \frac{1}{2\sigma_{\delta, \theta}^2} \sum_{i=1}^n \left(X_{i\delta_n} - \rho_\delta X_{(i-1)\delta_n} + \mathbb{M}(\rho_\delta - 1) \right)^2 - \frac{(\mathbb{M} - m_0)^2}{2u^2},$$

where c does not depend on n . It follows that the Bayesian estimator is now given by

$$(7.5) \quad \tilde{m}_n = \frac{(1 - \rho_\delta) \sum_{i=1}^n (X_{i\delta_n} - \rho_\delta X_{(i-1)\delta_n}) + m_0 \frac{\sigma_{\delta, \theta}^2}{u^2}}{(1 - \rho_\delta)^2 n + \frac{\sigma_{\delta, \theta}^2}{u^2}},$$

while the conditional MLE takes the form

$$(7.6) \quad \check{m}_n = \frac{\sum_{i=1}^n (X_{i\delta_n} - \rho_\delta X_{(i-1)\delta_n})}{(1 - \rho_\delta)n},$$

We may slightly modify the estimator (7.5) for obtaining

$$(7.7) \quad \bar{m}_n = \beta_n \bar{X}_n + (1 - \beta_n) m_0$$

with $\bar{X}_n = n^{-1} \sum_{i=1}^n X_{i\delta_n}$ and $\beta_n = \frac{(1 - \rho_\delta)^2}{(1 - \rho_\delta)^2 + \frac{\sigma_{\delta, \theta}^2}{nu^2}}$. Hence

$$\frac{1 + \beta_n}{1 - \beta_n} = \frac{4\theta u^2 n (1 - e^{-\theta\delta_n}) + 1 + e^{-\theta\delta_n}}{1 + e^{-\theta\delta_n}}$$

and, since $\text{Var}(\bar{X}_n) = \frac{(1 - e^{-2\theta\delta_n}) + \frac{2}{n} e^{-\theta\delta_n} (e^{-\theta n\delta_n} - 1)}{2n\theta(1 - e^{-\theta\delta_n})^2}$, asymptotically we get that, if $\delta_n \rightarrow 0$, $n\delta_n \rightarrow S > 0$,

$$\bar{m}_n \prec_{n \rightarrow \infty} \bar{X}_n \iff (m - m_0)^2 \leq \frac{(1 + 2u^2\theta^2 S)(\theta S - 1 + e^{-\theta S})}{\theta^3 S^2}$$

while if $\delta_n \rightarrow 0$, $n\delta_n \rightarrow \infty$, we get the equivalence : $\bar{m}_n \prec_{n \rightarrow \infty} \bar{X}_n \iff (m - m_0)^2 \leq 2u^2$. Again, the same results are obtained for predictors.

7.2. **Estimation of ρ .** In the case where m is known (one may set $m = 0$), we now choose $N(\rho_0, v^2)$ as a prior for $\rho = e^{-\theta\delta_n}$, with $0 < \rho_0 < 1$ and $v > 0$. Note that this prior is reasonable as soon as ρ_0 is not too far from 1 and v not too large. Using again the

TABLE 3. L^2 -prediction error (m unknown) for MLE predictor and percentage variation of L^2 -prediction error for others in the case where $\theta = 1$, $H = 1$, $u^2 = 1$ and $\delta = 0.1$.

	n=15			n=30			n=50			n=100		
MLE	0.548			0.499			0.488			0.464		
Mean (%)	2.76			2.82			1.61			0.35		
CMLE (%)	27.97			10.54			5.62			1.35		
	a=4	a=5	a=7	a=4	a=5	a=7	a=4	a=5	a=7	a=4	a=5	a=7
Bay (%)	-4.87	-8.52	5.77	-2.79	-4.72	4.78	-1.30	-2.63	2.39	-0.62	-1.08	1.08
CMA2 (%)	-1.04	-12.86	33.54	-5.3	-5.02	16.05	.03	-2.28	6.77	-.35	-0.99	2.00

conditional likelihood, one obtains the expression :

$$\tilde{L}(X_{\delta_n}, \dots, X_{n\delta_n}, \rho) = \frac{1}{(\sigma_{\delta,\theta}\sqrt{2\pi})^n} \exp\left(-\frac{1}{2\sigma_{\delta,\theta}^2} \sum_{i=1}^n (X_{i\delta_n} - \rho X_{(i-1)\delta_n})^2\right) \times \frac{1}{v\sqrt{2\pi}} \exp\left(-\frac{1}{2v^2}(\rho - \rho_0)^2\right).$$

Since $\sigma_{\delta,\theta}^2$ depends on ρ , we make the approximation $\sigma_{\delta,\theta} \sim \delta$ for obtaining the posterior distribution $\mathcal{N}(\frac{B}{A}, \frac{1}{A})$ where $A = \frac{1}{\delta_n} \sum_{i=1}^n X_{(i-1)\delta_n}^2 + \frac{1}{v^2}$ and $B = \frac{1}{\delta_n} \sum_{i=1}^n X_{(i-1)\delta_n} X_{i\delta_n} + \frac{\rho_0}{v^2}$, hence the ‘‘Bayesian’’ estimator takes the form

$$(7.8) \quad \tilde{\rho}_n = \frac{\sum_{i=1}^n X_{(i-1)\delta_n} X_{i\delta_n} + \frac{\rho_0 \delta_n}{v^2}}{\sum_{i=1}^n X_{(i-1)\delta_n}^2 + \frac{\delta_n}{v^2}}.$$

Comparison with the conditional MLE

$$(7.9) \quad \hat{\rho}_n = \frac{\sum_{i=1}^n X_{(i-1)\delta_n} X_{i\delta_n}}{\sum_{i=1}^n X_{(i-1)\delta_n}^2}$$

is rather intricate and will be illustrated numerically in the next section.

7.3. Simulation. For $\theta \in \{0.5, 1, 2\}$, $m = 5$, various sample sizes n and values of δ , 5000 replications of Ornstein-Uhlenbeck sample paths are computed from the autoregressive relation (7.1). First, for known θ but m unknown, we compare various predictors of $X_{n\delta+H}$, $H = h\delta$ and $H = 0.5, 1$ or 2 , defined by $\mathbf{m}(1 - e^{-\theta h\delta}) + e^{-\theta h\delta} X_{n\delta}$ where \mathbf{m} refers to estimators which are either :

- non bayesian : MLE with m_n defined in (7.3), Mean \bar{X}_n , CMLE with \check{m}_n defined in (7.6),
- or bayesian : Bayes with \hat{m}_n defined in (7.4), CMA1 with \tilde{m}_n defined in (7.5) ($u^2 = 1$) and CMA2 with \bar{m}_n defined in (7.7) ($u^2 = 1$).

Among all non bayesian estimators and in all cases, it emerges that MLE outperforms the other two, with a very poor behaviour of the CMLE toward the others, a fact already noticed by Cox (1991). For this reason, our results below do not report the obtained values for CMA1, because of its too high sensitivity toward CMLE. In Table 3, we give the rounded empirical L^2 -prediction error of the MLE, and for comparison, the percentage variations observed for the others predictors for $\theta = 1$ and $\delta = 0.1$. It appears that all errors decrease

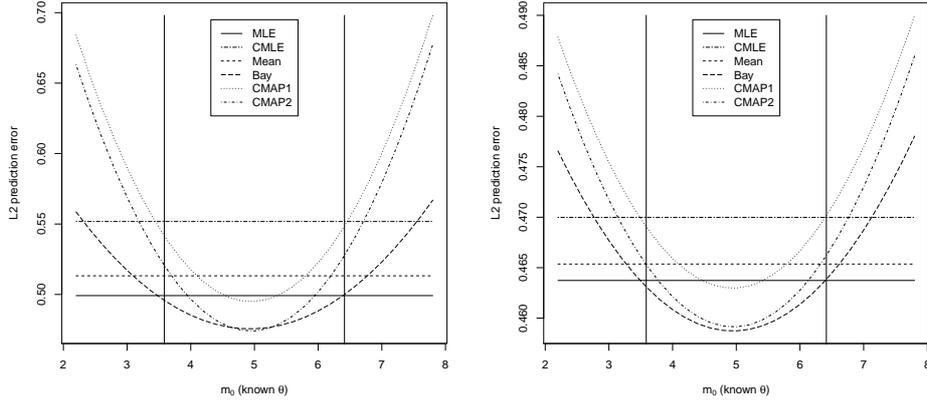


FIGURE 2. L^2 -prediction error for m unknown ($m = 5$) and $\theta = 1$ (known), $\delta = 0.1$ in terms of m_0 with $\mathcal{N}(m_0, 1)$ prior : MLE (plain), CMLE (twodash), Mean (dashed), Bayes (longdash), CMAP1 (dotted), CMAP2 (dotdash) when $u^2 = 1$. Vertical lines corresponds to $m_0 = 5 \pm \sqrt{2}u^2$. On the left : $n = 30$ ($S = 3, S + h * \delta = 4$), on the right : $n = 100$ ($S = 10, S + h * \delta = 11$).

as n increases, and Bayes predictors are highly competitive for small sample sizes and good choice of priors, namely $\mathbb{M} \sim \mathcal{N}(m_0, 1)$, with $m_0 \in \left] 5 - \sqrt{2 + (S + 2)^{-1}}, 5 + \sqrt{2 + (S + 2)^{-1}} \right[$ or asymptotically, $S = n\delta \rightarrow \infty$, $m_0 \in \left] 5 - \sqrt{2}, 5 + \sqrt{2} \right[$, see Section 7.1.1. By this way, errors are significantly reduced for $a = 4$ or 5 and n less than 50 , while a bad choice like $m_0 = 7$ damages them dramatically. It appears also that CMAP2 has the smallest errors but only on a small area around m , the Bayesian predictor (with \hat{m}_n defined in (7.4)) being more robust against the choice of m_0 . These results are confirmed in the Figure 2 where errors are given in term of m_0 : as expected, we obtain parabolic curves for bayesian predictors. Again, the Bayesian setting improves the errors for good choices of prior (especially for small values of δ and $n\delta$ where MLE is not so good) and otherwise deteriorate it.

TABLE 4. L^2 -prediction error (m unknown) for MLE predictor and percentage variation of L^2 -prediction error for others in the case where $\theta = 1, H = 1, u^2 = 1$ and $m_0 \in \{4, 5, 7\}$.

n	10			20			50			100			
	δ	0.1	0.2	0.5	0.1	0.2	0.5	0.1	0.2	0.5	0.1	0.2	0.5
MLE		.586	.531	.488	.531	.498	.464	.488	.464	.441	.464	.461	.444
Mean (%)		2.77	2.46	1.92	2.40	2.13	.44	1.61	.36	.16	.35	.16	-.06
CMLE (%)		47.01	17.69	5.73	17.56	7.38	1.40	5.62	1.36	0.24	1.35	.47	-.05
Bay_4 (%)		-5.06	-4.23	-1.31	-4.22	-1.84	-.65	-1.30	-.62	-.03	-.62	-.31	-.04
Bay_5 (%)		-10.48	-6.66	-2.68	-6.65	-3.41	-1.11	-2.63	-1.08	-0.23	-1.08	-0.37	-0.05
Bay_7 (%)		4.26	6.58	2.47	6.57	3.32	1.11	2.39	1.08	0.11	1.08	0.49	0.12
CMAP2_4 (%)		1.89	-1.70	.26	-1.70	-.13	-.30	.03	-.34	.13	-.35	-.18	-.10
CMAP2_5 (%)		-17.35	-9.23	-2.13	-9.23	-3.20	-.96	-2.28	-.98	-.09	-.99	-.26	-.12
CMAP2_7 (%)		46.63	26.12	7.21	25.96	10.04	2.15	6.77	2.02	.31	2.00	.75	.07

In Table 4, we compare the obtained errors for varying values of δ , while in Table 5 the influence of θ is measured. First it appears, that obtained errors depend only on $S = n\delta$, and not on the individual values of n and δ (see the bold errors). It is not a surprise since examination of L^2 -risks shows that leading terms are of order $n\delta$ for each estimators. Moreover errors are much larger as δ and—or θ are small. Again, it agrees with our theoretical framework since more observations are correlated (more δ is small), more covariances are important and the overall risk is degraded. Also, weak values of θ corresponds to variables with high variance since $\text{Var}(X_1) = (2\theta)^{-1}$, and prediction is more difficult in this case. Finally, errors are represented in term of n in Figure 3 (left) : not surprisingly, errors decrease and estimators are asymptotically equivalent.

TABLE 5. L^2 -prediction error (m unknown) for MLE predictor and percentage variation of L^2 -prediction error for other in the case where $n = 20$, $\delta = 0.1$, $u^2 = 1$, and $m_0 \in \{4, 5, 7\}$.

θ	0.5			1			2		
H	0.5	1	2	0.5	1	2	0.5	1	2
MLE	0.421	0.728	1.138	.34	.531	.677	.249	.303	.333
Mean (%)	0.51	1.32	2.23	1.81	2.4	4.07	1.88	2.72	2.85
CMLE (%)	15.03	27.59	48.31	10.76	17.56	27.37	6.04	9.76	10.65
Bay_4 (%)	-3.41	-5.9	-9.32	-2.72	-4.22	-6.09	-1.04	-0.92	-2.15
Bay_5 (%)	-5.27	-9.31	-15.56	-4.11	-6.65	-9.82	-1.99	-2.97	-3.58
Bay_7 (%)	2.18	4.31	5.68	4.04	6.57	9.24	1.82	1.7	3.87
CMAP2_4 (%)	-2.2	-3.33	-4.77	-1.07	-1.7	-1.96	0.67	1.85	0.33
CMAP2_5 (%)	-7.41	-12.86	-21.55	-5.5	-9.23	-13.28	-1.35	-2.1	-2.9
CMAP2_7 (%)	13.2	24.79	38.55	16.06	25.96	37.82	6.56	8.36	12.21

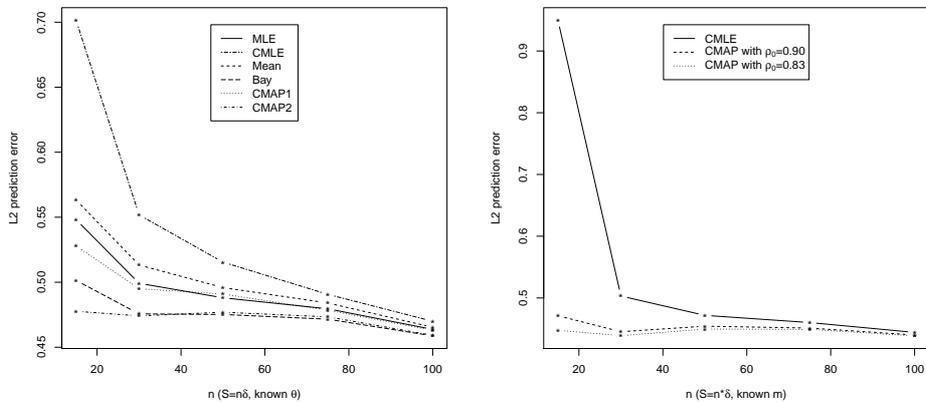


FIGURE 3. On the left : L^2 prediction error for unknown m , with prior $\mathcal{N}(5, 1)$, known θ ($\theta = 1$) in terms of n when $\delta = 0.1$: MLE (plain), CMLE (twodash), Mean (dashed), Bayes (longdash), CMAP1 (dotted), CMAP2 (dotdash). On the right : L^2 prediction error for unknown ρ , prior $\mathcal{N}(\rho_0, 10^{-2})$ in terms of n when $\delta = 0.1$: CMLE (plain), CBayes with $\rho_0 = 0.9$ (dashed), CBayes (dotted) with $\rho_0 = 0.83$.

TABLE 6. L^2 -prediction error (θ unknown) for MLE predictor and percentage variation of L^2 -prediction error for others in the case where $H = 1$, $\delta = 0.1$, $v^2 = 0.01$, and $\rho_0 \in \{0.5, 0.75, 0.85, 0.9\}$.

θ		0.5				1				2			
n = 20	CMLE	0.83				0.503				0.266			
ρ_0		0.5	0.75	0.85	0.9	0.5	0.75	0.85	0.9	0.5	0.75	0.85	0.9
“Bayes” (%)		-8.41	-14.67	-14.90	-13.55	-7.25	-12.05	-12.66	-11.48	-5.82	-6.56	-5.42	-3.21

θ		0.5				1				2			
n = 100	CMLE	0.679				0.444				0.242			
ρ_0		0.5	0.75	0.85	0.9	0.5	0.75	0.85	0.9	0.5	0.75	0.85	0.9
“Bayes” (%)		3.76	-0.25	-1.04	-1.15	0.89	-1.13	-1.22	-0.95	-0.44	-0.77	-0.46	-0.03

Concerning prediction when θ is unknown (m known), we have computed the two predictors derived from the estimators given by (7.7) (CMLE) and (7.8) (“Bayes”). The Figure 3 (right) that errors decrease with n and Bayes predictors are much better for small values of n . A noteworthy result is that errors are significantly improved for any choice of prior, at least for n small : see Table 6 for $n = 20$ and Figure 4 (left) for $n = 30$. This last conclusion may be tempered by the possibly bad behaviour of the CMLE in this framework. The Bayes predictor is more sensitive to the prior for $n = 100$ (and ρ_0 larger than 0.8).

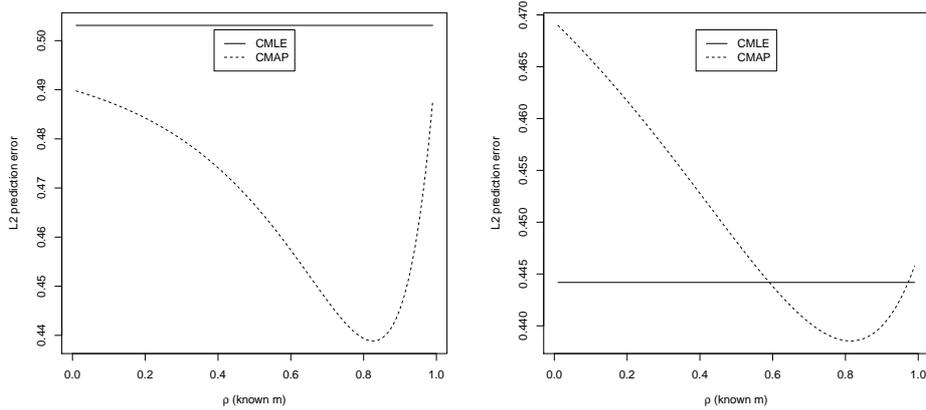


FIGURE 4. L^2 prediction error for $\rho = \exp(-\theta\delta)$ unknown ($\theta = 1$ and $\delta = 0.1$) and $m = 5$ (known) in terms of ρ_0 with $\mathcal{N}(\rho_0, 1)$ prior : CMLE (plain horizontal), Bayes predictor (dashed) when $v^2 = 10^{-2}$. On the left : $n = 30$ ($S = 3, S + h * \delta = 4$), on the right : $n = 100$ ($S = 10, S + h * \delta = 11$).

REFERENCES

- Bosq D (2012) A note on Bayesian prediction. Rend Circ Mat Palermo Special issue, to appear
- Bosq D, Blanke D (2007) Prediction and inference in large dimensions. Wiley series in probability and statistics, Wiley-Dunod
- Cox DD (1991) Gaussian likelihood estimation for nearly nonstationary AR(1) processes. Ann Statist 19(3):1129–1142
- Grenander U (1981) Abstract inference. Wiley, New York
- Lehmann EL, Casella G (1998) Theory of point estimation, 2nd edn. Springer, New-York
- Liptser RS, Shiryaev AN (2001) Statistics of random processes I, II, 2nd edn. Springer, New-York
- Thompson B, Vladimirov I (2005) Bayesian parameter estimation and prediction in mean reverting stochastic diffusion models. Nonlinear Analysis 63(5-7):e2367–e2375
- Yatracos YG (1992) On prediction and mean-square error. Canad J Statist 20(2):187–200

UNIVERSITÉ D'AVIGNON ET DES PAYS DE VAUCLUSE, LABORATOIRE DE MATHÉMATIQUES D'AVIGNON,
33 RUE LOUIS PASTEUR, 84000 AVIGNON, FRANCE

E-mail address: delphine.blanke@univ-avignon.fr

UNIVERSITÉ PIERRE ET MARIE CURIE - PARIS 6, LABORATOIRE DE STATISTIQUE THÉORIQUE ET
APPLIQUÉE, 4 PLACE JUSSIEU, 75252 PARIS CEDEX 5, FRANCE

E-mail address: denis.bosq@upmc.fr