BAYESIAN PREDICTION FOR STOCHASTIC PROCESSES. THEORY AND APPLICATIONS

DELPHINE BLANKE AND DENIS BOSQ

ABSTRACT. In this paper, we adopt a Bayesian point of view for predicting real continuous-time processes. We give two equivalent definitions of a Bayesian predictor and study some properties: admissibility, prediction sufficiency, non-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 for stochastic processes (see for example Kutoyants (2004) for the asymptotic point of view) while Bayesian prediction does not appear very much in statistical literature. Some authors have studied the case of linear processes (see Díaz, 1990; Sáfadi and Morettin, 2000, 2003) but continuous time is not often considered. 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 continuous-time processes. Note that we don't consider prediction for the linear model, a somewhat different topic which has been extensively studied in literature. In fact, our main goal is to compare efficiency of Bayesian predictors with non-Bayesian ones, especially if we have few data at our disposal. Various simulations illustrate the obtained results.

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 predictor, thus it is more convenient to substitute the conditional expectation with the conditional mode. An alternative method consists in computing the Bayesian estimator or the maximum a posteriori (MAP) and to plug it in $\mathbb{E}_\theta(Y|X)$. 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 simple case of Poisson process prediction. We compare the unbiased efficient predictor with the Bayesian and the MAP ones. Concerning diffusion processes, note that Thompson and Vladimirov (2005) obtain fine results for Bayesian prediction but without comparison with classical predictors.

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

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

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.

2. The prediction model

In the non Bayesian context, let (X, Y) be a random vector, defined on some Probability space and with values in a measurable space $(F \times G, \mathcal{F} \otimes \mathcal{G})$. In the following, F and G will be $\mathbb{N}$ or $\mathbb{R}^k$, $k \geq 1$. (X, Y) has distribution $(\mathbb{P}_\theta, \theta \in \Theta)$ where θ is the unknown parameter and Θ is an open set in $\mathbb{R}$. We suppose that $\mathbb{P}_\theta$ has a density $f(x, y, \theta)$ with respect to a σ-finite measure $\lambda \otimes \mu$.

One observes X and wants to predict Y. Actually, it is possible to consider the more general problem '*predict* $Z = \ell(X, Y, \theta)$ *given* X' (cf. Yatracos, 1992).

In this paper, we suppose that Z is real valued and denote $p(X)$ (or $q(X)$) a statistical predictor. Then, if Z, $p(X)$ and $q(X)$ are square integrable, a classical preference relation is

$$p(X) \prec q(X) \ \ (Z) \Longleftrightarrow \mathbb{E}_\theta(p(X) - Z)^2 \leq \mathbb{E}_\theta(q(X) - Z)^2, \ \theta \in \Theta$$

where '(Z)' means 'for predicting Z' and $\mathbb{E}_\theta$ is the expectation taken with respect to the distribution $\mathbb{P}_\theta$.

Now, let $\mathbb{E}_\theta(Z|X)$ be the conditional expectation of Z given X associated with the distribution $\mathbb{P}_\theta$. The next lemma is simple but important.

Lemma 2.1. *We have* $p(X) \prec q(X) \ \ (Z) \Longleftrightarrow p(X) \prec q(X) \ \ \big(\mathbb{E}_\theta(Z|X)\big)$.

Proof. The result directly follows from the Pythagoras theorem, since:

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

and

$$\mathbb{E}_\theta(q(X) - Z)^2 = \mathbb{E}_\theta\big(q(X) - \mathbb{E}_\theta(Z|X)\big)^2 + \mathbb{E}_\theta(\mathbb{E}_\theta(Z|X) - Z)^2.$$

□

This lemma shows that predicting Z or predicting $\mathbb{E}_\theta(Z|X)$ is the same problem.

Note that prediction theory has some similarity but also some difference with estimation theory. In the sequel, 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. **The Bayesian predictor.** In the Bayesian framework, we suppose that $\mathbb{T}$ is a random variable with prior distribution τ over Θ, and admitting a density $\varphi(\theta)$ with respect to a σ-finite measure ν (cf. Lehmann and Casella, 1998).

Thus, we may consider the scheme

$$(\Omega, \mathcal{A}, \mathbb{P}) \xrightarrow{(X,Y,\mathbb{T})} (F \times G \times \Theta, \mathcal{F} \otimes \mathcal{G} \otimes \mathcal{B}_\theta),$$

where $(\Omega, \mathcal{A}, \mathbb{P})$ is a probability space, $(X, Y, \mathbb{T})$ a random vector and $\mathcal{B}_\theta$ the σ-algebra of Borel sets over Θ. Now, we denote $\mathbb{Q}$ the distribution of $(X, Y, \mathbb{T})$ and we consider the following regularity assumption:

Assumption 3.1.
$\mathbb{Q}$ *admits a strictly positive density* $f(x, y, \theta)\varphi(\theta)$ *over* $F \times G \times \Theta$, *with respect to the* σ*-finite measure* $\lambda \otimes \mu \otimes \nu$. *In addition,* f *and* φ *are supposed to be continuous with respect to* θ *on* Θ.

Note that in practice, λ, μ and ν can be the Lebesgue measure or the counting measure. Also, remark that similar results can be derived under a more general version of Assumption 3.1, namely the existence of a common version $m(X,\theta)$ of $\mathbb{E}_\theta(Y|X)$ for all $\theta \in \Theta$ (see Blanke and Bosq, 2012).

Now, the Bayesian risk for prediction is

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

where $\mathbb{E}_\theta$ is expectation taken with respect to $\mathbb{P}_\theta$ and $\mathbb{E}$ is expectation taken with respect to $\mathbb{Q}$.

It follows that the Bayesian predictor is

$$p_0(X) = \operatorname*{argmin}_p r(p(X),Y) = \mathbb{E}(Y|X). \tag{1}$$

More precisely, we choose $p(X)$ under the form

$$p_0(X) = \int_G yf(y|X)\mathrm{d}\mu(y)$$

where

$$f(y|X) = \frac{\int_\Theta f(X,y,\theta)\varphi(\theta)\mathrm{d}\nu(\theta)}{\int_{G\times\Theta} f(X,y,\theta)\varphi(\theta)\mathrm{d}\mu(y)\mathrm{d}\nu(\theta)}$$

which ensures existence and uniqueness of $p_0(X)$ under Assumption 3.1. In the following, we set

$$m(X,\theta) = \mathbb{E}_\theta(Y|X) = \int_G yf_\theta(y|x)\mathrm{d}\mu(y), \theta \in \Theta$$

where

$$f_\theta(y|x) = \frac{f(x,y,\theta)}{\int_F f(x,y,\theta)\mathrm{d}\lambda(x)}.$$

Remark 3.1. *If Assumption 3.1 holds, the relation* $\mathbb{E}(Y|X) = \mathbb{E}\big(\mathbb{E}(Y|X,\mathbb{T})\big|X\big)$ *gives the following alternative form of* p_0*:*

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

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

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 $\mathbb{Q}$. If a strictly positive density does exist, the distribution of (X,Y) has marginal density $f(x,y) = \int_\Theta f(x,y,\theta)\varphi(\theta)\,\mathrm{d}\theta$ and, in fact, it suffices to compute $\operatorname*{argmax}_y f(x,y)$ (x fixed).

A related 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

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

where $\ell(x,\theta) = \int f(x,y,\theta)\,\mathrm{d}\mu(y)$, hence the MAP predictor

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

under 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 $\widetilde{\theta}(X)$, 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. Here, we suppose that $p_0(X)$ does exist and is defined by relation (1).

4.1. **Admissibility.** A Bayesian predictor is said to be unique if it differs, for any other Bayesian predictor, only on a set $\mathcal{N}$ with $\mathbb{P}_\theta(\mathcal{N}) = 0$ for all $\theta \in \Theta$ (see Lehmann and Casella, 1998, p. 323). Then, we have

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

Proof. If $p_0(X)$ is not admissible, there exists a predictor $p(X)$ such that

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

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

4.2. Y**-Sufficiency.** A statistic $S = S(X)$ is said to be Y-sufficient (or sufficient for predicting Y) 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 $\mathbb{Q}^S$, that does not depend on θ.
(b) X and Y are conditionally independent given S.

Note that this does not imply that $\mathbb{E}_\theta\big(Y|S(X)\big)$ is constant with respect to θ since the sufficient statistic is in the submodel generated by X (see example of the Poisson process in Section 5). If S is Y-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 Assumption 3.1 holds, then*

$$\mathbb{E}_\theta(Y|X) = \mathbb{E}_\theta\big(Y|S(X)\big), \quad \theta \in \Theta. \tag{2}$$

Proof. We have

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

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

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

Taking expectation and noting that $\mathbb{E}_\theta\big(Y \cdot \mathbb{E}_\theta(Y|X)\big) = \mathbb{E}_\theta\Big(\big(\mathbb{E}_\theta(Y|X)\big)^2\Big)$, entails

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

that is $\|\mathbb{E}_\theta(Y|X)\|^2_{L^2(\mathbb{P}_\theta)} = \big\|\mathbb{E}_\theta\big(Y|S(X)\big)\big\|^2_{L^2(\mathbb{P}_\theta)}$. This implies relation (2) since $\mathbb{E}_\theta\big(Y|S(X)\big)$ is the projection of $\mathbb{E}_\theta(Y|X)$ on $L^2_{S(X)}$. □

Note that, if (X, Y) has a strictly positive density of the form $f(x, y, \theta) = L(S(x), y, \theta)$, one obtains (2) by a direct computation. Concerning the Bayesian predictor, we have

Proposition 4.2. *If p_0 is unique and S is Y-sufficient, then*

$$p_0(X) = \mathbb{E}\big(Y|S(X)\big).$$

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

$$p_0(X) = p_1(X) = \mathbb{E}^{S(X)}\big(\mathbb{E}(Y|X)\big) = \mathbb{E}\big(Y|S(X)\big).$$

□

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

$$\mathbb{E}_\theta(Y|X) = A(X) + B(\theta)C(X) + D(\theta), \quad \theta \in \Theta \tag{3}$$

where A, $B \otimes C$, $D \in L^2(F \times \Theta, \mathcal{F} \otimes \mathcal{T}, \mathbb{Q}_{(X,\mathbb{T})})$, $\mathbb{Q}_{(X,\mathbb{T})}$ being the distribution of $(X, \mathbb{T})$. Then, the Bayesian predictor has also a special form:

Proposition 4.3. *Suppose that Assumption 3.1 is fulfilled. If $\mathbb{E}_\theta(Y|X)$ satisfies* (3)*, the associated Bayesian predictor is given by*

$$p_0(X) = A(X) + \mathbb{E}(B(\mathbb{T})|X) \cdot C(X) + \mathbb{E}(D(\mathbb{T})|X). \tag{4}$$

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 (3) entails $m(X,\mathbb{T}) = A(X)+B(\mathbb{T})\cdot C(X)+D(\mathbb{T})$, and Remark 3.1 gives $p_0(X) = \mathbb{E}(m(X,\mathbb{T})|X)$ hence (4) from the properties of conditional expectation. The last assertion is a special case of (4). □

4.4. **Unbiasedness.** A predictor $p(X)$ of Y is said to be unbiased if $\mathbb{E}_\theta\, p(X) = \mathbb{E}_\theta(Y)$, $\theta \in \Theta$. 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*

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

where $\mathbb{E}$ denotes here expectation taken from $\mathbb{Q}_{(X,\mathbb{T})}$.

Proof. See Lehmann and Casella (1998, p. 234). □

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

$$\mathbb{E}_\theta(p_0(X)) = \mathbb{E}_\theta(Y), \quad \theta \in \Theta \tag{5}$$

then, $p_0(X)$ is an unbiased estimator of $\mathbb{E}_\theta(Y)$ but it is not necessarily a Bayesian estimator of $\mathbb{E}_\theta(Y)$. Recall that the Bayesian interpretation of (5) is:

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

Now, we have the following result:

Proposition 4.4. *If the Bayesian risk satisfies*

$$\mathbb{E}\big(p_0(X) - m(X,\mathbb{T})\big)^2 = 0 \tag{6}$$

then $p_0(X)$ is unbiased for predicting Y. Conversely under Assumption 3.1, if

$$m(X,\theta) = A(X) + D(\theta),\ \theta \in \Theta \tag{7}$$

and if $p_0(X)$ is unbiased, then (6) *holds.*

Proof. Relation (6) implies $p_0(X) = m(X, \mathbb{T})$, $\mathbb{Q}_{(X,\mathbb{T})}$ a.s., that is

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

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 (7) and (4) in Proposition 4.3 imply

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

Now, since $p_0(X)$ is unbiased, we have

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

where the last equality follows from $\mathbb{E}(Y|X, \mathbb{T}) = m(X, \mathbb{T})$ and a conditioning on $\mathbb{T}$. But by (7),

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

By identification, it means that the Bayesian estimator of $D(\mathbb{T})$ is also unbiased. Then Lemma 4.2 gives

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

□

In the more general case where $\mathbb{E}_\theta(Y|X)$ has the form (3) 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 for some $\theta_0 \in \Theta$, let us define a 'Bayesian type' predictor by

$$p_0(X) = \alpha\, p(X) + (1 - \alpha)m(X, \theta_0), \quad (0 < \alpha < 1), \tag{8}$$

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* (8)*. Then, if $p_0(X)$ is unbiased, it follows that*

$$\mathbb{E}_\theta\big(m(X, \theta)\big) = \mathbb{E}_\theta\big(m(X, \theta_0)\big), \quad \theta \in \Theta, \tag{9}$$

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

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

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

and taking expectation in (8) yields

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

hence, since $p(X)$ is unbiased, (9) follows. Now, if $S(X)$ is a Y-sufficient statistic, Lemma 4.1 entails $m(X, \theta) = \mathbb{E}_\theta\big(Y|S(X)\big)$, thus, (9) implies

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

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

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

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

For some $\theta_0 \in \Theta$, consider the 'Bayesian type' predictor

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

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

$$p_0 \prec p \Longleftrightarrow |\theta - \theta_0| \leq \Big(\frac{1+\alpha}{1-\alpha}\Big)^{\frac{1}{2}} \cdot \Big(\mathbb{E}_\theta\big(\overline{\theta}(X) - \theta\big)^2\Big)^{\frac{1}{2}}. \tag{10}$$

Proof. We have

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

then, since p is unbiased,

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

thus

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

and (10) follows. □

Remark 4.1. *If $X = X_{(n)} = (X_1, \ldots, X_n)$ and $\mathbb{E}_\theta\big(\overline{\theta}(X) - \theta\big)^2 = \frac{v^2}{n}$ then the condition becomes*

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

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

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

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 $Y = N_{S+h}$ $(h > 0)$, $(S > 0)$. This a classical scheme but of interest, since in this case, there exists an unbiased efficient predictor (see Bosq and Blanke, 2007). Since Lemma 2.1 shows that it is equivalent to predict $m(X,\theta) = \theta h + N_S$, one obtains the unbiased efficient predictor $p(N_S) = \frac{S+h}{S} N_S =: N_S + \theta_S h$ (with $\theta_S = \frac{N_S}{S}$).

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

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

First, 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 $p_0(N_t,\, 0 \le t \le S) = \mathbb{E}(N_{S+h}|N_S)$. The same property holds for the Bayes estimator given by

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

and, from Proposition 4.3, the Bayesian predictor is

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

To compare $\widehat{p}_0$ with p, note that $\widehat{\theta}_S = \frac{S}{b+S}\cdot\theta_S + \left(1-\frac{S}{b+S}\right)\cdot\frac{a}{b}$. We deduce that

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

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

$$\widehat{p}_0 \prec p \Longleftrightarrow (\theta-\theta_0)^2 \le \Big(\frac{1}{S}+\frac{2}{b}\Big)\theta. \tag{11}$$

Solving (11) in θ, we get that $p_0 \prec p$ iff

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

with $\Delta = \big(\theta_0 + \frac{1}{2S} + \frac{1}{b}\big)^2 - \theta_0^2$. Also, from (11), a sufficient condition, holding for all S, is $(\theta-\theta_0)^2 \le \dfrac{2}{b}\theta$ which gives $\theta \in \Big]\theta_0 + \frac{1}{b} - \sqrt{\widetilde{\Delta}}, \theta_0 + \frac{1}{b} + \sqrt{\widetilde{\Delta}}\Big[$ with $\widetilde{\Delta} = \frac{1}{b}(2\theta_0 + \frac{1}{b})$, that is $p_0 \prec p$ if

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

Clearly, one obtains the same result for comparing $\widehat{\theta}_S$ with θ_S. For example, if one chooses $a=1$, $b=\frac{1}{\theta_0}$ (so that $\mathbb{E}(\tau)=\theta_0$) then $\theta_1 = \frac{2-\sqrt{3}}{b}$ and $\theta_2 = \frac{2+\sqrt{3}}{b}$. If b is small, $\theta_2-\theta_1$ 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

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

hence $\widetilde{\theta}_S = \frac{N_S+a-1}{b+S}$ where we choose $a \ge 1$ 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

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

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. With that method, 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 empirical L^2-error of prediction:

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

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

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

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 (11) 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, it appears as less sensitive to the prior: indeed this overall error is governed by the probabilistic one, much more important in this case.

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 5.1, the L^2-error of prediction is plotted as a function of a for $\theta = 1$ and $S = 20$. As expected by (11), 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 related results of 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.

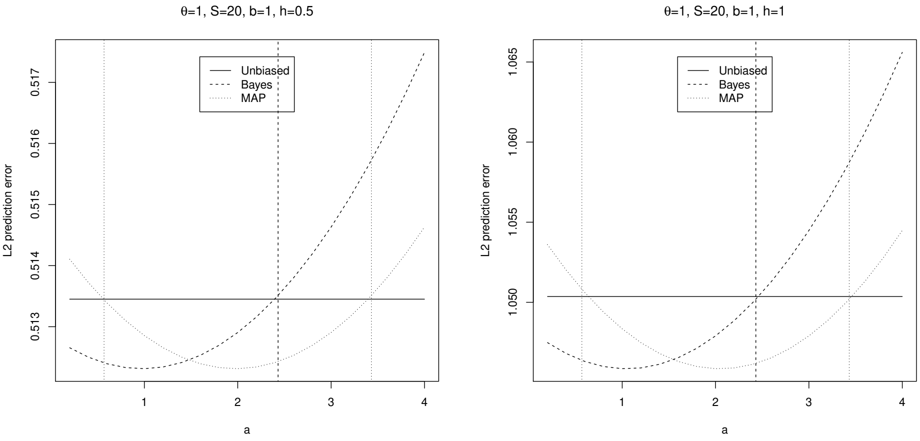

FIGURE 1. L^2 prediction error for $\theta = 1$ in terms of a with $\Gamma(a, 1)$ prior: UP (plain horizontal), BP (dashes), MAP (dots) for $S = 20$. Vertical lines corresponds to $a = 1 + \sqrt{S^{-1} + 2}$ (dashes) and $a = 2 \pm \sqrt{S^{-1} + 2}$ (dots). On the left : $h = 0.5$, on the right : $h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

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

	θ=0.5			**θ=5**			**θ=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)

6. Bayesian inference for the Ornstein-Uhlenbeck process

Consider a stationary version of the Ornstein-Uhlenbeck process (O.U.) defined by $X_t = m + \int_{-\infty}^{t} e^{-\theta(t-s)}\,\mathrm{d}W(s)$, $t \in \mathbb{R}$, $(m \in \mathbb{R},\ \theta > 0)$ where W is a standard bilateral Wiener process. Set $X_{0,t} = X_t - m$, $t \in \mathbb{R}$, then the likelihood of $X_{(S)} = (X_t,\ 0 \le t \le S)$ with respect to $X_{0,(S)} = (X_{0,t},\ 0 \le t \le S)$ is given by

$$L\big(X_{(S)}; m, \theta\big) = \exp\Big(-\frac{\theta m^2}{2}(2 + \theta S) + \theta m (X_0 + X_S + \theta \int_0^S X_t\,\mathrm{d}t)\Big) \tag{12}$$

(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 (12), it follows that the posterior density of $\mathbb{M}$ given $X_{(S)}$ is $\mathcal{N}\big(\dfrac{B}{A}, \dfrac{1}{A}\big)$ where $A = \theta(2 + \theta S) + \frac{1}{u^2}$ and $B = \theta Z_S + \frac{m_0}{u^2}$ with $Z_S = \big(X_0 + X_S + \theta \int_0^S X_t\,\mathrm{d}t\big)$. 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

$$\widehat{m}_S = \alpha_S\, m_S + (1-\alpha_S)m_0 \tag{13}$$

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

Asymptotic efficiency. The MLE m_S is efficient (cf. Bosq and Blanke, 2007, p. 28) and $\widehat{m}_S$ is asymptotically efficient since, from (13),

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

with $\alpha_S^2 \to 1$ as $S\to\infty$, $(1-\alpha_S)^2 = \mathcal{O}(S^{-2})$ 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 by 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

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

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*

$$\widehat{p}_{0,S} \prec p_S \iff |m-m_0| \le \Big(\frac{1}{\theta(2+\theta S)} + 2u^2\Big)^{\frac{1}{2}}$$

and $|m-m_0| \le u\sqrt{2}$ *implies* $\widehat{p}_{0,S} \prec p_S$ *for all* $S>0$.

The proof is straightforward since one has $\mathbb{E}_m(m_S-m)^2 = \big(\theta(2+\theta S)\big)^{-1}$. 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

$$L(X_{(S)}) = \exp\Big(-\frac{1}{2}(X_S^2 - X_0^2 - S) - \frac{\theta^2}{2}\int_0^S X_t^2\,\mathrm{d}t\Big),$$

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 $\varphi(x_{(S)}) = \frac{1}{\sqrt{\alpha v^2}}\exp\Big(-\frac{\theta_0^2}{2v^2} + \frac{\beta^2}{2\alpha}\Big)$ where $\alpha = \int_0^S x_s^2\,\mathrm{d}s + \frac{1}{v^2}$ 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 $\mathcal{N}\big(\frac{\beta}{\alpha}, \frac{1}{\alpha}\big)$, hence the Bayesian estimator of θ: $\widehat{\theta}_S = \frac{\beta}{\alpha} = \frac{\frac{1}{2}(S-X_S^2+X_0^2)+\theta_0 v^{-2}}{\int_0^S X_t^2\,\mathrm{d}t + v^{-2}}$ when the MLE is $\theta_S = \frac{\frac{1}{2}(S-X_S^2+X_0^2)}{\int_0^S X_t^2\,\mathrm{d}t}$, consequently

$$\widehat{\theta}_S = \gamma_S\,\theta_S + (1-\gamma_S)\theta_0 \quad \text{with} \quad \gamma_S = \frac{\int_0^S X_t^2\,\mathrm{d}t}{\int_0^S X_t^2\,\mathrm{d}t + v^{-2}}, \tag{14}$$

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

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

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

hence the Bayesian predictor $\widehat{p}_0(X_{(S)}) = \exp(-\frac{2\beta-h}{2\alpha} \cdot h) \cdot X_S$. 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 $\widetilde{p}(X_{(S)}) = e^{-\widehat{\theta}_S \cdot h} \cdot X_S$.

Finally, one may consider alternative priors, as well as, the translated exponential distribution with density $\varphi(\theta) = \eta \exp\big(-\eta(\theta-\theta_0)\big)\mathbb{1}_{]\theta_0,+\infty[}(\theta)$, $(\eta > 0,\, \theta_0 \geq 0)$. 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 $b = \int_0^S x_t^2 \,\mathrm{d}t$, the Bayesian estimator is given by $\widehat{\theta}_S = \int_{\theta_0}^\infty \theta\psi(\theta)\,\mathrm{d}\theta \Big/ \int_{\theta_0}^\infty \psi(\theta)\,\mathrm{d}\theta$ and can be numerically computed. 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

$$X_{n\delta} - m = e^{-\theta\delta}\big(X_{(n-1)\delta} - m\big) + \varepsilon_{n\delta}, \quad n \in \mathbb{Z} \tag{15}$$

and

$$\mathrm{Var}\,(\varepsilon_{n\delta}) = \frac{1 - e^{-2\theta\delta}}{2\theta} =: \sigma_{\delta,\theta}^2 \tag{16}$$

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 conditions as $n \to \infty$: $\delta_n \to 0$ and $n\delta_n \to \infty$ or $\delta_n \to 0$ and $n\delta_n \to S > 0$ 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 $X_0 - m, \varepsilon_{\delta_n}, \dots, \varepsilon_{n\delta_n} \sim \mathcal{N}(0, (2\theta)^{-1}) \otimes \mathcal{N}(0, \sigma_{\delta_n,\theta}^2)^{\otimes n}$, one may deduce that $(X_0 - m, X_{\delta_n} - m, \dots, X_{n\delta_n} - m)$ has the density

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

This yields

$$m_n = \frac{X_0 + X_{n\delta_n} + (1 - e^{-\theta\delta_n})\sum_{i=1}^{n-1} X_{i\delta_n}}{n(1 - e^{-\theta\delta_n}) + 1 + e^{-\theta\delta_n}} \tag{17}$$

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

$$L(X_0, X_{\delta_n}, \dots, X_{n\delta_n}, \mathbb{M}) = \Big(\frac{\theta}{\pi}\Big)^{\frac{1}{2}} \frac{1}{\sigma_{\delta_n,\theta}\sqrt{2\pi}} \times \exp\Big(- \theta(X_0 - \mathbb{M})^2$$

$$- \sum_{i=1}^{n} \frac{\big(X_{i\delta_n} - e^{-\theta\delta_n} X_{(i-1)\delta_n} - \mathbb{M}(1 - e^{-\theta\delta_n})\big)^2}{2\sigma^2_{\delta_n\,\theta}} \Big) \times \frac{1}{u\sqrt{2\pi}} \exp\big(-\frac{1}{2u^2}(\mathbb{M} - m_0)^2\big)$$

giving

$$\widehat{m}_n = \frac{X_0 + X_{n\delta_n} + (1 - e^{-\theta\delta_n}) \sum_{i=1}^{n-1} X_{i\delta_n} + (1 + e^{-\theta\delta_n})\frac{m_0}{2\theta u^2}}{n(1 - e^{-\theta\delta_n}) + (1 + e^{-\theta\delta_n})(1 + \frac{1}{2\theta u^2})}. \tag{18}$$

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 $\mathbb{E}_m(m_n - m)^2 = \frac{1+e^{-\theta\delta_n}}{2\theta\left(n(1-e^{-\theta\delta_n})+1+e^{-\theta\delta_n}\right)}$ yielding the equivalence: $\widehat{p}_{0,n} \prec p_n \Leftrightarrow (m - m_0)^2 \leq 2u^2 + \frac{1+e^{-\theta\delta_n}}{2\theta\left(1+e^{-\theta\delta_n}+n(1-e^{-\theta\delta_n})\right)}$. Asymptotically, we get if $\delta_n \xrightarrow[n\to\infty]{} 0$, $n\delta_n \xrightarrow[n\to\infty]{} S > 0$, $\widehat{p}_{0,n} \prec_{n\to\infty} p_n$ is equivalent to $(m - m_0)^2 \leq 2u^2 + \frac{1}{\theta(\theta S+2)}$. The condition $S \to \infty$ implying in turn the equivalence $\widehat{p}_{0,n} \prec_{n\to\infty} p_n \Leftrightarrow (m - m_0)^2 \leq 2u^2$, which are the same results as in the continuous case (cf. Proposition 6.1). If $n\delta_n \to S > 0$, 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 $(X_{\delta_n}, \dots, X_{n\delta_n}, \mathbb{M})$:

$$\widetilde{L}(X_{\delta_n}, \dots, X_{n\delta_n}, \mathbb{M}) = \frac{1}{(\sigma_{\delta_n,\theta}\sqrt{2\pi})^n} \exp\Big(- \frac{1}{2\sigma^2_{\delta_n,\theta}} \sum_{i=1}^{n} \Big((X_{i\delta_n} - \exp(-\theta\delta_n) X_{(i-1)\delta_n})$$

$$+ \mathbb{M}(\exp(-\theta\delta_n) - 1)\Big)^2\Big) \times \frac{1}{u\sqrt{2\pi}} \exp\Big(- \frac{1}{2u^2}(\mathbb{M} - m_0)^2\Big),$$

where $\sigma^2_{\delta_n,\theta}$ is defined by (16). Now:

$$\ln \widetilde{L} = c - \frac{1}{2\sigma^2_{\delta_n,\theta}} \sum_{i=1}^{n} \Big(X_{i\delta_n} - \exp(-\theta\delta_n) X_{(i-1)\delta_n} + \mathbb{M}(\exp(-\theta\delta_n) - 1)\Big)^2 - \frac{(\mathbb{M} - m_0)^2}{2u^2},$$

where c does not depend on n. Since we are in the Gaussian case, the conditional mode and the conditional expectation coincide and it follows that the Bayesian estimator is now given by

$$\widetilde{m}_n = \frac{(1 - \exp(-\theta\delta_n)) \sum_{i=1}^{n} (X_{i\delta_n} - \exp(-\theta\delta_n) X_{(i-1)\delta}) + m_0 \frac{\sigma^2_{\delta_n,\theta}}{u^2}}{(1 - \exp(-\theta\delta_n))^2 n + \frac{\sigma^2_{\delta_n,\theta}}{u^2}}, \tag{19}$$

while the conditional MLE takes the form

$$\breve{m}_n = \frac{\sum_{i=1}^n (X_{i\delta_n} - \exp(-\theta\delta_n)X_{(i-1)\delta})}{(1-\exp(-\theta\delta_n))n}, \tag{20}$$

We may slightly modify the estimator (19) for obtaining

$$\overline{m}_n = \beta_n \overline{X}_n + (1-\beta_n)m_0 \tag{21}$$

with $\overline{X}_n = n^{-1}\sum_{i=1}^n X_{i\delta_n}$ and $\beta_n = \dfrac{(1-\exp(-\theta\delta_n))^2}{(1-\exp(-\theta\delta_n))^2 + \frac{\sigma^2_{\delta_n,\theta}}{nu^2}}$. 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 $\mathrm{Var}\,(\overline{X}_n) = \frac{(1-e^{-2\theta\delta_n}) + \frac{2}{n}e^{-\theta\delta_n}(e^{-\theta n\delta_n}-1)}{2n\theta(1-e^{-\theta\delta_n})^2}$, asymptotically we get that, if $\delta_n \to 0, n\delta_n \to S > 0$,

$$\overline{m}_n \prec_{n\to\infty} \overline{X}_n \Longleftrightarrow (m-m_0)^2 \le \frac{(1+2u^2\theta^2 S)(\theta S - 1 + e^{-\theta S})}{\theta^3 S^2}$$

while if $\delta_n \to 0, n\delta_n \to \infty$, we get the equivalence: $\overline{m}_n \prec_{n\to\infty} \overline{X}_n \Leftrightarrow (m-m_0)^2 \le 2u^2$. 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 conditional likelihood, one obtains the expression:

$$\widetilde{L}(X_{\delta_n}, \dots, X_{n\delta_n}, \rho) = \frac{1}{(\sigma_{\delta_n,\theta}\sqrt{2\pi})^n} \exp\Big(-\frac{1}{2\sigma^2_{\delta_n,\theta}} \sum_{i=1}^n \big(X_{i\delta_n} - \rho X_{(i-1)\delta_n}\big)^2\Big) \\ \times \frac{1}{v\sqrt{2\pi}} \exp\big(-\frac{1}{2v^2}(\rho-\rho_0)^2\big).$$

Since $\sigma^2_{\theta,\delta}$ 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 $A = \dfrac{1}{\delta_n}\displaystyle\sum_{i=1}^n X^2_{(i-1)\delta_n} + \dfrac{1}{v^2}$ and $B = \dfrac{1}{\delta_n}\displaystyle\sum_{i=1}^n X_{(i-1)\delta_n}X_{i\delta_n} + \dfrac{\rho_0}{v^2}$, hence the 'Bayesian' estimator takes the form

$$\widetilde{\rho}_n = \frac{\sum_{i=1}^n X_{(i-1)\delta_n}X_{i\delta_n} + \frac{\rho_0\delta_n}{v^2}}{\sum_{i=1}^n X^2_{(i-1)\delta_n} + \frac{\delta_n}{v^2}}. \tag{22}$$

Comparison with the conditional MLE

$$\widehat{\rho}_n = \frac{\sum_{i=1}^n X_{(i-1)\delta_n}X_{i\delta_n}}{\sum_{i=1}^n X^2_{(i-1)\delta_n}} \tag{23}$$

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 (15). 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 $\mathfrak{m}(1-e^{-\theta h\delta}) + e^{-\theta h\delta}X_{n\delta}$ where $\mathfrak{m}$ refers to estimators which are either:

- non Bayesian: MLE with m_n defined in (17), Mean $\overline{X}_n$, CMLE with $\breve{m}_n$ defined in (20),
- or Bayesian: Bayes with $\widehat{m}_n$ defined in (18), CMAP1 with $\widetilde{m}_n$ defined in (19) ($u^2 = 1$) and CMAP2 with $\overline{m}_n$ defined in (21) ($u^2 = 1$).

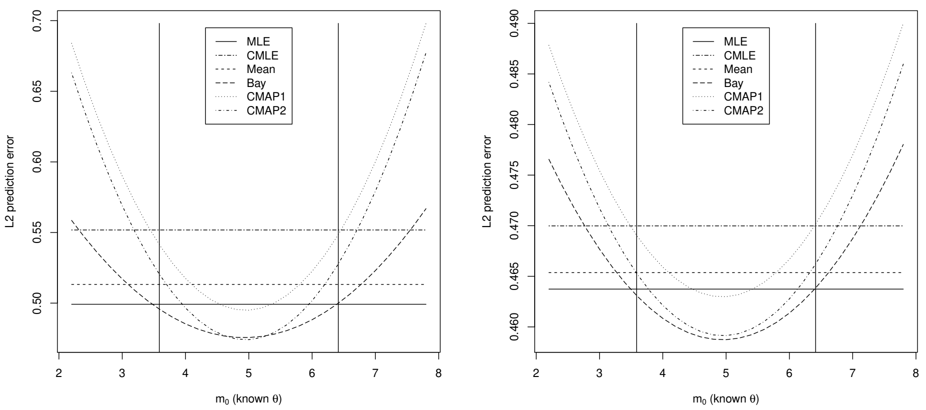

FIGURE 2. L^2-prediction error for m unknown ($m = 5$) and $\theta = 1$ (known), $\delta = 0.1$ in terms of m_0 with $\mathcal{N}(m_0, 1)$ prior: MLE (plain), CMLE (twodash), Mean (dashed), Bayes (longdash), CMAP1 (dotted), CMAP2 (dotdash) when $u^2 = 1$. Vertical lines corresponds to $m_0 = 5 \pm \sqrt{2u^2}$.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

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 following results do not report the obtained values for CMAP1, because of its too bad behaviour governed by the 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, in the case of $\theta = 1$ and $\delta = 0.1$. It appears

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**		
MLE	**0.548**			**0.499**		
Mean (%)	2.76			2.82		
CMLE (%)	27.97			10.54		
	$\mathbf{m_0 = 4}$	$\mathbf{m_0 = 5}$	$\mathbf{m_0 = 7}$	$\mathbf{m_0 = 4}$	$\mathbf{m_0 = 5}$	$\mathbf{m_0 = 7}$
Bay (%)	-4.87	-8.52	5.77	-2.79	-4.72	4.78
CMAP2 (%)	-1.04	-12.86	33.54	-.53	-5.02	16.05

	n=50			**n=100**		
MLE	**0.488**			**0.464**		
Mean (%)	1.61			0.35		
CMLE (%)	5.62			1.35		
	$\mathbf{m_0 = 4}$	$\mathbf{m_0 = 5}$	$\mathbf{m_0 = 7}$	$\mathbf{m_0 = 4}$	$\mathbf{m_0 = 5}$	$\mathbf{m_0 = 7}$
Bay (%)	-1.30	-2.63	2.39	-0.62	-1.08	1.08
CMAP2 (%)	0.03	-2.28	6.77	-0.35	-0.99	2.00

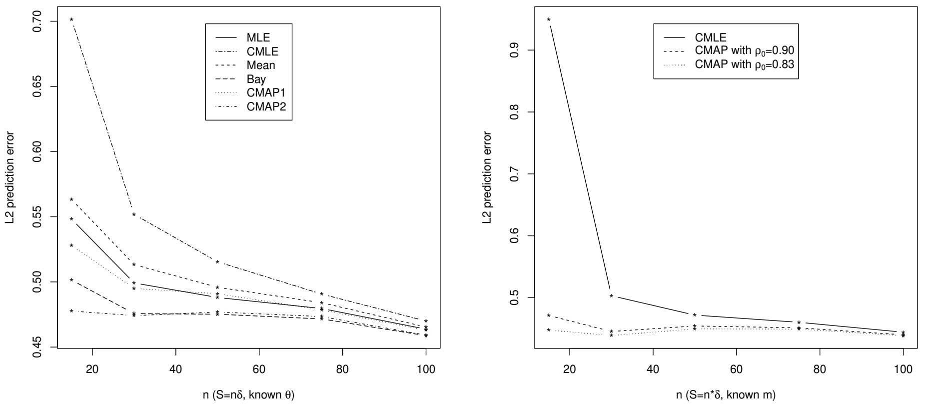

FIGURE 3. On the left: L^2 prediction error for unknown m, with prior $\mathcal{N}(5,1)$, known θ ($\theta = 1$) in terms of n when $\delta = 0.1$:MLE (plain), CMLE (twodash), Mean (dashed), Bayes (longdash), CMAP1 (dotted), CMAP2 (dotdash). On the right: L^2 prediction error for unknown ρ, prior $\mathcal{N}(\rho_0,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

that all errors decrease 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, $S = n\delta \to \infty$, $m_0 \in \left]5 - \sqrt{2}, 5 + \sqrt{2}\right[$, see Section 7.1.1. By this way, errors are significantly reduced for $m_0 = 4$ 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 $\widehat{m}_n$ defined in (18)) being more robust against the choice of m_0. These results are confirmed in Figure 7.1 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 deteriorates 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

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

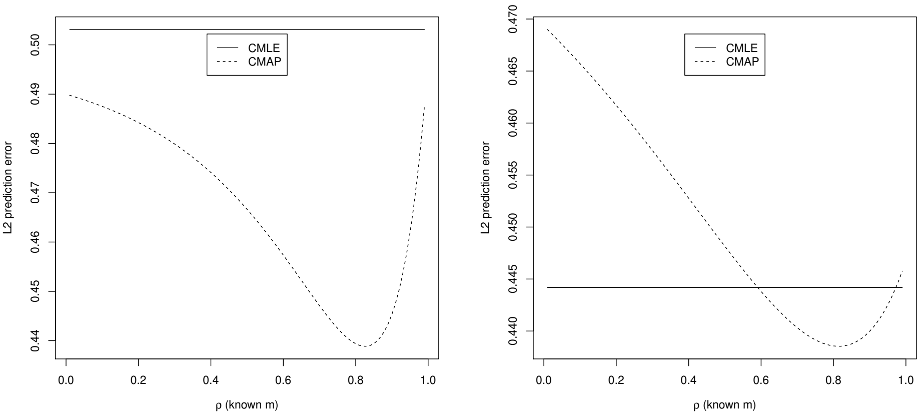

FIGURE 4. L^2 prediction error for $\rho = e^{-\theta\delta}$ unknown ($\theta = 1$, $\delta = 0.1$) and $m = 5$ (known) in terms of ρ_0 with $\mathcal{N}(\rho_0, 10^{-2})$ prior: CMLE (plain horizontal), Bayes predictor (dashed). On the left : $n = 30$ ($S = 3, S + h * \delta = 4$), on the right: $n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

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 / **n = 20**	**0.5**				**1**				**2**			
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 / **n = 100**	**0.5**				**1**				**2**			
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

In Table 4, we compare the obtained errors with 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 type 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, the errors are much larger as δ and—or θ are small. Again, it agrees with our theoretical framework since more δ is small, more important is the correlation, implying a degradation of the overall risk. Also, low values of θ corresponds to variables with high variance ($\mathrm{Var}\,(X_1) = (2\theta)^{-1}$), and prediction is more difficult in this case. Finally, errors are represented in term of n in Figure 7.2 (left): not surprisingly, errors decrease and estimators are asymptotically equivalent.

Concerning prediction when θ is unknown (m known), we have computed the two predictors derived from the estimators given by (21) (CMLE) and (22) ('Bayes'). Figure 7.2 (right) that errors decrease with n and Bayesian 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 7.3 (left) for $n = 30$. This last conclusion may be tempered by the possibly bad behaviour of the CMLE in this framework. Finally for $n = 100$, the Bayesian predictor is more sensitive to the prior (Figure 7.3, right).

References

Blanke D, Bosq D (2012) Bayesian prediction for stochastic processes. ArXiv e-prints url: `http://arxiv.org/abs/1211.2300v1`, `math.ST/1211.2300v1`

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

Díaz J (1990) Bayesian forecasting for AR(1) models with normal coefficients. Comm Statist Theory Methods 19(6):2229–2246

Grenander U (1981) Abstract inference. Wiley, New York

Kutoyants YA (2004) Statistical Inference for Ergodic Diffusion Processes. Springer Series In Statistics, Springer

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

Sáfadi T, Morettin PA (2000) Bayesian analysis of threshold autoregressive moving average models. Sankhyā Ser B 62(3):353–371

Sáfadi T, Morettin PA (2003) A Bayesian analysis of autoregressive models with random normal coefficients. J Stat Comput Simul 73(8):563–573

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

Avignon University, Laboratoire de Mathématiques d'Avignon, 33 rue Louis Pasteur, 84000 Avignon, France

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

Sorbonne Universités, UPMC Univ Paris 06, EA 3124, Laboratoire de Statistique Théorique et Appliquée, F-75005, 4 place Jussieu, Paris, France

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